Clinical biochemistry

Clinical Biochemistry

Edited by

prof. MUDr. Jaroslav Racek, DrSc.

MUDr. Daniel Rajdl, Ph.D.

Authors

Vladimír Bartoš, Milan Dastych, Milan Dastych jr., Tomáš Franěk, Milan Jirsa,

Marta Kalousová, Tomáš Karlík, Petr Kocna, Viktor Kožich, Michaela Králíková,

Alena Krnáčová, Pavlína Kušnierová, Richard Pikner, Věra Ploticová,

Richard Průša, Jaroslav Racek, Daniel Rajdl, Václav Senft, Vladimír Soška,

Drahomíra Springer, Kristian Šafarčík, Ivan Šebesta, Radka Šigutová,

Zdeněk Švagera, Eva Táborská, Libor Vítek, František Všianský,

Jiří Zadina, David Zeman, Tomáš Zima

Technical collaborator

Martin Navrátil

Published by Charles University

Karolinum Press

www.karolinum.cz

ebooks@karolinum.cz

Prague 2016

First edition

ISBN 978-80-246-3497-5 (pdf)

ISBN 978-80-246-3164-6 (epub)

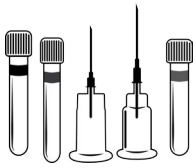
ISBN 978-80-246-3498-2 (mobi)

The publication has been partly created within project Klinická biochemie – inovovaná, interaktivní výuka e-learningem, reg. number: CZ.1.07/2.2.00/15.0048 and is co-funded by the European Social Fund and the state budget of the Czech Republic.

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Conntent

Intentionally Blank Page



CHAPTER 1

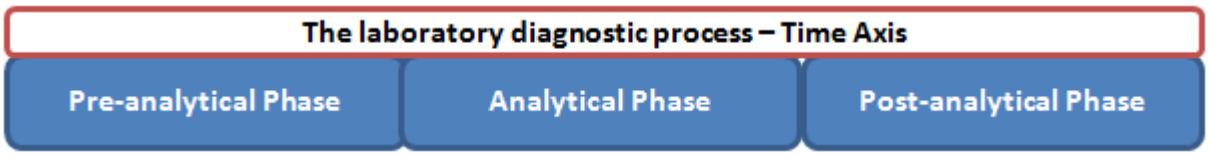
1. Pre-Analytical Effects on Laboratory Examinations

Authors: RNDr. Zdeněk Švagera, Ph.D.; Mgr. Radka Šigutová

Reviewer: RNDr. Jiří Zadina, CSc.

1.1. Introduction

The laboratory diagnostic process to obtain a result can be divided into three phases: the pre-analytical, analytical and post-analytical phases (see the diagram).



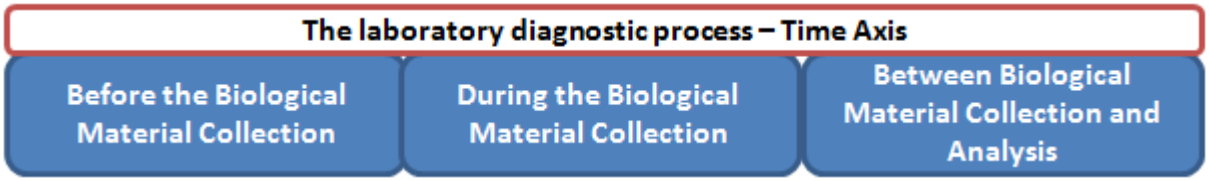
The pre-analytical phase is defined as the period from the physician's indication of the test up to the laboratory analysis of the biological material. In other words, this phase involves an individual's preparation for collection of the biological material, the collection itself, storage of the collected sample and its transport to the laboratory, and preparation of the sample for the assay. The importance of this phase is also supported by many publications mentioning the fact that up to 46 – 68 % of erroneous results are caused by failure to follow or respect the pre-analytical phase rules. That is why the primary task of the laboratory is to provide clients with all necessary instructions (on patient preparation, sample collection, biological material storage and transport, pre-analytical sample treatment) so as to minimize the risk of errors that could consequently cause harm to the patient. All this information is summarized in the manuals of testing laboratories.

The pre-analytical phase is followed by the analytical phase, involving the sample analysis itself. Each laboratory must have an established quality control system to ensure the validity of the issued results. The analytical phase ends with the post-analytical phase, defined as the period from obtaining the lab result to its hand-over to the physician.

It is necessary to keep in mind that biological samples constitute a risk of infection, and therefore personal protective equipment (rubber gloves, protective coat) should be used for work with biological material (material collection, lab work with the sample). In addition, a face mask and safety goggles must be used for highly infectious samples such as HIV or hepatitis C. If clothes or skin is contaminated by the biological material, the affected area should be washed and then disinfected. In the event of injury, the wound must be treated (let it bleed for several minutes, wash with soap, disinfect) and medical attention sought.

1.2. Pre-Analytical Phase and Its Sources of Variability

As mentioned in the introduction, the pre-analytical phase, i.e. before the analysis of the sample (specified parameter) in the laboratory, can be a source of many errors. Therefore, it is necessary to explain what factors affect the pre-analytical phase most. These factors can be divided chronologically:



1.3. Before the Biological Material Collection

Factors affecting the pre-analytical phase before the biological material collection can be further divided into controllable and uncontrollable factors. Controllable factors include, for example, adherence to some daily regimen, dietary habits, etc. Uncontrollable factors are variables such as age, gender, race, etc.

1.3.1. Controllable Factors before the Biological Material Collection

Food is an important controllable factor before the biological material collection. Blood should ALWAYS be collected after a fasting period. Where this is not the case, increased levels of some metabolites can be observed due to ingested nutrient metabolism. Glucose, triacylglycerol, free fatty acid and lipoprotein levels are elevated. People whose diet is rich in fats will primarily have an elevated serum triacylglycerol concentration on the one hand, and a decreased serum nitrogen substance concentration on the other. Protein-rich food leads to increased ammonia and urea levels. At the same time, postprandial hormones (e.g. insulin, which reduces potassium and phosphate levels) are released. Food composition may also affect the pH of urine. For example, vegetable and fruit consumption makes urine more alkaline, while meat, fat and protein-rich food makes it more acidic. Some metabolite levels may also be influenced by the consumption of certain beverages (caffeine increases the glucose level in blood). Alcohol also significantly affects biochemical assays. After alcohol ingestion, the blood lactate concentration increases almost immediately, while hydrogencarbonate and glucose levels go down. Long-term alcohol burden in the body leads to liver damage, which is manifested by increased alanine aminotransferase (ALT), aspartate aminotransferase (AST) and gamma-glutamyltransferase (GGT) levels. Triacylglycerol and cholesterol concentrations are also elevated.

Another factor that may affect the final result is physical strain before the collection. The impact on the result will depend on the type of physical activity: either a short-term activity, with high-intensity anaerobic metabolism of the body, or a long-term (endurance) activity where the body predominantly employs aerobic metabolism. Medium physical exertion increases the glucose level and insulin secretion is stimulated. Muscular activity also increases levels of AST, lactate dehydrogenase (LD) and creatine kinase (CK) enzymes as well lactate and fatty acid levels. Long-term strenuous activity results in a decrease in blood sugar, an increase in creatinine, and multiple-fold increase in lactate levels. Cholesterol and triacylglycerol levels are also reduced.

Another controllable factor before the biological material collection is mechanical trauma; for example, muscle trauma, including intramuscular injections, causes the release of enzymes (CK, ALT, AST) and muscle tissue proteins (e.g. myoglobin). Cycling may cause mechanical trauma to the prostate, which may manifest itself by the release of prostatic serum antigen leading to a false positive result for this test. Marathon running and heart valve defects lead to the mechanical haemolysis of erythrocytes.

A very common problem, which is very difficult to control, is the effect of drugs. Drugs may affect the level of some monitored analytes; for example, acetylsalicylic acid (aspirin) increases serum AST and ALT and urine protein levels, furosemide increases serum glucose, amylase (AMS) and alkaline phosphatase (ALP), and decreases sodium cation levels. Drugs may also interfere with the analytical assay procedure. For example, since vitamin C has strong reduction properties, it causes a false decrease in the level of analytes detected using peroxide. Drugs may also affect the rate of metabolism or monitored analyte elimination, or damage certain organs – the hepatotoxicity of narcotic agents being an example.

Stress is also a major factor. Stress situations cause the release of stress hormones such as renin, aldosterone, somatotropin, catecholamines, cortisol, glucagon and prolactin. This is why blood collection for prolactin assays should be performed within three hours after waking up. Another example might be the 60% drop in cholesterol compared with the initial level within 24 hours after acute myocardial infarction. It takes many weeks before its concentration reverts to normal. For this reason, blood collection for cholesterol, HDL and LDL cholesterol assays is not recommended when patients with suspected acute myocardial infarction are being hospitalized. In contrast, slight stress may increase cholesterol concentration. Post-operative stress decreases the concentration of thyroidal hormones and transferrin, and secondarily increases the concentration of ferritin.

1.3.2. Uncontrollable Factors before the Biological Material Collection

Uncontrollable factors before biological material collection include age, gender, race and biological rhythms. A further uncontrollable factor which might be included here is pregnancy. However, since this example of influence on the pre-analytical phase is too specific, it will not be described in this communication. Except for biological rhythms and pregnancy, these effects do not require any special attention as they are beyond our control and are considered

through reference limits for the relevant analyte.

Age is a very important uncontrollable factor, since most monitored analytes are age related. An older person will have higher cholesterol levels than a younger person. Children and adolescents exhibit higher total alkaline phosphatase activity than adults due to a higher production of the bone isoform of this enzyme as the body grows. The reason is that the assay includes total alkaline phosphatase activity, including the bone isoform. Attention must also be paid to the higher total ALP level in pregnant women due to the higher production of the placental isoform of this enzyme.

Gender also has a major influence on the result of the assay. It is commonly known that many parameters depend on the hormone set and physical constitution. For example, men have higher levels of creatine kinase (CK), ALT, AST, ALP, uric acid, urea, haemoglobin, ferritin, iron and cholesterol than women.

Furthermore, the non-Caucasian population is increasing in the Czech Republic. For example, the CK and AMS activity or the granulocyte count rise in ascending order from Caucasian through Asian to African-American populations (African-Americans have up to twice as much CK activity and Asians have a higher salivary amylase activity and a higher total bilirubin concentration).

Other effects that should be considered are biological rhythms with their different time periods, either occurring within a single day (circadian) or cycles taking roughly a year to complete (circannual). Circadian changes vary for different parameters; for example, there is up to 50 % change in iron levels during the day. Other parameters such as AST, ALT, LD and ALP show changes in the range of tens of percent. Maybe the most notable circadian change occurs in cortisol – about 250 % with minimal levels in the evening. An example of circannual rhythm is the change in vitamin D concentration, with maximum levels in summer months due to skin exposure to intense sunlight.

1.4. During the Biological Material Collection

Factors influencing the pre-analytical phase during the biological material collection are primarily related to the work of the sample-collecting nurse, who has to keep in mind the basic sampling principles that may affect the result of the test. In particular, such principles include collection timing, selecting the appropriate collection set, the patient position during the collection, venostasis and local metabolism effects, as well as the effect of infusion and transfusion in the hospital environment.

1.4.1. Collection Timing

Collection timing plays a very important part in the strategy to obtain valid results. Most often, collections take place in the morning when we can be sure that the patient has fasted (provided the patient respects general pre-collection recommendations) and the circadian rhythm effect mentioned in the chapter above is limited. A different example is blood sugar monitoring (blood sugar profile) or pharmacotherapy monitoring, where samples are taken based on drug elimination half-life.

1.4.2. Patient Position during the Collection

Patient position during the collection is also important. It must be kept in mind that the difference in protein concentration when comparing a standing vs. sitting position for 15 minutes is 5 – 8 %, and about 10 – 20 % for a standing vs. recumbent position. In the standing position, water transfers from the intravasal to the interstitial space, which subsequently leads to a rise in high-molecular substances, primarily proteins, lipoproteins and protein-bound substances such as calcium cation and hormones (cortisol, thyroxin), or some drugs. In general, biological material should always be collected in the same position, preferably the standard sitting position, which is not always possible in hospitalized patients, though.

1.4.3. Use of Tourniquet and Local Metabolism Effect

The effect of local metabolism when a tourniquet is used for collection is also interesting. The evidence shows that one minute after constricting the arm with a tourniquet there is already a significant transfer of water and ions from the vessel to the interstitium, with a subsequent rise in protein and blood protein-bound substance concentration. Long-term constriction or overcooling of the arm leads to a change in local metabolism due to hypoxia, which results in a rise in partial carbon dioxide pressure and potassium and lactate concentration, which in turn results in a drop in pH. In addition, there are homeostasis changes connected with the release of the tissue factor. Exercising the arm is

not recommended, or it is even forbidden, as it primarily causes an increase in potassium concentration. For these reasons, the period for which the arm is constricted should not exceed one minute, and the tourniquet should be released immediately after the venipuncture.

1.4.4. Choosing the Collection System and the Effect of Anticoagulants

The choice of the collection system is also very important. Options include a closed or an open sample collection system. The open collection system consists of a classical needle and a Luer-taper syringe. Following venipuncture, freely flowing blood is taken directly into the test tube or by gently pulling the plunger. Collection into a closed system is the preferred option today as it minimizes the risk of contaminating the collecting person through the blood, and collection tubes are colour coded depending on the added preservative or anticoagulant. Another advantage of the closed system is that the ratio of anticoagulant (preservative) to collected blood is maintained.

As mentioned above, anticoagulant (heparinate, citrate, oxalate, etc.) can be chosen depending on the required test. Nevertheless, attention needs to be paid when choosing the anticoagulant for cation tests, since the anticoagulant must not contain the cation being determined. For example, the use of EDTA with potassium will lead to highly pathological potassium concentrations in the sample! EDTA is not suitable for determining bivalent cation concentration as it acts as a chelating agent, binding these cations to form a complex, and it results in finding a falsely low concentration of these ions. In some cases, another substance (preservative) such as sodium fluoride is added to the anticoagulant in order to determine glucose concentration. The addition of sodium fluoride will cause glycolysis inhibition in red blood cells, thus preventing a drop in glucose concentration over time.

In addition, we must keep in mind that if a collection set containing an anticoagulant is used, we should gently mix the collected blood immediately after the collection. Without mixing, the anticoagulant effect is limited and undesired blood clotting will occur. A suitable needle lumen should be selected for blood collection to avoid red blood cell haemolysis.

1.4.5. Effect of Infusion and Transfusion

Patients in a critical condition have to receive transfusion and infusion products containing high concentrations of selected substances and low concentrations of others. Infusion may therefore affect the determination of some substances, usually by direct contamination during collection or just due to their properties. For example, the infusion of glucose with potassium results in a false increase in glucose and potassium levels. The infusion of lipid emulsion causes serum chylosis and Hartmann infusions containing high lactate concentration (>15 mmol/l) cause a false increase in lactate concentration. On the other hand, Plasmalyte infusion causes a false normalisation of ion concentration in the collected sample. This is why certain rules should be followed during the sample collection following an infusion. Ideally, collect blood from the other arm, i.e. where the infusion was not applied, or stop the infusion for 15 minutes and then take the sample.

With respect to the pre-analytical phase, the age of transfusion must be taken into account. With the growing age of the erythrocyte concentrate, sodium and glucose concentrations decrease due to red erythrocyte metabolism, whereas, in contrast, potassium and lactate concentrations increase.

1.5. Between Biological Material Collection and Analysis

This period includes the time from the collection of biological material until its analysis in the laboratory, and involves handling the sample following the collection, its subsequent transport to the laboratory, and centrifuging or pre-treatment before the analysis.

In general, if anticoagulated blood is taken (collection container with anticoagulant), the test tube should gently be shaken immediately after the collection. If non-anticoagulated blood is taken, wait about 30 minutes before transporting the sample to allow sample clotting (exact time required for clotting is indicated by the manufacturer of the collection set). Immediate transport of the biological material after the collection may cause haemolysis and sample deterioration. The problem of haemolysis interfering with the assay is not only related to the release of erythrocyte content into the serum or plasma with a subsequent increase in the concentration of these substances in the tested material, but also to the release of haemoglobin, whose colour interferes directly with a photometric assay or with the agent used for the assay. Take care – haemolysis may also occur due to sample overcooling, high centrifuge speed or a narrow sampling needle. The following table describes the effect of haemolysis on selected biochemical assays.

Increase in concentration or activity	Potassium, magnesium cation, lactate dehydrogenase, aspartate aminotransferase, alanine aminotransferase, creatine kinase, acid phosphatase
Decrease in concentration or activity	Gamma-glutamyltransferase, alkaline phosphatase, amylase

Storage of the sample before the transport and the very transport of the biological material are very important and must be given adequate attention, especially if samples are transported from practitioners in the periphery and brought to a specialized laboratory. The transport time will vary; however, always avoid exposing the sample to extreme conditions (heat/freezing) during the transport, minimize shaking the sample and avoid complete deterioration which will occur if the sample is spilled. This is why samples have to be transported in temperature-controlled transport boxes protected against spillage. Some samples (tissues) must be transported frozen even at very low temperatures (-80°C) on dry ice. If the maximum time before sample processing is exceeded or transport conditions are not adhered to, some substance concentrations in the material for testing will change. One example is a decrease in glucose concentration or an increase in lactate concentration due to the anaerobic glycolysis of blood elements. Some analytes in biological material are thermolabile at room temperature (most parameters) and some, paradoxically, at 4°C (e.g. ALT activity decreases or potassium concentration increases due to the ATPase inhibition in the erythrocyte). Some analytes are photo-sensitive (e.g. bilirubin and porphyrins), and their amount drops unless transported and stored in the dark. For these reasons, some analytes have specific recommendations for storage and transport. For example, the recommendations for a plasma ammonia assay are as follows: carry out the anaerobic collection, prevent haemolysis, maintain the anticoagulant to blood ratio and transport in a transport container or on melting ice; analyse within 20 minutes after the collection.

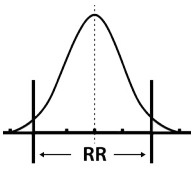
As soon as the samples are delivered to and received by the laboratory, they are either analysed directly (when whole blood is used), or must be centrifuged to obtain serum or plasma. The required conditions must be adhered to during centrifuging to achieve perfect serum (plasma) separation from erythrocytes and perfect leukocyte sedimentation in the plasma. If the speed (relative centrifugal force) is too high during centrifuging, the cells may break and their content may get released. Many analytes require centrifuging at lower ambient temperatures (cooled centrifuges), for hormone assays, for example.

Urine analysis requires a chapter to itself, since it requires the use of collected, first morning or single random specimens. Very often, patients are not instructed about the collection rules; they typically collect urine for a longer or a shorter time than required; moreover, obtaining an exact reading of the quantity of urine collected over the collection period, usually 24 hours, is always problematic. Nor it is possible to ensure the required storage of the collected urine in the fridge or the urine pre-treatment needed to stabilize the tested parameter. First morning urine collection poses a similar problem, since it has to be delivered for sediment analysis within one hour of collection. There is often a delay in delivering the collected urine to the laboratory, which leads to false negative or false positive results (increase in the bacteria count, increase in the pH value due to the urease of bacteria and cell element degradation).

In general, the transport and storage conditions required for transported samples/material must be followed. Material transport in extreme (very hot, very cold) conditions requires special care.

Ten pieces of advice for obtaining correct results

- Instruct the patient (why they are being tested, diet, physical strain)
- Time the collection correctly
- Fill in the order slip correctly
- Choose the right collection procedure
- Choose the right test tube
- Take the recommended amount of material
- Do not spill any biological material
- Label the test tubes correctly
- Ensure appropriate storage for biological material before transport
- Ensure appropriate transport to the lab



CHAPTER 2

2. Reference Values and Methods for Their Determination

Authors: Ing. Vladimír Bartoš, Ph.D.; RNDr. Pavlína Kušnierová, Ph.D.; Ing. František Všianský

Reviewer: prof. MUDr. Richard Průša, CSc.

2.1. Introduction

Laboratory test results are indisputably a very important source of information when choosing the right treatment for the patient. Among other things, they help determine or specify the diagnosis, select, optimize and monitor the therapeutic procedure, or determine the prognosis of the patient's condition.

When the results are interpreted, in most cases, the extent to which a result is consistent with values that might be reasonably expected in the selected reference population is considered. In other words, a specific result is compared with the limits defining the interval of result values obtained in the same laboratory test of a sample of the reference population, i.e. with the reference interval.

2.2. Basic Terms and Definitions

Reference interval (reference range): A generally accepted definition of this term is: the determination of limits between which 95 % of reference values fall (an interval encompassing up to 99 % is used for certain parameters).

This is usually the interval between the lower and the upper reference limits, which are typically the 2.5^{th} and 97.5^{th} percentile of a set of values obtained through the analysis of a sufficiently homogeneous and large sample of the defined reference population.

Sometimes, from a clinical point of view, only the upper or the lower reference limit may be important, which corresponds to the 95^{th} percentile, or the 5^{th} percentile, respectively.

Examples of some specific variables and their assignment to the reference interval type are shown below (Table 1).

Reference values: These are the values (results of measuring the relevant variable) obtained from a selected group of individuals with a defined health condition.

Reference value distribution: Distribution of individual measurement results corresponding to some of the statistical distributions. The relevant type of distribution (normal, also called Gaussian, log-normal, Laplace distribution, etc.) is tested using suitable statistical methods.

Reference population: The set of all individuals meeting certain criteria concerning their health status or other defined requirements (age, gender, race). Absence of a certain disease is usually required; it is clear, however, that this defined notion of „health" is very relative and in its way imperfect.

Reference individual: An individual selected from the set of reference population.

Reference population selection (reference sampling group): A randomly selected part of the reference population from whom reference sampling values are obtained by measurement in order to estimate the reference limits. It is unrealistic to make measurements on the entire reference population, and therefore only a randomly selected sample is measured. Sampling characteristics obtained by measurement of this sample are thereby a more or less plausible estimate of actual values.

Likelihood estimation: Agreement between the estimate and the actual value typical of the entire population. This depends on many factors, the primary one being the size of the reference population sample. At least 120 reference

individuals are typically required as a minimum; the confidence interval of estimates contracts considerably with the size of the reference sampling group. The likelihood estimation is further affected by pre-analytical aspects (method of collection, transport and storage of samples), the measurement method and the statistical evaluation method used.

Reference Interval Type	Measured Variable (Parameter)
LRL – URL	Na^+, K^+, Ca^{2+}, glucose, transferrin, thyroid hormones
below URL	bilirubin, AST, ALT, CK, troponin I, tumour markers
above LRL	cholinesterase, prealbumin

Table 2.1. Examples of reference interval types relative to the clinical importance of reference limits (LRL – Lower Reference Limit, URL – Upper Reference Limit)

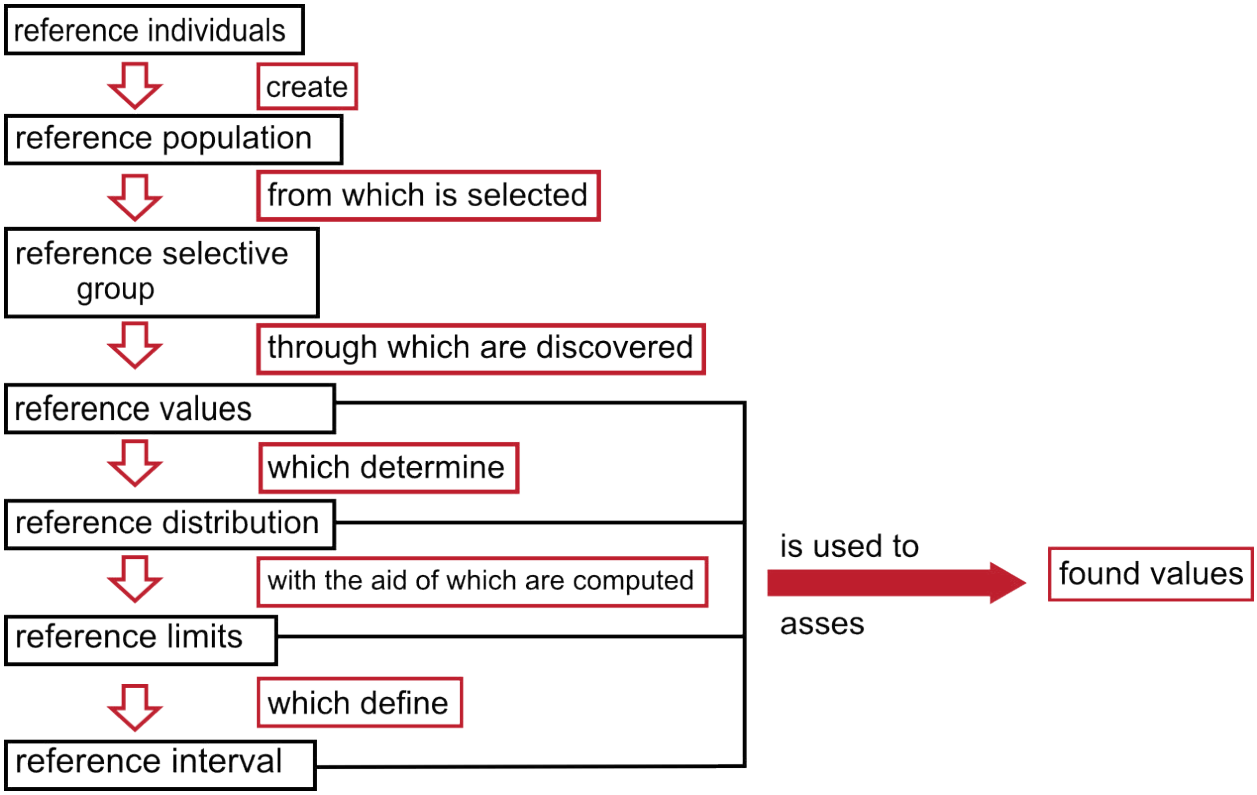


Figure 2.1. Relationships between terms in connection with determining the reference interval

2.3. Options for Determining Reference Intervals

One of two methods is usually chosen to determine reference intervals from actual measurement results.

The first method, referred to as the direct (inductive) method of estimating reference limits, is applied when the possibility exists of obtaining a sufficiently large reference sampling population on which the studied variable is measured, providing the relevant pre-analytical conditions are adhered to. Selecting the requirements for reference individuals is quite a complicated task which may substantially affect the resulting reference limit values. In the NORIP study (Malmø, 27/4-2004), some of these requirements were proposed together with the pre-analytical conditions that should be followed when determining reference intervals. The conditions that reference individuals should meet include:

- Be feeling subjectively well;
- Have reached the age of 18;
- Not be pregnant or breast-feeding;
- Not have been hospitalized or seriously ill during the last few months;

- Not have had more than 24 g of alcohol in the last 24 hours;
- Not have given blood as a donor in the last 5 months;
- Not have taken prescribed drugs during the last 2 weeks;
- Not have smoked during the last few hours prior to sampling.

The second method, also referred to as the indirect (deductive) method, is applied if no suitable reference sampling population is available, and only the results of the variable obtained by measurement of a „mixed" population comprising healthy and ill individuals can be used.

2.3.1. Direct Method of Reference Interval Estimation

This method is based on the statistical processing of measured reference values. Data obtained by measurement can be evaluated either parametrically or non-parametrically, depending on whether parameters characterizing the reference value distribution are used to establish the reference limits, which are then used to derive the numerical values of corresponding quantiles determining the reference limits. Although the non-parametric procedure is more universal, in this case the quantile likelihood estimates are lower.

2.3.1.1. Parametric Procedure

Use of the parametric procedure is only legitimate in selections coming from a normal (Gaussian) distribution, or from distributions that can be transformed into a normal distribution. Only in these cases is it possible to use the parameters of this distribution, i.e. the sample mean and sample standard deviation, as the best estimates of position and variance characteristics of this distribution, and to derive from them the relevant percentile values.

An important step in selecting the reference limit method is to confirm the assumption of normality for the distribution of obtained reference values, with the primary requirement being the symmetry of distribution. Various statistical tests such as Kolgomorov-Smirnov, Anderson-Darling or D'Agostino-Pearson tests are used to evaluate this normality. However, it should be kept in mind that different statistical tests can have different predicative abilities, and so are more likely to inform us that there is something wrong with the anticipated normality of the distribution. These tests are often used in combination with graphical methods; this has the advantage of enabling a comparison between the distribution of measured reference values (usually shown in a histogram) and a normal distribution with the same parameters using a frequency function diagram (probability density). Common statistical programs also enable the plotting of diagrams called rankit-plots, which are used to compare sampling distribution quantiles with normal distribution quantiles. If the sampling distribution matches the normal distribution, the dependence is linear.

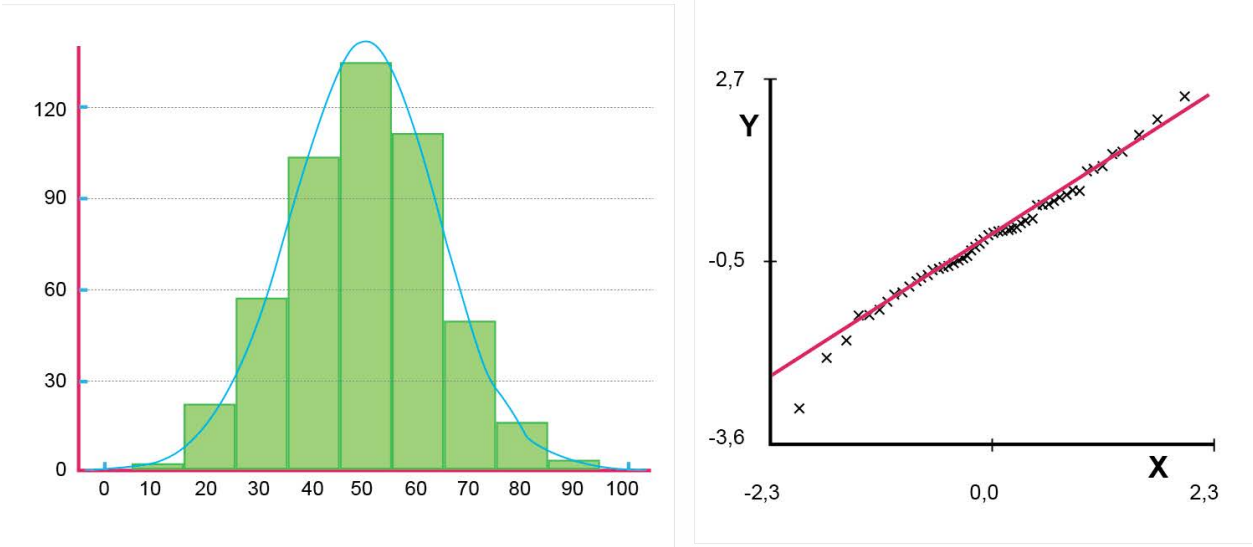


Figure 2.2. The histogram of distribution of the reference values obtained (x-axis: measurand value, y-axis: frequency of occurrence) and its approximation using the frequency function with the same parameters, and the rankit plot comparing empirical distribution quantiles (y-axis) and normal distribution quantiles (x-axis)

So that the parameter estimate is unbiased, a test must be conducted to ensure that the reference values obtained by measurement do not contain outlying results or gross errors. Some common tests for outliers (Grubb's test, Dean-Dixon test, etc.) or some graphical techniques can be used for this purpose. If the data contain outliers or gross errors,

the next step must be their exclusion.

The normal distribution is clearly characterized by its parameters, the mean μ and the variance σ^2. Only estimates of these parameters can be determined from the sampling dataset, which are the sample mean arithmetic mean and the sample variance s^2. Both parameters can be computed from the following equations:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

The sample standard deviation s is then computed by a simple root extraction of the variation value. Knowing these parameters, the 2.5th and the 97.5th percentile can be easily estimated. Assuming normal distribution, these quantiles can be computed as

$$x_{(2.5)} = \bar{x} - 1.96\ s$$
$$x_{(97.5)} = \bar{x} + 1.96\ s$$

As a general rule for normal distribution, the total area under the density curve equals 1, and the probability that a random variable acquires values from a certain interval is equal to the area defined under the curve of density above this interval.

For example, for an interval with limits < μ – 1.96 σ; μ +1.96 σ >, the size of this area is exactly 0.95.

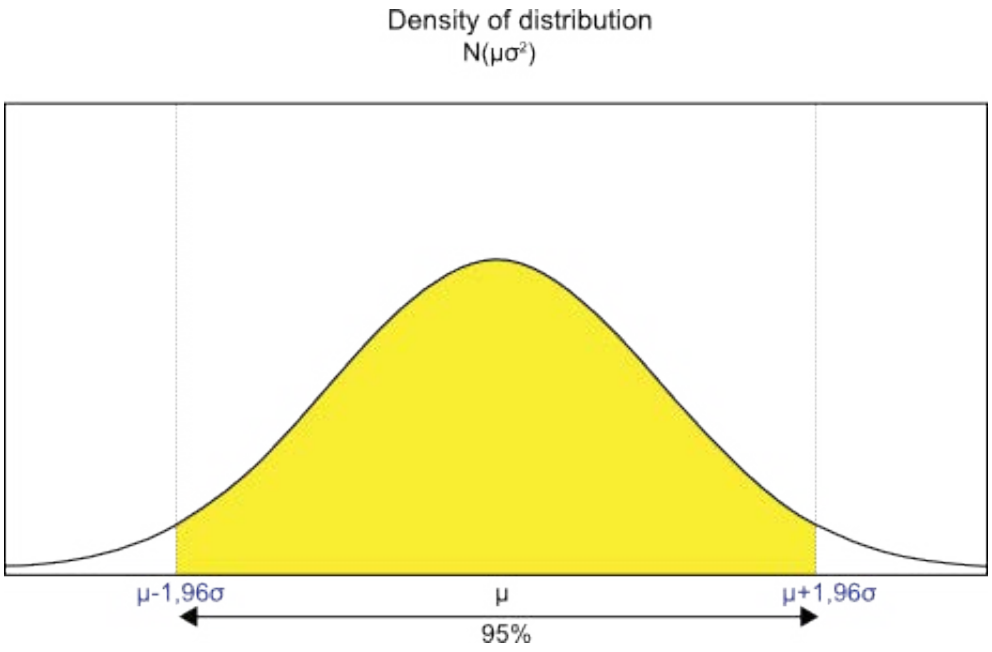


Figure 2.3. Diagram of the probability density of the random variable X with normal distribution N(μ,σ2), representing the confidence level of the variable occurrence within the intervals of μ -1.96σ to μ+1.96σ

In practice, the coefficient 1.96 is often rounded to 2, so the reference interval limits are then determined more easily as arithmetic mean ± 2 s.

Greater coefficients can be used to determine a reference interval with a coverage of probability greater than 95 %. For example, the coefficient 2.57 corresponds to a probability of 99 %.

In practice, only a limited set of biologically significant variables meets the condition of normal distribution. This type of distribution can be expected only in analytes with a relatively narrow biological distribution, for example during

serum ion concentration measurement (Na^+, K^+, Cl^-, Ca^{2+}, etc.).

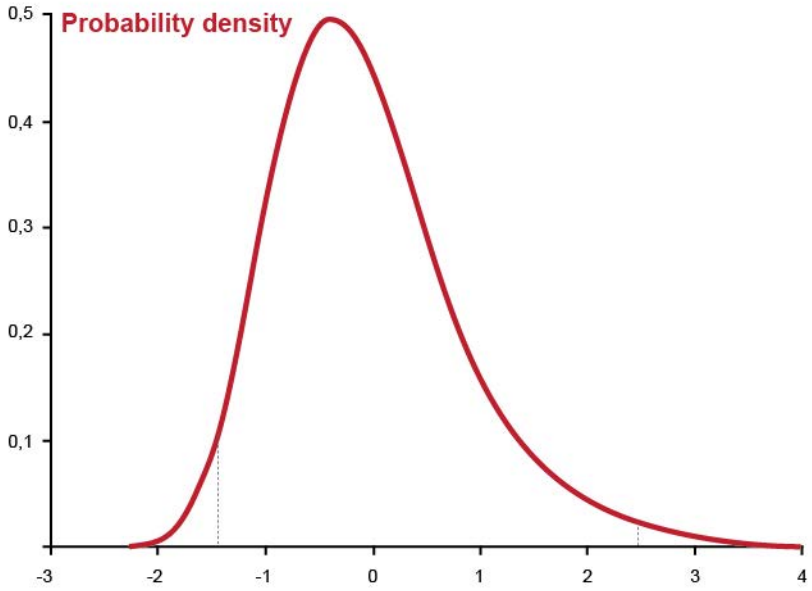


Figure 2.4. Diagram of probability density of the random variable X with asymmetric distribution skewed to the right ("right-tailed" distribution)

However, in practice, we generally encounter variables with an asymmetric distribution of values, i.e. distribution skewed towards greater values. Examples of such variables are concentrations of glucose, creatinine, urea, AST, ALT, CK enzymes, thyroid-stimulating hormone, etc.

This skewed asymmetric distribution is often modelled using log-normal distribution, which can be easily transformed into normal distribution using logarithmic transformation by simple logarithmic calculation of the measured values. If this transformation is insufficiently effective, another transformation such as a power transformation or Box-Cox transformation can be used.

To determine the reference interval, a similar procedure to selection from the normal distribution is used: the normality of transformed data is verified, sample parameters (transformed mean and standard deviation) are estimated and then used to calculate the transformed reference limits LRL_T and URL_T

$$LRL_T = \bar{x}_T - 1.96 \cdot s_T$$
$$URL_T = \bar{x}_T + 1.96 \cdot s_T$$

The inversion function used for the transformation is then used to determine the reference range limits. For example, if logarithmic transformation is used, the computed limit values are exponentiated (exp is inverse to the log function).

$$LRL = \exp(LRL_T)$$
$$URL = \exp(URL_T)$$

2.3.1.2. Non-Parametric Procedure

The non-parametric procedure is primarily used where a sufficiently large reference sampling group is available. In addition, this method of determinig the reference interval limit is more general, without any requirements for data distribution.

The below procedure is followed when determining reference limits:

- Measured data are sorted into ascending order by size;
- Each value is assigned a sequence number from 1 to N with the minimum value having 1 and the maximum value having N;

- The elements in the sorted set are assigned sequence numbers determining the 2.5th and 97.5th percentile - n(2.5) and n(97.5);
- The LRL is assigned the value of the element corresponding to the sequence number of the 2.5th percentile, and the URL is assigned the value of the element corresponding to the sequence number of the 97.5th percentile (if either of n(2.5) or n(97.5) is not an integer, the reference limit value is obtained by linear interpolation between the values of elements with sequence numbers bilaterally closest to this number);
- The median of selected reference values is considered the reference interval mean.
- The following steps can be taken to determine the sequence numbers of corresponding percentiles (IFCC recommendation):

$$n_{(2.5)} = 0.025 \cdot (N+1)$$
$$n_{(97.5)} = 0.975 \cdot (N+1)$$

or, alternatively, in accordance with the steps recommended by Linnet (Linnet K., Clin Chem 2000,46, 867-9):

$$n_{(2.5)} = 0.025 \cdot N + 0.5$$
$$n_{(97.5)} = 0.975 \cdot N + 0.5$$

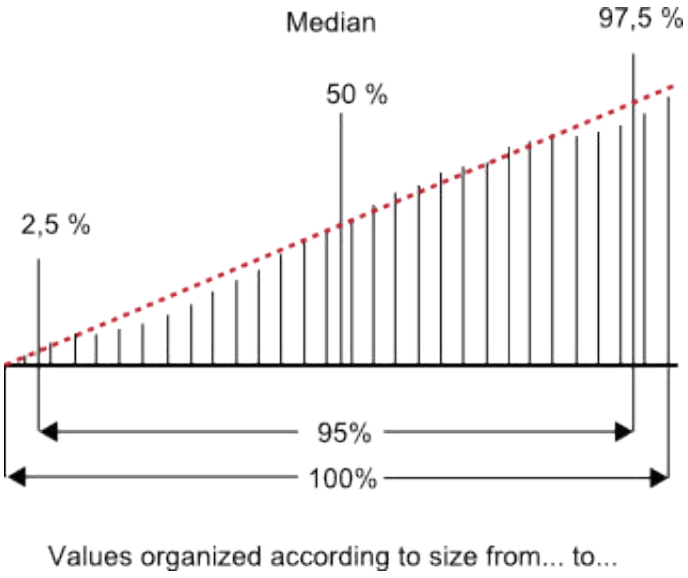


Figure 2.5. Ascending sorting of selected reference values by size

2.3.1.3. *Efficiency of Reference Limit Estimates*

Since we only work with the reference sampling population, whatever the method we use, only reference limit estimates and not their actual values are obtained. The measure of likelihood of these estimates can be, for example, the size of their 90% confidence interval. This is closely related to the range of the reference sampling population, i.e. the number of reference individuals included in the selection.

If a non-parametric procedure is used, confidence intervals are generally wider than those in the parametric procedure. To reach a comparable confidence level for both procedures, the sampling range in the non-parametric procedure would have to be about twice the size as that for the parametric procedure. It is important to note that it is indeed possible, nonetheless, to improve the efficiency of a non-parametric procedure, i.e. to reach the same quality with a sampling range of about 5-15 % lower. This procedure is referred to as the “bootstrap” principle. This involves repeating the non-parametric procedure, always on a subset of reference values (range n) randomly selected from the entire sampling population with the original range N. For example, the number of repetitions might be k=100. Each repetition yields a pair of specific limits – the lower LRL_i and the upper URL_i. The average of all partial lower limits is considered the final lower reference limit estimate, and the average of partial upper limits is considered the upper reference limit estimate.

2.3.2. Indirect Method of Reference Interval Estimation

This method of estimating reference interval limits can be used if there is no real possibility of obtaining a sufficiently large and representative reference sampling population, but results of variable measurement obtained in normal, routine laboratory operation are available. This method requires a relatively high number of data, in the order of thousands. On the one hand, data are obtained from a "mixed" population containing both healthy and ill individuals without the possibility of meeting pre-defined sampling requirements. On the other hand, this is a population of actually examined individuals, which, if large enough, enables the carrying out of their stratification by age or gender, for example.

Real result processing is based on the assumption that in a sufficiently large sample, the proportion of healthy to ill individuals is substantially higher, and that if an appropriate result evaluation method is used, the reference interval limit estimates obtained will correspond to a healthy population. However, reference intervals obtained using this method will be wider than those obtained using the direct method, with a possible shift towards values typical of an ill population. In addition, this method prevents any objective estimation of the probability density distribution type in the healthy population.

Brief overview of the method (Baadenhuijsen):

- Histogram plotting (about 50 classes, at least 1200 values);
- Data smoothing using the Golay-Savicky filter;
- Plotting the derivation of measured value frequency to concentration logarithms;
- Calculation of intercept (a) and slope (b) of the straight line plotted in the linear part of the dependence;
- $\mu = a/b$; $\sigma = -1/b$;
- Reference limits $RL = \mu \pm 1.96\ \sigma$.

2.4. Importance of Reference Interval when Interpreting Results

The reference interval is used in practice in interpreting measurement results by defining the limits of a "normal" finding, i.e. the range within which the result of a healthy individual is supposed to lie. To provide more descriptive information, laboratory result reports often supplement the numerical result with a graphical representation of the result position against reference limits. This is usually only a schematic illustration of whether the measurement result lies within or outside the reference interval (example: the use of * = result; within or outside brackets = reference interval). However, should a physician automatically interpret a measurement result lying outside the reference range as a pathological finding, regardless of how far the result is from the reference limit, it would be a grave mistake.

It should be borne in mind when evaluating laboratory test results against the reference interval that the reference interval is plotted using estimates of its limits. This means that these limits are not points but intervals, and so for each limit there is a confidence interval to express the area where the actual limit is located with a certain confidence level.

In the first place, be cautious when interpreting results which occur near reference limits, regardless of whether they are within or outside the reference interval. In addition, measurement result uncertainty, an attribute inherent in every analytical method, plays a role in such borderline situations. Due to such uncertainty, even a measured result has to be understood not as a point corresponding to its value but also as an interval where the measurement result occurs with 95% probability.

It follows from the above that there can be a non-zero probability in these cases that the real measurement result can be on the opposite side of the actual reference limit.

In addition, it must be realized that interpreting laboratory results using the 95% reference value interval also includes the following proviso: even in a healthy individual there is a 5% probability that their result will be lower (2.5% probability) or higher (2.5% probability) than the lower or upper reference limit.

CHAPTER 3

3. Analytical Properties of the Laboratory Method, Quality Control

Authors: Ing. Vladimír Bartoš, Ph.D.; Ing. František Všianský; RNDr. Pavlína Kušnierová, Ph.D.

Reviewer: Ing. Václav Senft

3.1. Introduction

Inherent in every measurement, and therefore in every method used in biochemistry laboratories, is a typical set of properties, generally referred to as the performance characteristics of the method. Their level indicates the options for measurement made using this method, which is why they are a determining factor in the usability of the method for the required application.

The level of analytical properties of the method is also a determining factor in the use of the method for clinical purposes. Therefore, every physician who uses measurement results should be aware of the basic analytical properties of the method used.

The process, the aim of which is, besides determining the functional characteristics of the method, to comprehensively evaluate the suitability of the method for the intended clinical purpose, is called method validation. Realization of this process is an integral part of method development. Basic analytical properties of the method are also checked before the method is first used in the laboratory, and they are likewise regularly checked during routine use of the method. This process is known as verification of the method.

The set of operations carried out in the laboratory aimed at ensuring the adequate likelihood of measurement results is wider in scope. These activities are generally referred to as **quality control**.

This set of activities is primarily intended to assure the quality of the analytical process in the clinical laboratory. In line with general trends, even medical laboratories implement comprehensive quality management systems to manage all laboratory operations; their aim is not only to maintain but also gradually improve the quality of laboratory services provided. The implementation of these mechanisms in laboratory management is inspected and certified by independent bodies in certification or accreditation processes, and is also increasingly required by healthcare payers.

3.2. Performance Characteristics of the Analytical Method

The basic analytical properties of the method undoubtedly include two terms referred to as **precision** (in Czech: preciznost) and **trueness** (in Czech: pravdivost). Their combined projection in a specific measurement result then constitutes the property of the result referred to as **accuracy** (in Czech: přesnost).

3.2.1. Precision

Precision is the closeness of agreement between independent measurement results obtained under pre-specified conditions.

Measurement precision is expressed numerically by measures of imprecision, which define dispersion between independently obtained results, such as standard deviation, variance, or coefficient of variation.

Precision is the evaluation of the impact of *random errors* in measurement that can never be eliminated and whose magnitude is inherent in a certain method or specific measurement procedure. These errors are caused by accidental effects (instability of instruments, fluctuation of ambient measurement conditions such as temperature, variations in the operator's actions, etc.). Their action results in differences between repeated measurement results. If there is a sufficient number of repetitions, they are uniformly dispersed around their average value due to the randomness of their

origin. Minimum deviations are most frequent and their number decreases with the increasing value of deviations. The distribution of deviation frequency corresponds to the normal (Gaussian) distribution. The measure of dispersion, i.e. the imprecision of results, is the standard deviation *s*. Since it is expressed in units of the measured quantity (measurand) and depends on its magnitude, use of a relative expression of standard deviation, i.e. the coefficient of variation *CV* expressed in % (sometimes also referred to as the relative standard deviation), is preferred.

The precision of the method is not the same throughout the method's working range. The dependence of precision on the measurand magnitude is called *precision profile*, and is an important scale of quality of the measurement method. The *CV* reaches lowest values in the middle area of the measurement range, whereas values grow towards the ends (growth is significant especially in the area of very low values of the measurand). The magnitude of imprecision is affected by the actual working range of the method. One condition for its selection might be that *CV* should not exceed the required level, 10 % for example - see Figure 6.

The specified measurement conditions in the precision definition could be conditions of repeatability or reproducibility.

The condition of **measurement repeatability** covers the same measurement procedure, operating staff, measuring system, working conditions, the same site and repeating the measurement on the same or similar object over a short period of time.

The condition of **measurement reproducibility** covers different sites, operating staff, measuring systems, and repeating the measurement on the same or similar object. The relatively large freedom in the setting of these conditions always requires detailed specification as to which factors were variable.

Specific conditions of reproducibility involving measurements made using the same procedure, on the same site, and by repeating the measurements on the same or similar objects, but over a more extended period of time, are referred to as conditions of **intermediate measurement precision**. Measurement over a longer period of time may involve the effect of other variable factors such as changes in calibration or calibrators, use of a different lot of reagents or a change of operators. Precision determined under these conditions is the best measure of the quality of method execution in specific laboratory conditions.

3.2.2. Trueness

Trueness is the closeness of agreement between the average value obtained from a large series of measurement results and either the actual value or an accepted reference value x_0. The measure of method trueness is its bias b:

$$b = \bar{x} - x_0$$

or, in the relative expression:

$$b = \frac{\bar{x} - x_0}{x_0} \cdot 100\ (\%)$$

The actual (true) value of the measured quantity (measurand) is in practice inaccessible on principle, and could only be obtained by perfect measurement. Therefore, it is replaced by an accepted reference value as the best practical approximation of the actual value of the quantity. The reference value is usually obtained based on a quantity measurement using a generally accepted reference method or other generally recognized process (by measurement in selected reference laboratories, etc.).

Method trueness is determined by the existence of a systematic error incidental to measurement. This kind of error may affect the measurement result either in a constant way (results are shifted always by the same value), proportionally (always by the same multiple), or in a combination of these two ways. In this connection, we speak of the constant and the proportional components of systematic error.

While random errors cannot be avoided during measurement (only their magnitude can be influenced), systematic measurement errors can sometimes be eliminated or at least partially corrected by appropriate adjustment.

3.2.3. Accuracy

Accuracy is the closeness of agreement between the result of a measurement and the true value of a measurand. This property applies to one measurement result and is actually the current expression of the combination of the precision and trueness of a method. It is the contribution of the Random Error (RE) and Systematic Error (SE) that occurred at the moment of a specific measurement. The sum of these contributions is referred to as the Total Error (TE) of measurement.

As mentioned above, the measure of the contribution of the random error component in a given method is expressed by an estimate of standard deviation *s*; the measure of the contribution of the systematic error component is the deviation *b*. By using these two parameters, the method can be characterized by estimating the total analytical error TE_A occurring in the measurement:

$$TE_A = k \cdot s + b$$

The estimate confidence level is expressed by the coefficient *k* that is equal to the corresponding quantile of the selected one-sided confidence interval (1.65 for 95 % and 2.33 for 99 %).

Besides the two aforementioned characteristics, the clinical applicability of the method is also affected by the natural diversity of the observed parameter in the normal population, called the biological variability of the parameter (tested analyte). With respect to the origin of the contribution to overall biological variability, interindividual (among different individuals of the population) and intraindividual (within the same biological individual) variability is distinguished, and if coefficients of variation are used for their relative expression, they are usually denoted as CV_G or CV_I, respectively. The specific values of both contributions for many significant biological parameters have been monitored and published in the technical literature.

In terms of assessing a method's clinical applicability, it is desirable that the relative analytical precision CV_A of the method should be optimally better than a half of intraindividual biological variability, i.e.

$$CV_A \leq 0{,}5 \cdot CV_I$$

It is required for the acceptable relative analytical trueness B_A of the method that it should be better than a quarter of the total biological variability, i.e.

$$B_A \leq 0{,}25 \cdot \sqrt{CV_I^2 + CV_G^2}$$

Thus, for the total analytical error TE_A of the method to be acceptable there follows the requirement that:

$$TE_A \leq 1{,}65 \cdot (0{,}5 \cdot CV_I) + 0{,}25 \cdot \sqrt{CV_I^2 + CV_G^2}$$

The values of the analytical precision of the method and the intraindividual variability of the analyte are used to compute the Critical Difference *CD* (sometimes also referred to as the Least Significant Change LSC) between two consecutive patient results. This is the difference between two measurement results that can be, depending on the aforementioned characteristics, indicated as significant at the selected confidence level with certain probability, usually 95 %. It is a parameter that indisputably plays an important role in the clinician's decision-making process when a laboratory result changes over time.

$$CD = 1.96 \cdot \sqrt{2 \cdot (CV_I^2 + CV_A^2)} = 2.77 \cdot \sqrt{CV_I^2 + CV_A^2}$$

3.2.4. Relationship between the Precision and Trueness of a Method

As mentioned above, the mutual interaction of the two characteristics over a certain period of measurement results in a specific accuracy level of the measurement result (see Fig 1).

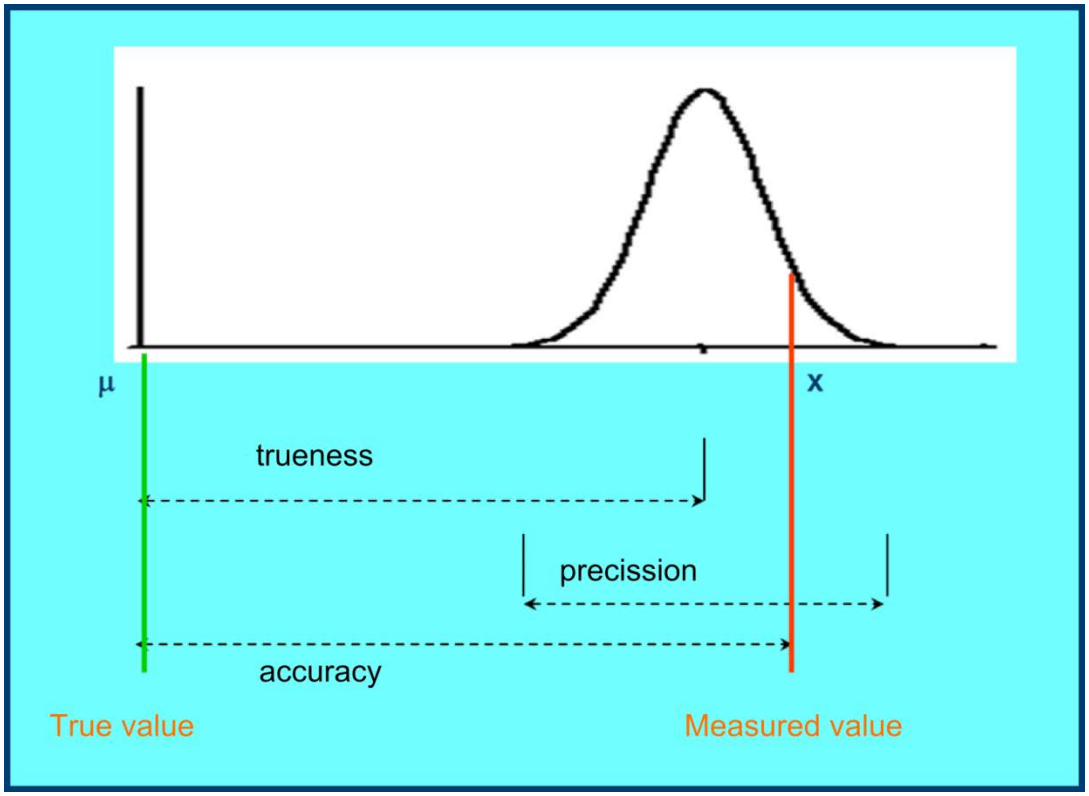


Figure 3.1. The relationship between the terms of precision, trueness and accuracy of measurement

The overall quality of a method can be evaluated in relation to the levels of both characteristics. This relationship is illustrated using the example of target shooting shown in Figure 2. As the quality of the method improves, so another property of the method namely uncertainty (see below), also improves.

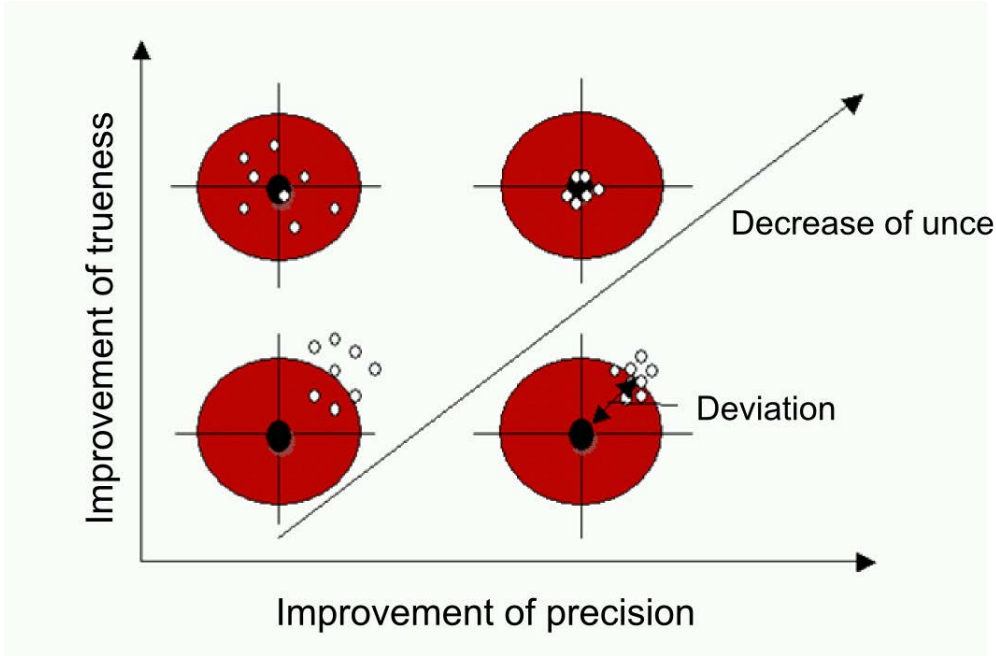


Figure 3.2. Relationship between precision and trueness of the method and their effect on the uncertainty of the method

3.2.5. Measurement Result Uncertainty

Uncertainty is a parameter associated with the result of measurement which characterizes the measure of dispersion of values that could reasonably be attributed to the measurand.

The concept of uncertainty has replaced the previous concept of error in contemporary metrology. Unlike the

previous concept, this latest better describes the fact that a measurement result is just an estimate of reality with a degree of uncertainty. This means that it is not a point estimate differing from reality by a specified error; rather that the measurement result lies with certain confidence within a bounded interval of possible values. Knowing the uncertainty allows measurement results to be better compared between each other or with reference intervals, for example. An estimate of the measurement result uncertainty obtained using a method is also part of its validation, where it is used to assess whether the method is adequate enough for the required purpose.

The principle of determining the uncertainty consists in evaluating effects that can affect the measurement result, and subsequently in estimating an interval for which we can state with a specified measure of confidence that it contains the actual measurand value. The total uncertainty is the result of the composite action of many sources. The effect of each one is their individual contributions to the resulting uncertainty. This contribution is referred to as a component of uncertainty. The quantification of the contribution of certain components can be obtained from statistical distributions of measurement series results characterized by their experimental standard deviation (A-type uncertainty components). The quantification of the contribution of other components is obtained from probability functions based on experience or other information (B-type uncertainty components). The numerical value of each uncertainty component transformed into the standard deviation is called **standard uncertainty** and denoted as u_x (the index x expresses its relation to the component x). Insignificant standard uncertainty contributions are ignored, and the rest are aggregated under the propagation of uncertainty rule into the **combined uncertainty** u_C:

$$u_C = \sqrt{u_1^2 + u_2^2 + \ldots + u_n^2}$$

Note: The above algorithm is a very simplified method of measurement uncertainty estimation and does not constitute the only possible procedure.

In practice, measurement results are accompanied by the **expanded uncertainty** U_C, which is actually a combined uncertainty multiplied by the expansion coefficient (coverage factor) k so that the uncertainty estimate corresponds to the required confidence level. The coefficient value for 95% probability is 1.96, but the rounded value 2 is usually used in practice:

$$U_C = 2 \cdot u_C$$

The specific measurement result is then presented as a value determined by measurement and is accompanied by expanded combined uncertainty, for example: 132 ± 6 nmol/l.

In addition to uncertainty sources associated with the analytical method itself, the resulting uncertainty of a measurement result also includes contributions from sources inherent primarily in the pre-analytical phase of the test. See Figure 3.3.

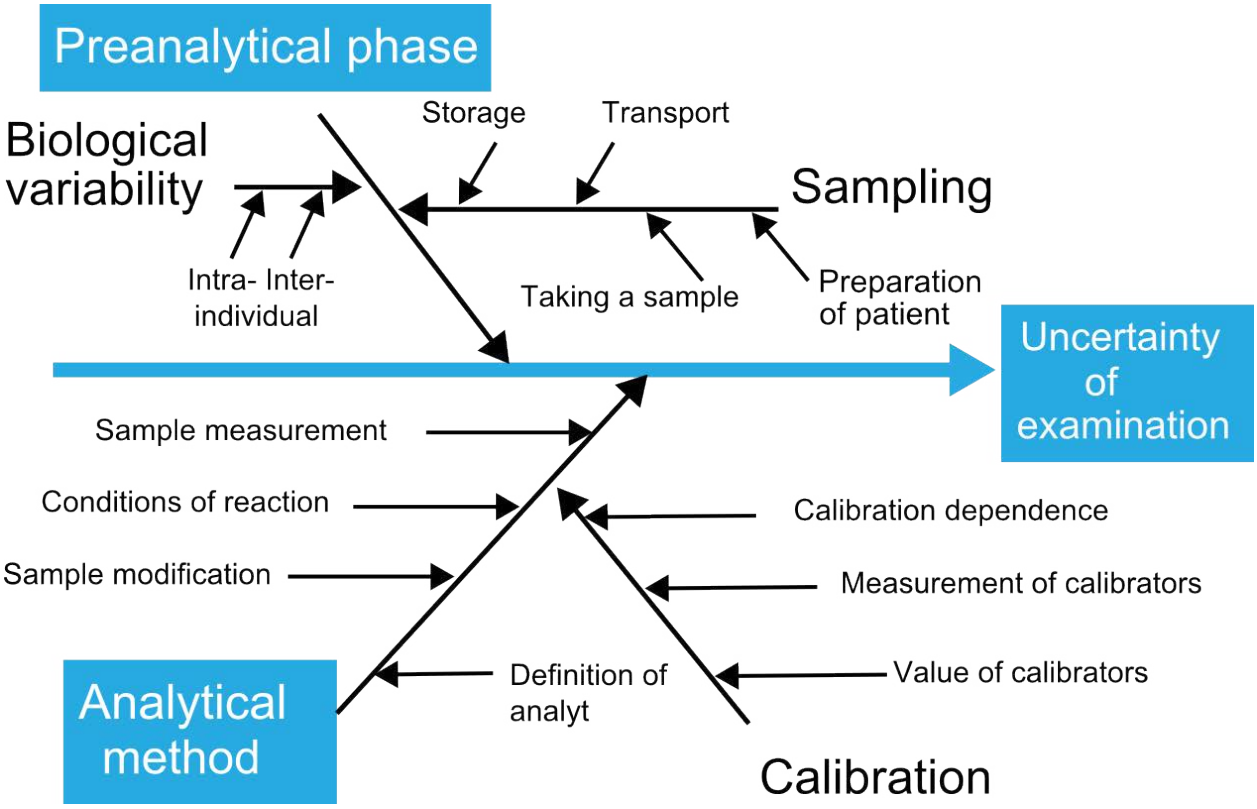


Figure 3.3. Main sources of laboratory test uncertainty in a clinical laboratory

The importance of indicating uncertainty consists not only in the expression of a certain measure of the indeterminacy of a result, which may occur with a given probability within the entire bounded interval, but also the result uncertainty has to be taken into account when interpreting the result against decision limits, especially if the result is close to these limits.

3.2.6. Traceability of the Method

The traceability of a method is the property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty.

This is a very important property of the method in respect of reaching its optimum trueness. If within the method each of the quantity inputs included in the measurement model is traceable to a basic SI unit, then the traceability of the method can be considered optimally ensured and the results of measurement are qualified as comparable on a worldwide scale. This means in practice that particularly the calibrators used in the measurement process are required to be traceable to standards of higher metrological quality. This principle is ensured by the existence of a hierarchical structure of reference materials and methods, through which the working calibrator is gradually traced to the very highest standard, to an SI unit in optimal cases. This hierarchy also comprises different entities responsible for performing individual steps.

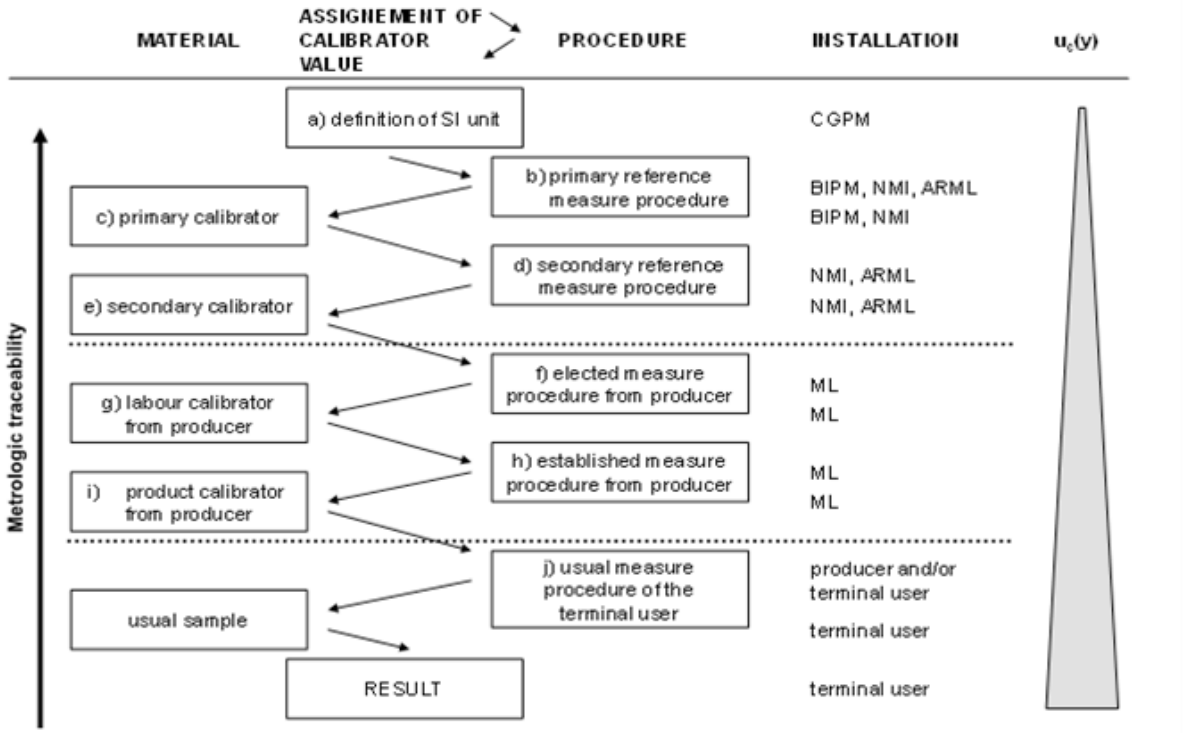


Figure 3.4. Hierarchy of calibration and metrological traceability to SI

Note: according to ČSN EN ISO 17511:2004;

CGPM – General Conference on Weights and Measures; BIPM – International Bureau of Weights and Measures;

NMI – National Metrology Institute; ARML – Accredited Reference Measurement Laboratory; MCL – Manufacturer's Calibration Laboratory; ML – Manufacturer's Laboratory; Mf - Manufacturer

Several types of standards (reference materials) occur at different levels:

- Primary reference material – substances of maximum attainable purity; the materialized measurement unit is produced by weighing or measuring its volume;
- Secondary (certified) reference material – usually a matrix (protein matrix content); it is accompanied by a certificate declaring the analyte content, determined using a method traceable to the primary reference material (usually a reference method with minimum uncertainty);
- Working calibrator – intended for routine method calibration; their values are derived from certified reference materials or manufacturer's working calibrators traceable to them.

Testing the traceability of the method is an integral part of the method validation process.

3.2.7. Sensitivity of the Method

In practice, several characteristics referred to as sensitivity are used. They relate to the ability of a method to distinguish between certain levels of the measurand (mainly in the area of very low values). Various attributes or separate terms are used to distinguish between the characteristics (analytical sensitivity, functional sensitivity, detection limit, or limit of determination).

Caution! It must be kept in mind that these characteristics describe the method in terms of their analytical properties. However, the term "diagnostic sensitivity" is also often used in clinical practice to characterize the method in terms of its clinical usability, which is usually its ability to distinguish healthy individuals from ill ones. Diagnostic sensitivity is described in detail in chapter 4 Sensitivity and Specificity of the Method, Interrelations, Laboratory Screening.

3.2.7.1. Analytical Sensitivity

Analytical sensitivity is the ratio of change in the response of measuring equipment (measuring device output signal) to the corresponding change in the stimulus (input signal, measurand). In other words, analytical sensitivity points to the size of change in the detected response caused by a change in the measurand.

It is constant in linear methods and corresponds to the slope (gradient) of the calibration curve. Analytical sensitivity in non-linear methods will naturally change with the change of the calibration curve slope. The analytical sensitivity of the method depends primarily on the principle of response detection used.

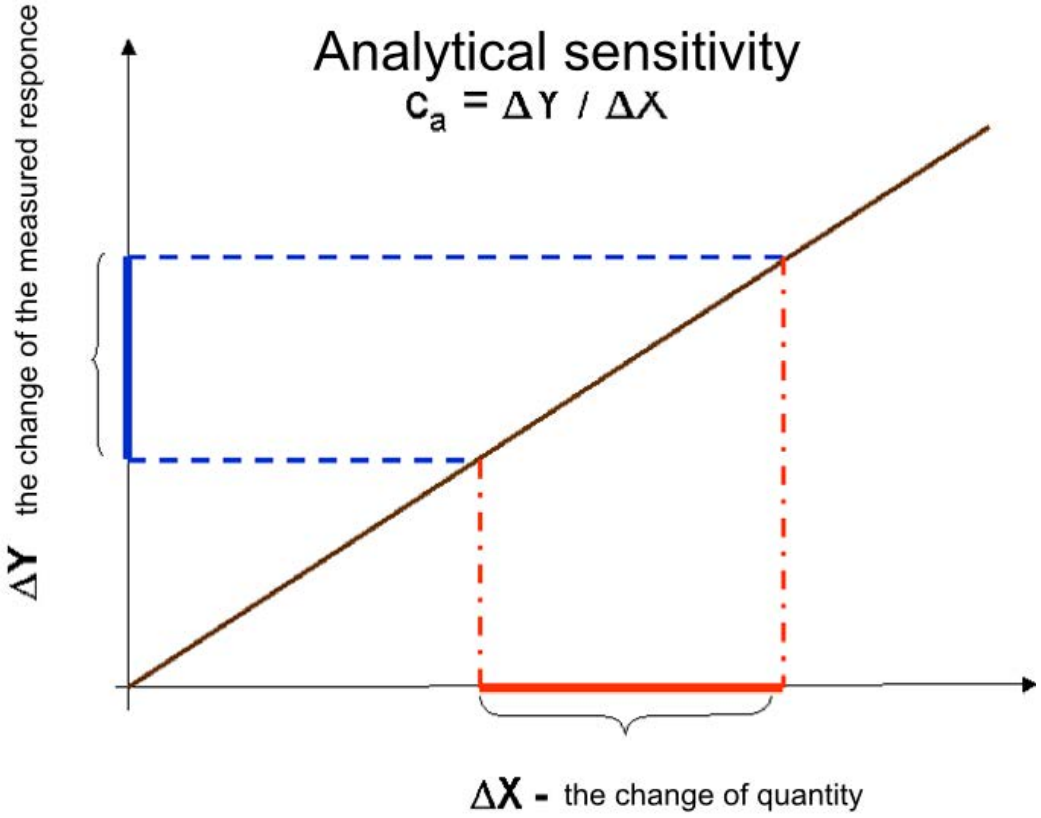


Figure 3.5. Determining the analytical sensitivity c_a of the method

3.2.7.2. Functional Sensitivity

Functional sensitivity of the method is the measurand value (e.g. the measured analyte concentration in the sample) where the repeatability of measurement is at a pre-defined level (usually CV = 20 %). It is simply found from the precision profile, which is the dependence of the precision of the method on the measurand magnitude (see Figure 6).

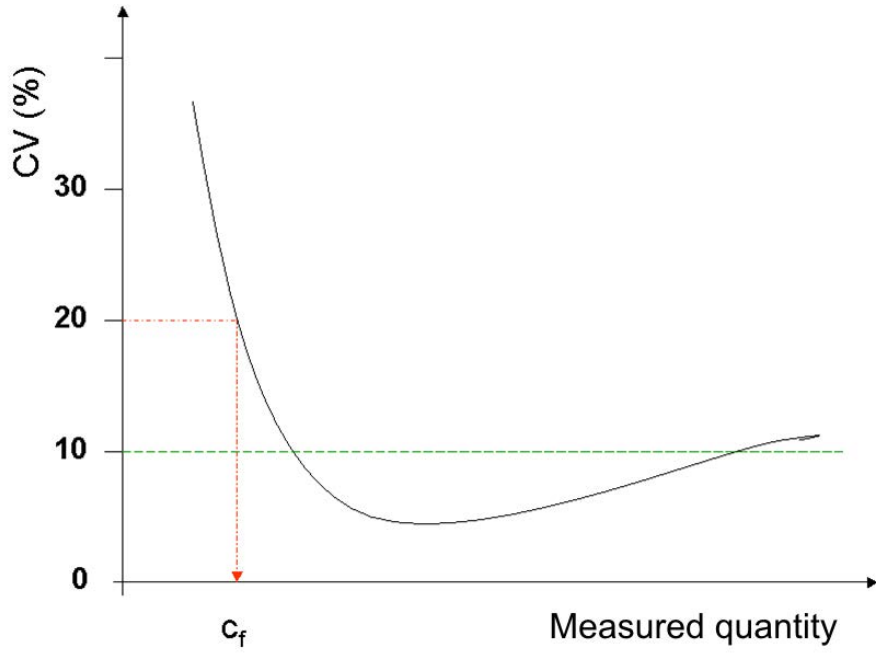


Figure 3.6. Determining functional sensitivity c_f from the precision profile of the method

3.2.7.3. L_D – Limit of Detection

The limit of detection is the lowest amount of analyte in a sample that can be detected (distinguished from zero), but not necessarily quantified as an exact value with certain uncertainty. This limit value separates the area of noise of the detection device from a detectable signal.

It is usually determined from the data of sample measurement with zero measurand content (blank) under conditions of repeatability (5 – 10 repetitions). The measured response values are used to compute the average y_B and the standard deviation s_B. The measurand value corresponding to the signal is then read from the calibration dependence:

$$y = y_B + 3.29 \cdot s_B$$

3.2.7.4. L_Q – Limit of Quantification

In contrast to the above, the limit of quantification defines the lowest amount of analyte in a sample (measurand value) that can be quantitatively determined as an exact value with the required measure of uncertainty.

The calculation is similar to L_D, only the coefficient 3 is replaced by 6 or 10, depending on what measure of uncertainty, or confidence level, is required. The measurand value corresponding to the calculated signal level is then also read from the calibration dependence:

$$y = y_B + k \cdot s_B \; where \; k = 6 \; or \; 10$$

The results of measurement under the limit L_D cannot be issued, and results between L_D and L_Q can be only considered as qualitative results.

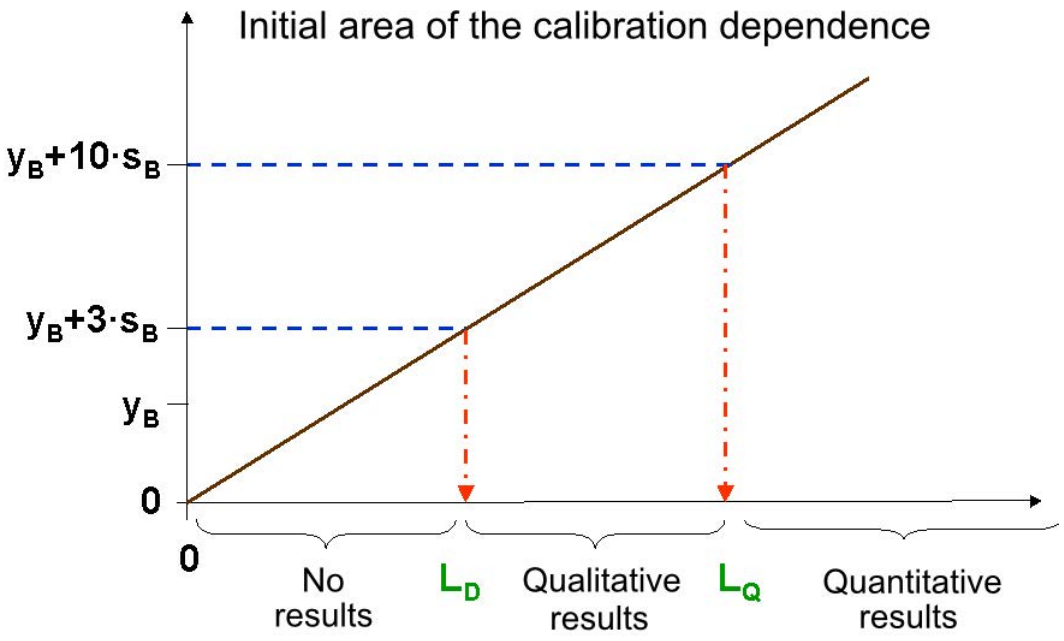


Figure 3.7. Determining the limits of detection and quantification

3.2.8. Linearity of the Method

This is understood as the linear dependence between two random variables, i.e. between the signal (measured response of the analytical method) and the measurand. In other words, it is the ability of a method to provide a detectable measurement response that is directly proportional to the measurand magnitude within the given range (e.g. the analyte concentration in the sample). It is advantageous if the method works in the area of linear calibration dependence because only two measured points, the blank and one calibrator are then sufficient to plot the dependence (a two-point calibration is made). After reading the blank signal, the linear calibration dependence is then defined practically by a single number, the calibration factor *F* that can be easily determined from the quotient of the known measurand value in the calibrator x_K and the corresponding calibrator signal response magnitude y_K. This factor and the measured sample response *y* then can be used to easily determine the measurand value *x* in the sample.

$$x = F \cdot y = \frac{x_K}{y_K} \cdot y$$

If the linear area of the method is limited from above and samples with a greater measurand value need to be analyzed, then appropriate dilution of the sample should be selected.

3.2.9. Working Range

The working range of the method is a closed interval of measurand values that can be determined using a measuring procedure, whereby the interval is most commonly limited by the lower and upper limits of quantification (provided measurement uncertainty in this area is acceptable) or the range of linearity of the method. See Figure 8.

In practice there are, however, also many methods with non-linear behaviour of the calibration dependence. Two-point calibration is obviously not sufficient in such cases. The calibration curve is usually plotted based on the measurement of at least 6 calibrators, while the calculation of the measurand value in an unknown sample is much more complicated in this case since a more complex mathematical apparatus has to be used. The working range of the method is then limited in the upper area of the calibration dependence by other limiting factors (the limit of quantification in the area near 0).

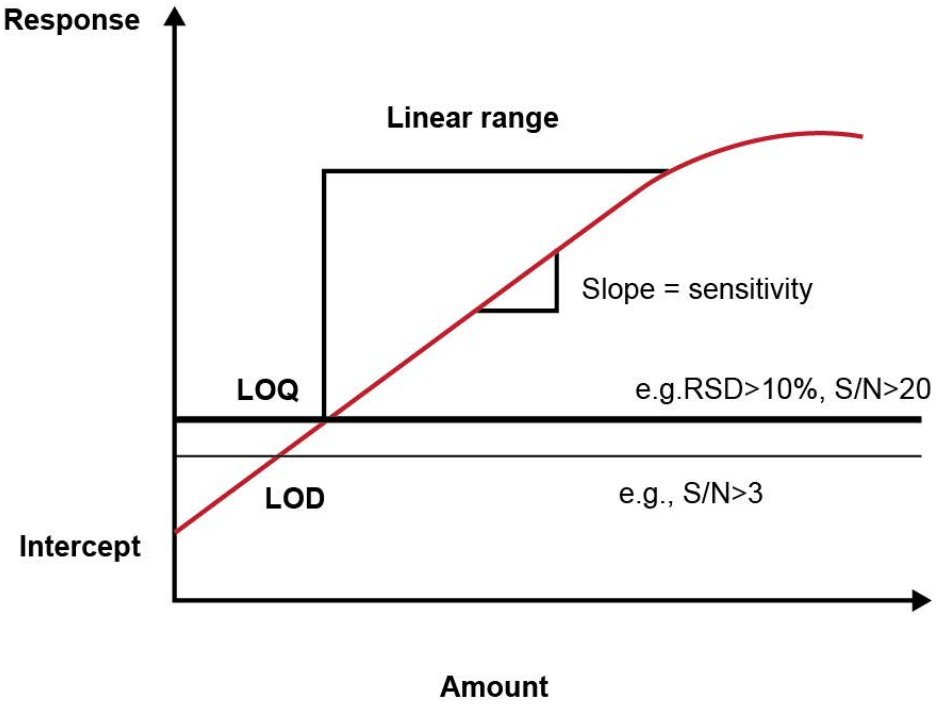


Figure 3.8. Effect of the limit of quantification and linearity on the working range of the method

3.2.10. Analytical Specificity

Analytical specificity is the ability of the measuring procedure to determine only the measurand intended for determination.

Analytical specificity is usually expressed as non-specificity, i.e. as the effect of any sample component other than the analyte that causes a change in the signal and measuring device indication, thereby introducing a systematic error.

The analytical specificity of the method can be worsened, for example, by the presence of various interferents, components commonly present in the biological matrix of analyzed samples. The effects of elevated triacylglycerol, bilirubin or haemoglobin levels are most commonly studied as potential sources of interferences. Also drugs and their active agents may have a marked effect on the course of reactions during the analysis.

Cross-reactions with substances having a structure very similar to the studied analyte occur in some method types such as immunoanalytical methods. These are usually various precursors or metabolites of the analytes, aggregated molecules or their grafts containing identical active sites or antigenic determinants.

Immunoassays as methods based on the use of specific antibodies can also be affected by the presence of various heterophilic antibodies in the analyzed samples. The effect of HAMA (Human Anti-Mouse Antibodies) is often mentioned because their occurrence in the population is growing. The measurement result may be affected in these cases both positively as well as negatively.

The effect of cross-reacting or interfering substances is also connected to contributing to measurement uncertainty. Determining the content of any interfering analyte present in the biological sample matrix increases measurement uncertainty due to an increase in the systematic error.

The action of cross-reactions or interferences on measurement results is difficult to evaluate. Experimental procedures do exist but they are almost impracticable in routine clinical laboratory conditions. They are based on the presence of a sufficient amount of individual potentially cross-reacting substances or interferents. Testing their effect should be part of the method validation. One common way of presenting this information is, for example, stating what amount of some interfering quantity does not cause a measurement deviation greater than 10 %.

3.2.11. Recovery

Recovery is a property of the method indicating the method's ability to cover the entire analyte present in the sample with the measured signal, and so it relates to the overall trueness of the method. It is a measure of efficiency of the method.

Recovery also identifies any influence of the differences between the composition of samples and standards on the assay results.

In mathematical terms, recovery is defined as the ratio or percentage of substances obtained under specified conditions from the total amount of substance that should be theoretically determined. It is obvious from the relative expression that calculated values of this characteristic should tend to 100 %.

$$R(a) = \frac{Q(a)}{Q} \cdot 100\ (\%)$$

where:

R(a)…Recovery of analyte assay in the matrix under specified conditions

Q(a)…Amount of analyte determined by the assay

Q ... Actual amount of the analyte in the matrix

3.2.12. Robustness of the Method

This property describes the ability of a method to provide acceptable measurement results despite small deviations in the measurement procedure or sample composition. There are many factors that may change slightly in the course of the process; typically these include a slight change in temperature or reaction environment, a change in incubation times, change of operator, etc. It is obvious that the aim is to have an analytical method whose response to the slight deviations that can commonly occur in practice is minimal.

3.2.13. Comparability of the Method

As a general requirement, methods used for measuring the same quantity should be mutually comparable. Primarily this is a precondition for the compatibility of results obtained using different methods. Comparability of the method is also important if the method used in the laboratory is changed and a new method is compared with a previous one. It is clear that existing differences may be a source of problems when results are interpreted.

Different regression analysis procedures are used to compare the results of two methods when pair data obtained by parallel analysis of a set of samples are evaluated using both methods. The output is the evaluation of linear regression and the significance of the differences of their parameters from optimum values (the regression line intercept close to 0 and its slope close to 1). It is performed jointly by evaluating the level of correlation between the results of both methods. In the optimal case, the correlation coefficient r should have a value close to 1, i.e. the pair data should be assembled as close as possible to the regression line. Another form of output is a graphical presentation of the comparison. Results obtained using the comparative method (reference/original method) are plotted against each other on the x-axis of the diagram, and results obtained using the compared (new) method on the y-axis – see Figure 9.

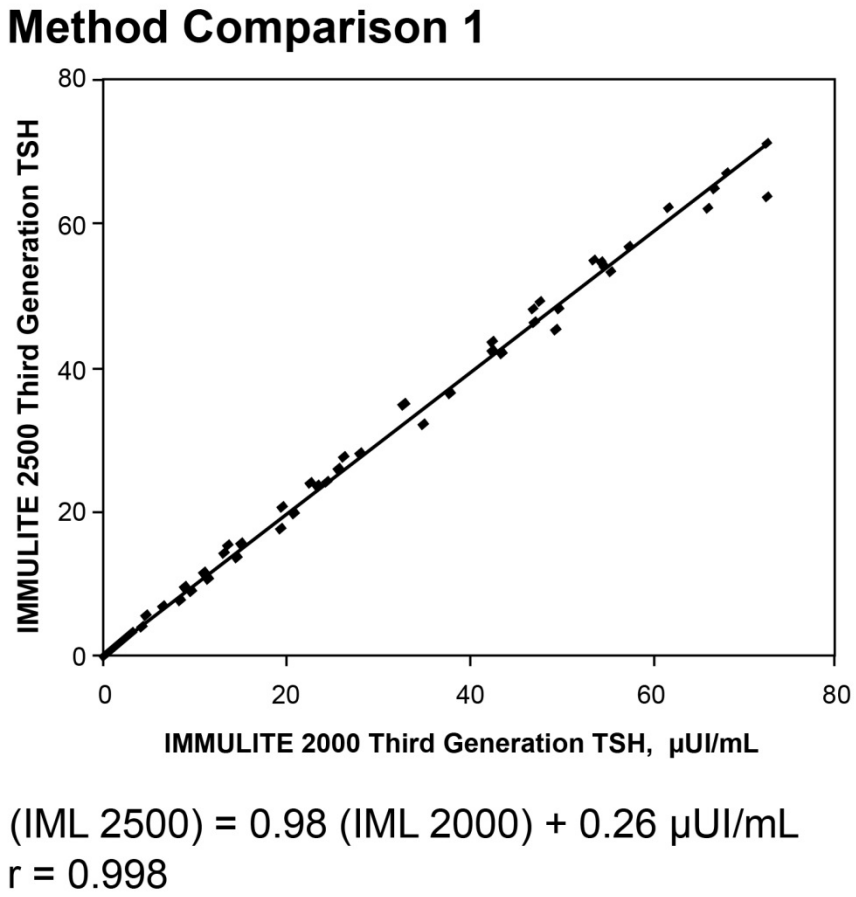


Figure 3.9. Example of graphic comparison of two methods together with regression dependence

3.3. Validation and Verification of Methods

As mentioned in the introduction to the chapter, the basic characteristics of the analytical method are tested at the stage of method development with the aim of confirming that the method complies with the requirements for its intended use. This process is called **method validation** and also includes the evaluation of the method in terms of the intended clinical use. This is a very demanding and complex process that has to be planned in detail beforehand and the process steps and results have to be carefully documented. The process output is a validation report that clearly concludes whether the method is suitable for the given intended use.

European legislation requires that manufacturers or importers of in vitro diagnostic medical devices (IVD MD) should affix the CE marking to the devices before putting them on the European market. Requirements for the conformity assessment may vary depending on the diagnostic device classification, but they always also involve diagnostic agent validation. The use of commercial diagnostic devices is a prevalent trend in the current routine practice of clinical biochemistry laboratories. If a laboratory develops their own analytical method or modifies a commercial method, the requirement for method validation is transferred to the laboratory personnel.

Method verification is performed to a much lesser extent, and only verifies whether the diagnostic agent also fulfils the selected declared properties when used in a specific laboratory. The method is verified in this way not only before putting it into practice, but also subsequently at appropriate intervals, usually once a year. The basic characteristics of the method verified in the verification process primarily include repeatability and the intermediate precision of the method, including the estimate of uncertainty.

3.4. Quality Control

The term quality control generally refers to a set of activities performed in a laboratory with the aim of ensuring the sufficient likelihood of measurement results. These activities in practice only impact upon ensuring the analytical confidence of the methods used. Depending on who takes part in such activities, quality control is further divided into:

- Internal Quality Control (IQC) – performed only by lab personnel;
- External Quality Assessment (EQA) – external entities are called in to organize and evaluate.

3.4.1. Internal Quality Control

This type of control is performed by continuously monitoring the intermediate precision and/or trueness of the method, and is achieved through the continuous measurement of the control samples analyzed in every assay series. Knowledge of the target measurand value in the control sample is required only to monitor the trueness of the method. This option is offered, for example, by various commercial control samples with a test certificate. In other cases, samples without a test certificate can also be used for precision monitoring, while an alternative procedure can be chosen for monitoring trueness.

A series of assays usually refers to the length of one working shift, for continuous or automatic shift operation, or one assay made in a group (batch) of samples, for manual methods.

The number of control samples is not strictly defined and depends on knowledge of the overall quality level of the given method. In practice, 2-3 samples with different measurand value are commonly used. It is a good idea to use one control sample with this value within the reference interval, and the other one outside the reference interval, above the upper limit, for example.

The results of control sample measurements are assessed in connection with results found at an earlier stage. In principle, this means monitoring the magnitude and character of current control sample measurement errors, with the aim of identifying situations that are unacceptable in terms of stipulated quality requirements and avoiding issuing inadequate results by taking operative measures.

The commonest form of this monitoring is visualising the results in diagrams known as control charts. This method is based on evaluating the measurand value fluctuation around the central line of the chart (target/declared measurand values); the allowable range is given by the interval between the lower and upper control limit. These limits are usually an appropriate multiple of the standard deviation derived from the intermediate precision of the method reached in the previous period. An example of a control chart is shown in Figure 10.

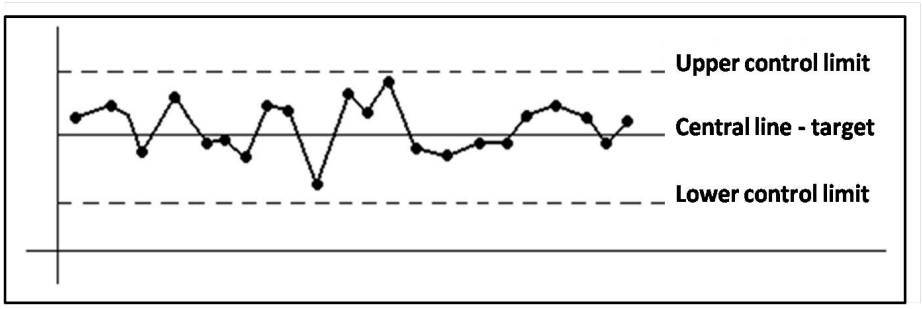


Figure 3.10. Example of a control chart without suspicious results

Two limits are plotted in the chart in most cases – the warning limits at the ±2s level and the control limit (requiring intervention) at the ±3s level. If the deviation of the control measurement result exceeds the value 3s, this is a signal to search out the possible cause of the deviation and to take operative intervention; one such intervention might be to repeat the whole series of sample measurement preceding the assay of controls.

The control chart also shows other situations that may require intervention, for example, if control results repeatedly deviate from the centre line always in the same direction, or if consecutive results have the same trend (see Fig. 11).

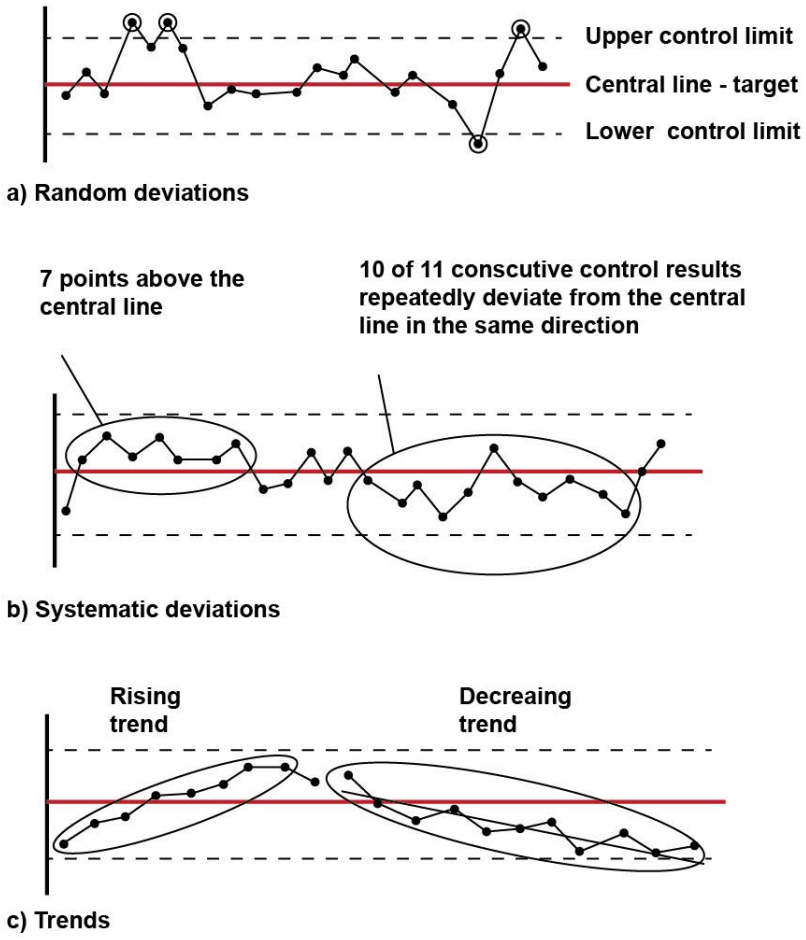


Figure 3.11. Control chart with different types of anomalies

Algorithms aimed at obtaining the maximum probability of error detection with the minimum probability of false series rejection (combined Westgard rules) have been developed to ensure the greatest possible objectivity in assessing internal control results. Today, these algorithms are part of control software analyzers and are also often built into laboratory information systems (LIS), which contributes making evaluation simpler. Some manufacturers of reagents or commercial control samples also offer an option to automatically submit control measurement results through the Internet to an assessment centre, where they are evaluated in the context of other users of the same control materials (example: BioRad, Randox, etc.). This method of the internal quality control in some respects comes close to an external assessment.

3.4.2. External Quality Assessment

The principle of the External Quality Assessment (EQA) is the evaluation of an analytical method by the mutual comparison of measurement results obtained through the analysis of identical samples in different laboratories. This evaluation is organized and performed under predefined and harmonized conditions (in accordance with applicable standards and terms of the region where the evaluation is made).

The EQA arrangement and evaluation is provided by a generally recognized entity – EQA provider. SEKK s.r.o. is the provider for clinical biochemistry tests in the Czech Republic. In addition, there are many foreign EQA schemes. The measurement itself is of course carried out in the participating laboratory. Control measurements are usually provided several times a year as part of various EQA schemes.

The EQA process consists of the following steps:

- Distribution of identical control samples to laboratories; the samples must be as commutable as possible

(i.e. have properties similar to the properties of samples analyzed on a routine basis), must be homogeneous and stable enough so that the requirement for all participants having the same conditions is met;

- Participating laboratories perform assays concurrently within a specific scheme, and send their results to the scheme provider on a prearranged date;
- The results are evaluated by the provider using predefined principles and criteria made known to the participants in advance;
- The evaluation of results is summarized in a pre-arranged format in writing and distributed to the participants; the information is presented as confidential, with anonymous participants identified by codes.

In general, several samples are distributed in one EQA round (mostly 2 samples, but there are control schemes where 5 samples are distributed in one round). The agreement of the participant's result with the assigned value (target value) is assessed during the evaluation. Several types of assigned value can be distinguished depending on the level of traceability of the measurement of a variable and the availability of control material. If a reference method for measuring the variable exists, a Certified Reference Value (CRV) can be the assigned value. Another alternative is the Consensus Value from Experts (CVE), a value obtained by measurement in selected expert laboratories, or the Consensus Value from Participants (CVP), a value computed either from the results of all participants in a specific EQA round or from the results of a group of participants using the same measurement procedure.

To assess agreement with the assigned value, maximum tolerance limits are specified and must not be exceeded if the participant is to be rated as successful. These limits express the acceptable difference D_{max} in percent from the assigned value (they define the neighbourhood of the assigned value where results are considered acceptable). This value is compared with the deviation $D_{\%}$ computed simply as the relative participant's error from the participant's result x and the assigned value X:

$$D_{\%} = \frac{x - X}{X} \cdot 100\ (\%)$$

Despite various providers' efforts towards unification, the specific level of allowable deviations differs. Some of the following possible approaches can be applied to determine them:

- Based on biological variances (intra- and inter-individual variability of the measurand within the population is taken into account in the calculation);
- Based on contemporaneous analytical possibilities (e.g., as a value of interlaboratory precision reached in 20 % of the best laboratories per EQA round).

From the limit applicable to a certain analyte in the EQA scheme, the limit for an internal check can also be derived, and compliance with the limit should ensure successful EQA evaluation. In simple terms, the limit for the internal check (limit of acceptable intermediate precision of the method) should be one third of the EQA limit.

The evaluation involves a simple assessment whether the laboratory deviation $D_{\%}$ is lower than D_{max}. For an EQA round participant to be successful, the absolute value of the ratio of both values, referred to as the P-score, is lower than 1:

$$|P| = \left|\frac{D_{\%}}{D_{max}}\right| \leq 1$$

Some EQA schemes use a different assessment method based on the z-score to assess participant success. This score measures the deviation of participant's result *x* from the assigned value *X* and the detected interlaboratory precision of the EQA round participants' results expressed as a standard deviation. For a participant to be successful, it is required that the absolute value *z* be lower than 3 (values under 1 are considered very good and under 2 acceptable):

$$|z| = \left|\frac{x - X}{s}\right| \leq 3$$

The successful participation of a laboratory in EQA schemes is a mark of the quality of its work, and is required not only by authorities auditing the functioning of the laboratory's quality management system, but also by healthcare payers. Beside this somewhat repressive aspect, EQA schemes have a significant educative character consisting, for example, in the comments from authorized experts (supervisors) attached to audit round results, the provision of consultations, publication of articles in journals, etc.

3.4.3. Quality Management System in Clinical Laboratories

The responsible approach to ensuring the quality of clinical laboratory work cannot be reduced to simply ensuring the quality of the analytical process. In line with general trends, comprehensive Quality Management Systems (QMS) are also put into practice in clinical laboratories. In addition to ensuring the quality of the actual measurement of various quantities, emphasis is given to managing all laboratory operations with an eye to quality.

This concerns not only the actual management of the laboratory, including the provision of high-quality resources (personnel, hardware, laboratory environment and resources used for analysis), the maintenance of proper supplier--customer relationships and responsible behaviour towards laboratory service customers, but also organizing work to ensure that all pre- and post-analytical analysis requirements are met, information is obtained and stored in the laboratory information system, etc.

In order for QMS to function properly, all lab processes and operations should be responsibly described and documented, and subsequently maintained as controlled documents. This creates a situation where only currently valid documents exist in the laboratory; outdated versions are filed and any records related to lab operations required later continue to remain retrievable. The availability of uniform and complete documents to all laboratory personnel for their work is a precondition for the reduction to a minimum of serious errors caused, for example, by the flawed transfer of information.

QMS documentation typically has the following structure:

- Quality Manual describing the whole system;
- Guidelines and rules characterizing the basic laboratory processes;
- Laboratory Manual, a document primarily informing laboratory clients about the services offered and conditions of service provision;
- Standard Operating Procedures describing the steps of individual analytical methods and the operation of technical equipment;
- Records on the use of technical equipment and the quality of measurements.

3.4.4. Certification and Accreditation

As in other industries, health care providers have also been recently under growing pressure over their ability to prove the care taken over the quality of their work. These pressures can be enforced by regulating mechanisms within the society, or they can reflect an entity's effort to succeed in the market environment, or they might simply represent a voluntary effort to present their qualifications for achieving the commensurate quality of their work.

Today, two major internationally recognized and supported mechanisms are in force attesting to the implementation of a functioning QMS, or that certify a certain entity's capacity to perform specified activities. The first is the certification of QMS compliance with the requirements stipulated by the internationally recognized standard. The second is accreditation, a process confirming whether the competence of a certain entity also complies with international standard requirements.

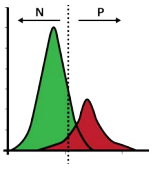
The general requirements for a functional QMS are summarized in the standards of the ISO 9000 series, specifically in ČSN EN ISO 9001. This standard specifies requirements for a QMS implemented at any entity so that the system is set correctly and works towards ensuring the quality (and improvement) of such entity's operations. At present there are many certification bodies, i.e. legal entities, who assess an applicant's QMS compliance with the requirements of the standard mentioned above. There are also many companies (different organizations in general, but also including some clinical laboratories) that are holders of such certificate. Nevertheless, this method is not particularly popular in Czech healthcare or, more particularly, in medical laboratories.

The requirements specified in ČSN EN ISO 15189 particular to medical laboratories are better suited to the specific conditions of the work such undertakings carry out. For this reason, in the clinical laboratory sector the process of accreditation to certify laboratory proficiency in accordance with standard requirements has been adopted to a much greater extent. Within any one country accreditation is usually provided by a single national accreditation body; in the Czech Republic this is the Český institut pro akreditaci, o.p.s. (Czech Accreditation Institute or ČIA).

A similar process is the system for testing how the system of laboratory work has been set up, which is organized by the Národní autorizační středisko pro klinické laboratoře (National Authorization Centre for Clinical Laboratories or NASKL). This body of the Czech Medical Association performs educational audits of labs registered in the Clinical

Laboratory Register. Two consecutive audits I and II test whether the laboratory complies with requirements specified in national accreditation standards. The content of these requirements is very similar to those stipulated by ISO 15189. Both the above forms are continuous processes repeated over time in the form of re-audits and regular monitoring activities.

Proof of care of laboratory quality is currently a necessary condition laid down by some healthcare payers prior to concluding healthcare payment contracts with healthcare providers. For example, Všeobecná zdravotní pojišťovna (VZP) health insurance company recognizes only one of the previous alternatives, i.e. the ČIA accreditation or the NASKL certificate for audit II. The requirement for functioning processes that ensure laboratory quality thus fulfils the analogous role of regulatory mechanisms, virtually allowing the provision of clinical laboratory services only to those laboratories that have undergone these processes successfully and hold the required certificates.



CHAPTER 4

4. Diagnostic Sensitivity, Specificity of the Method, Methods of Determination, Interrelations, Laboratory Screening

Authors: Ing. František Všianský; Ing. Vladimír Bartoš, Ph.D.

Reviewer: Ing. Václav Senft

4.1. Introduction

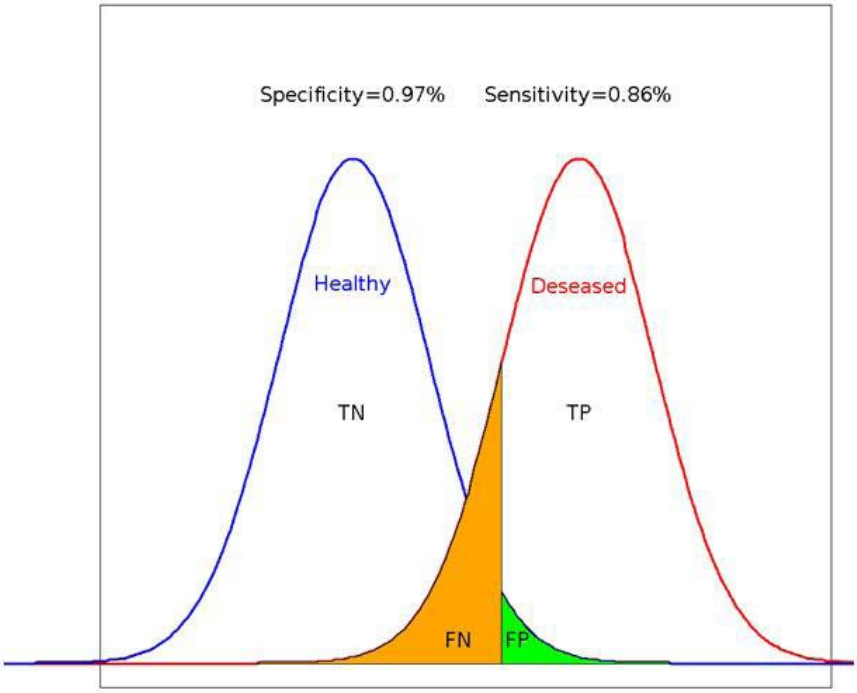
Apart from the basic analytical characteristics of the method, persons evaluating laboratory methods must also know the clinical (diagnostic) characteristics of the method in accordance with the patient diagnosis. These characteristics allow the efficiency of biochemical methods to be compared in respect of their ability to distinguish between the healthy and the diseased.

Calculations of these clinical characteristics are based on a four-fold table, where a set of individuals is divided into a group of healthy and diseased individuals based on a definitive method, and the individuals have been examined by the tested laboratory test. Depending on the laboratory method result, the individuals can be divided into 4 groups:

- The diseased with a positive test result (True Positive – TP);
- The diseased with a negative test result (False Negative – FN);
- The healthy with a positive test result (False Positive – FP);
- The healthy with a negative test result (True Negative – TN).

	Healthy	**Diseased**
Positive test	FP	TP
Negative test	TN	FN

This situation can also be shown on a graph where the x-axis represents the measurand value, the tested analyte concentration, for example, and the y-axis represents the relative frequency of this quantity occurrence in the group of healthy individuals (left-hand curve) and in the group of diseased individuals (right-hand curve). The picture is only illustrative; mutual position and the forms of the two curves will naturally vary for each method.



4.2. Diagnostic Sensitivity and Specificity of the Method

These terms are often confused with analytical sensitivity and specificity (often just because only sensitivity and specificity are mentioned). However, they characterize the smallest amount of analyte the method is able to capture and whether the method specifically detects just one substance. Therefore, they are not relevant at all to the ability of the method to distinguish between the healthy and the diseased.

4.2.1. Diagnostic Sensitivity

This parameter is used to diagnose the disease, and evaluates the ability of a laboratory method to detect the presence of a disease. Diagnostic sensitivity takes values from 0 to 1 (0 – 100 %). The higher the sensitivity, the better is the ability of the test to detect the presence of a disease.

$$Senzitivity = \frac{SP}{health} = \frac{SP}{SP + FN}$$

4.2.2. Diagnostic Specificity

This is used to rule out a disease in a group of healthy individuals, and expresses the probability of a negative result in a healthy person. Again, this parameter's values may range from 0 to 1 (0 – 100%). It is advisable that the laboratory method should have the greatest possible diagnostic specificity value; the greater the specificity, the better is the ability of the test to filter out healthy individuals from the studied set.

$$Specificity = \frac{SN}{health} = \frac{SN}{SN + FP}$$

4.2.2.1. Cut-Off

This is the laboratory test result value used in practice to distinguish individuals with the presence of a disease from individuals without the disease. Consequently this means that if the cut-off is exceeded, some action is taken (diagnosis is established, treatment is started, etc.). The cut-off is set to have an optimum sensitivity to specificity ratio for the given purpose (as high a sensitivity as possible for screening and, conversely, a high specificity to prevent a risky therapeutic intervention in a healthy person).

The chosen discrimination cut-off value then corresponds to specific numbers of individuals included in the groups

– True Negative (TN), False Negative (FN), True Positive (TP) and False Positive (FP). For the given method, the chosen cut-off value will always correspond not only to the specific level of sensitivity and specificity attained, but also to likelihood or predictive values (see below).

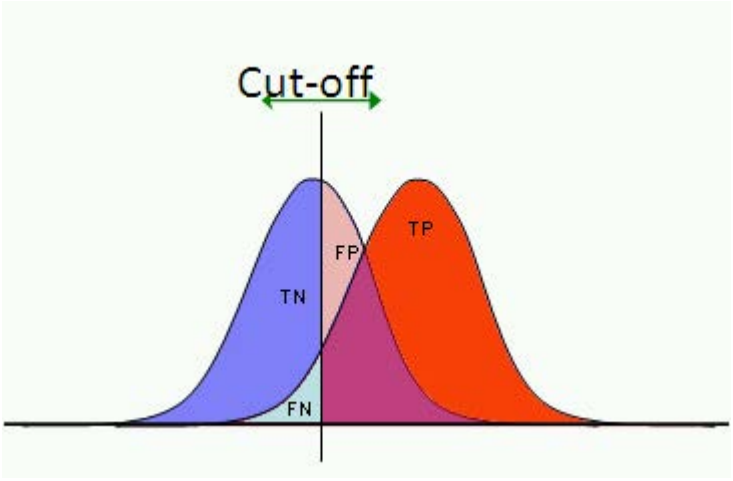


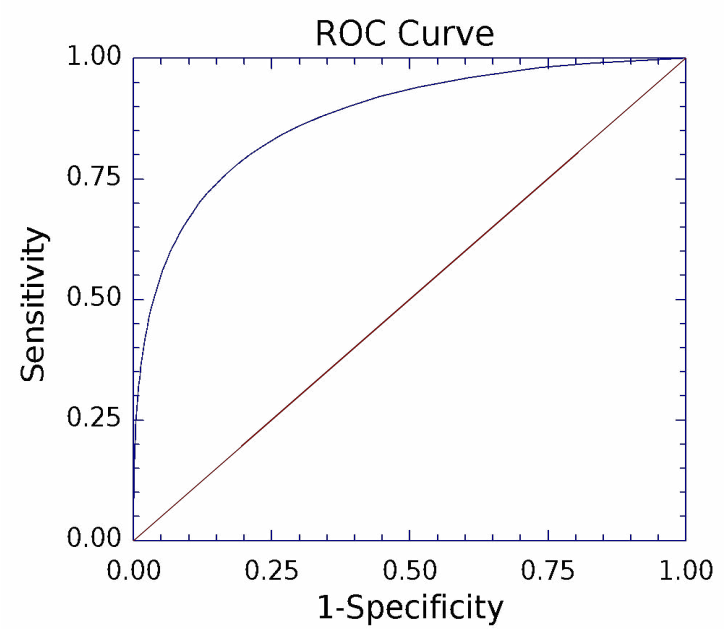
Figure 4.1. The cut-off (e.g. selected according to expert-based consensus) is represented by the vertical line in the picture.

4.2.2.2. ROC Analysis

It is obvious from the calculation of the diagnostic sensitivity and specificity values of a method that both these characteristics of a method behave as communicating vessels. If the sensitivity of a method falls, its specificity rises and vice versa. The selection of a certain cut-off value is decisive for the corresponding values of both characteristics.

The relationship between diagnostic specificity and sensitivity is evaluated by the ROC analysis. (ROC stands for **R**eceiver **O**perating **C**haracteristic. The term dates back to World War Two when this method was used to assess the ability of radars to discriminate real signals from noise.) In practice, this relationship is graphically represented using the ROC curve.

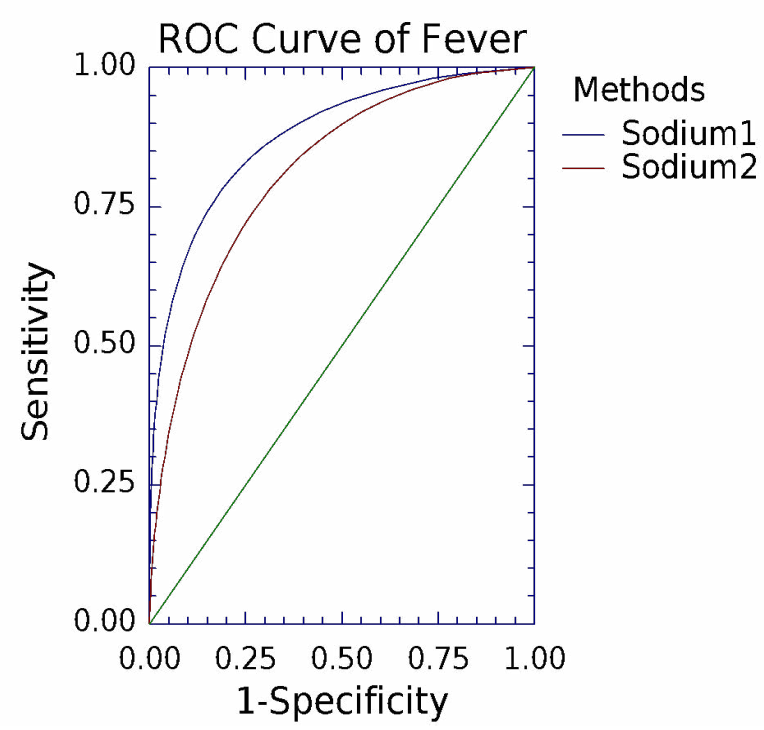
The ROC curve is usually represented as the relationship between the true positive rate or diagnostic sensitivity (ranging between 0 and 1, or 0 and 100 if expressed as a percentage) on the y-axis and the false positive rate, i.e. 1-specificity (or 100-specificity if expressed as a percentage) on the x-axis. It is sometimes also expressed directly as the diagnostic specificity, in which case it directly concerns the relationship between the sensitivity and specificity of a method. However, the range of values on the x-axis is then shown contrariwise, i.e. from 1 to 0 (or 100 – 0 %). The curve itself represents the set of points corresponding to the specificity and sensitivity always determined for the chosen cut-off value. The sensitivity for certain specificity and vice versa, attained using the method with the chosen cut-off, can then be read on the axes.



The evaluation criterion used is the area under the ROC curve (AUC, Area Under the Curve). The area equals 1.0 for the theoretically most efficient method. For an inefficient method, the curve coincides with the diagonal and the AUC equals 0.5. Methods can be classified in the following way depending on the AUC size.

AUC	**Rating**
0.90 – 1.00	Excellent
0.80 – 0.89	Good
0.70 – 0.79	Fair
0.60 – 0.69	Poor
0.50 – 0.59	Fail

Comparing ROC curves for different methods also allows the clinical usability of methods to be mutually evaluated. A clinically more suitable (better discriminating) method has the ROC curve located more to the upper left corner, while the curve of a worse discriminating method is close to the diagonal.



A suitable statistical program can be used to test whether the AUCs differ from each other, i.e. whether the efficiency of one method is significantly better than the efficiency of another.

4.3. Other Clinical Characteristics

4.3.1. Prevalence

This is the probability of the occurrence of a disease in a population. Sometimes also referred to as an a-priori or pre-test probability.

4.3.1.1. Positive predictive value, PV^+

This expresses the probability of the occurrence of a disease with a positive test result. It is influenced by the prevalence of the disease. Where prevalence is low, even an "excellent" diagnostic test has an elevated number of false positive results.

$$PV^+ = \frac{TP}{(FP+TP)}$$

4.3.1.2. Negative predictive value, PV^-

Expresses the probability of the absence of a disease with a negative laboratory test result. It is also dependent on the disease prevalence.

$$PV^- = \frac{TN}{(FN+TN)}$$

4.3.2. LR – Likelihood Ratio

4.3.2.1. Positive likelihood ratio, LR^+

This expresses the ability of a laboratory test to capture the presence of a disease with a result higher than the cut-off.

It indicates how many times the probability of an increase above the cut-off in diseased individuals (relapse) is higher than the probability of an increase above the cut-off in a group of individuals without the disease.

$$LR^+ = \frac{sensitivity}{(1-specificity)}$$

4.3.2.2. Negative likelihood ratio, LR^-

It indicates how many times the probability of a result below the cut-off in a group of individuals without a disease is higher than the probability of a result below the cut-off in a group of diseased individuals.

$$LR^- = \frac{(1-sensitivity)}{specificity}$$

4.3.2.3. LR interpretation

- LR = 1 Test without any diagnostic value (ROC is exactly the diagonal of the graph; the method cannot discriminate between the healthy and the diseased);
- LR+ > 1 The test result is usually considered a highly positive result;

- LR- < 1 The test result is usually considered a highly negative result.

4.3.2.4. Odds ratio

The odds ratio is the ratio of the chance of a positive test in diseased individuals to the chance of a positive test in healthy individuals. Or, the odds ratio can be understood as the chance of a disease occurrence when the test is positive to the chance of a disease occurrence when the test is negative.

$$\text{Odds ratio} = \frac{(TP \cdot TN)}{(FP \cdot FN)}$$

The odds ratio is used for the retrospective evaluation of studies where the numbers of patients with whom the study objective was/was not achieved are known. For example, the study objective can be to differentiate patients with the risk (chance) of a disease from patients without such risk (chance).

OR = 1 means the inability of a test to divide patients according to the test result into groups with different chances. The higher the odds ratio, the higher is the discriminating capability of the test.

4.4. Laboratory Screening

Screening is an organized activity with a widespread impact, intended to find people with a significant risk of a certain disease among other people in the population. The aim is to find such people at an early or subclinical stage of the disease so that they might benefit from their participation in the screening programme thanks to a timely intervention. Depending on the selected population screened for the disease, we distinguish between universal screening (congenital hypothyrosis screening) and selective screening (in the at-risk population – mammographic screening).

In order to divide a population into healthy and at-risk or potentially ill individuals, appropriately selected and, if possible, simple screening tests are used, either clinical or laboratory tests or a combination thereof (sometimes even several laboratory tests supplemented by clinical tests – Down syndrome screening).

The screening test divides the population of tested people into two groups:

- A group of individuals with a positive test and a high probability that they have the disease or are at risk;
- A group of individuals with a negative test and a high probability that they do not have the disease or are not at risk.

Following the primary screening test, individuals with a positive screening result undergo a highly specific diagnostic test, which greatly reduces the number of people without the condition that had been formerly evaluated as diseased.

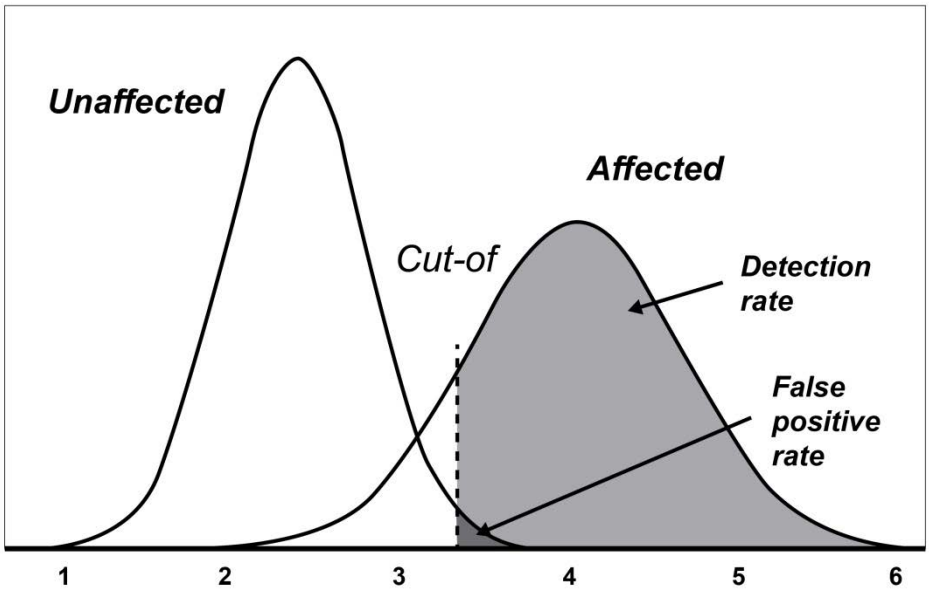
As in every laboratory test, the test used for screening purposes has its clinical characteristics, i.e. its diagnostic specificity and sensitivity.

The proportion of healthy individuals in the total number of individuals in the first group is the False Positive Rate (FP or FPR) of the screening test. The *specificity of the screening test* is the difference 1 – FP, in other words the ability to detect genuinely healthy individuals as being healthy.

The proportion of diseased individuals captured by the test in the total number of the diseased is referred to as the *sensitivity of the screening test* (DR – Detection Rate).

A suitable screening test is required to have the highest sensitivity possible and sufficient diagnostic specificity (a minimum false positive rate) at the same time. A test with low non-specificity (a false positive rate) is acceptable provided the subsequent diagnostic test in the second round rules out individuals with a false positive screening test result.

In quantitative tests there is a relationship between the two parameters, which is given by the mutual position of test value distributions for the healthy and affected populations (the overlap) and the position of the discriminating cut-off limit, which defines from what screening test value the result will be taken as a positive result.

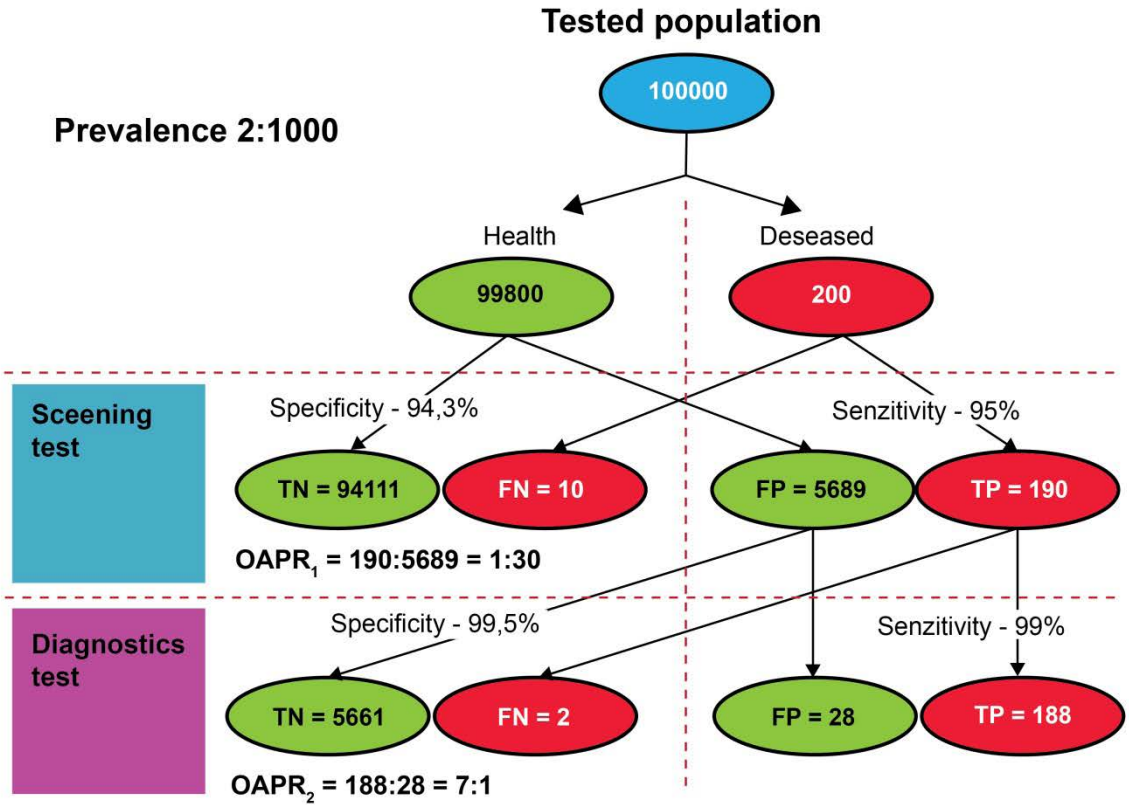


These parameters (the specificity or sensitivity of the test) do not give us an entirely objective assessment of the efficiency of a screening programme. To achieve this, the prevalence of the disease must be taken into account. The probability that an individual with a positive screening test is genuinely ill is described by the positive predictive value PV+. The PV+ is greatly influenced by prevalence: if the disease prevalence is low, the PV+ will be low even in the case of a highly sensitive and specific test.

A characteristic of screening test efficiency describing disease prevalence is a parameter referred to as the OAPR (the Odds of being Affected given a Positive Result), which indicates what the chance is of a diseased individual being captured by the screening test. The parameter is calculated as a ratio of truly positive individuals with positive test results to the number of negative individuals with a positive test result:

$$OAPR = \frac{TP}{FP} = \frac{Prev.senzitivity}{(1 - Prev)(1 - specificity)}$$

The OAPR parameter is expressed as a ratio of two figures (example 1:20) and expresses the disease prevalence in a group of people selected by screening.


Tested population
Prevalence 2:1000
100000
Health
Deseased
99800
200
Sceening test
Specificity - 94,3%
Senzitivity - 95%
TN = 94111
FN = 10
FP = 5689
TP = 190
OAPR1 = 190:5689 = 1:30
Diagnostics test
Specificity - 99,5%
Senzitivity - 99%
TN = 5661
FN = 2
FP = 28
TP = 188
OAPR2 = 188:28 = 7:1

CHAPTER 5

5. Basic Urine Tests

Author: MUDr. Daniel Rajdl, Ph.D.

Reviewers: Doc. MUDr. Milan Dastych, CSc. MBA; MUDr. Marie Šolcová

5.1. Summary

Kidneys are organs which play a fundamental role helping to maintain an optimum internal environment throughout the body. The intermediary for such role is **urine**, and it is urine which therefore shows how the kidneys are functioning and reflects the blood composition. Urine is produced by filtering blood through the glomerular filtration membrane of nephron ([1]) and by subsequent modifications (reabsorption of important substances such as water and excretion of waste substances such as urea). These modifications take place in renal tubules. This chapter deals with basic urine tests that should always be an integral part of kidney examination. Further renal function tests are described in the chapter Kidney Function Tests.

5.2. Sample Collection and the Pre-Analytical Phase

Urine is one of the most easily obtainable biological materials and urine tests bring valuable information. At the same time, it is a potentially infectious biological material which requires the use of gloves and respect for all the rules of working with biological material. Most mistakes that may eventually irrevocably impair the urine test result stem from the method of collecting and storing the sample and by failure to adhere to the time to analysis. The basic types of clinically used samples are shown in *Table 5.1*.

Urine Type	Typical Use	Comment
First morning urine	Chemical urine tests and sediment Pregnancy tests	See also random specimen
Second morning urine	Microalbuminuria, glycosuria	Microalbuminuria is expressed in mg/mmol creatinine
Random specimen	Chemical urine tests and sediment, bacteriology (cultivation + sensitivity), amylase, etc.[1]	Taken any time during the day
Catheterized specimen	Bacteriology (cultivation + sensitivity)	
24-hour urine specimen	Creatinine clearance Waste in urine	Frequent mistakes in collection Example: ion waste (e.g. Na, K) protein, nitrogen, hormones
Short-term collection (about 4 hrs)	Fractional excretion	See Kidney Function Tests
Pediatric specimen[2]	Chemical urine tests and sediment (little children with nappies)	Warn parents about the difficulty of the task

[1] Nephron is the basic functional unit of the kidney. It consists of the glomerulus (a tuft of capillaries with a selectively permeable filtering membrane; blood is filtered here to form primary urine) and tubules (the basic elements are the proximal tubule, loop of Henle, distal tubule and collecting ducts). The function of tubules is to absorb, excrete and metabolize, etc.

Table 5.1. Urine samples: Basic types and use

„Non-collected" (single) urine samples in general are taken after carefully cleaning the genitals (normal washing is enough, immoderate use of disinfectant is not acceptable), a midstream specimen of urine is caught in the test tube (at least about 2-3 seconds from the start of urination). Collecting a correct sample from women during menses (without catheterization) is impossible.

Urine is usually collected over 24 hours (shorter intervals are also possible) using the following steps:

- After getting up in the morning, the patient urinates normally, i.e. not into the collection container.
- After that time, the patient collects all voided urine in the container(s) provided.
- The next morning after getting up (exactly 24 hrs later) the patient urinates into the collection container.

Then the entire volume of urine collected is mixed (if several containers are used, the contents should be emptied into one big container), measured (accurate to 100 ml) and a representative sample is taken.

Tests requiring urine collection (e.g. creatinine clearance) are being abandoned at present as collection is often subject to considerable error (incomplete collection, poor sample mixing, inaccurate measurement of the sample volume, etc.) and is inconvenient for patients ([2]). So, if there is a relatively reliable alternative not involving urine collection, it should be preferred.

The time from sample collection to analysis, time to analysis (see below for specific examples), should be adhered to for each urine analyte. If this is not possible (e.g. if timely transport of the sample to the laboratory cannot be provided), preservatives that extend this time can sometimes be useful. The simplest and commonest preservatives are cold and dark, i.e. a refrigerator. Storage in a refrigerator prevents significant bacterial growth for about 24 hours. Before analysis, allow the sample to reach room temperature (otherwise amorphous phosphates and urates may precipitate in the sediment, and increased specific gravity may occur). Chemical preservatives are usually only suitable for one or a few analyses, and, as a downside, may interfere with other assays; in practice, their use is generally reserved for specific applications (see Pre-Analytical Phase). There are also commercially available test tubes that (according to the manufacturer) are able to provide reliable sample preservation for up to 72 hours (at room temperature).

Failure to deliver urine samples to the laboratory (and analyze them) in time may significantly affect the results. Some alterations relate to bacterial multiplication in the sample: bacteria metabolize glucose (which then decreases in the sample), some bacteria have the enzyme urease (which metabolizes urea in the sample to CO_2 and NH_3 [ammonia] – which binds H^+ to form NH_4^+, thus increasing the pH of the sample). In addition, ketone bodies may evaporate or bilirubin may oxidize (when exposed to light). Refer to the chapter Pre-Analytical Phase for more details.

5.3. Physical Properties of Urine

Physical properties of urine include density, appearance (colour, turbidity), odour or even taste ([3]). Density will be dealt with below; the other physical properties of urine usually have a very approximate or negligible diagnostic significance.

Appearance of urine can be the main reason why the patient visits the doctor. Normal urine is straw-yellow and the intensity of this colour usually roughly corresponds to urine density. Normal urine is also clear, without turbidity (unless there has been sample collection error or artificial alterations due to urate/phosphate precipitation when stored in the refrigerator). The commonest pathological urine colours are summarized in *Table 5.3*.

Colour	Cause	Comments
Light yellow to colourless	Normal after larger fluid intake; polyuria in diabetes mellitus, diabetes insipidus	Increased urine volume per 24 hrs in polyuria; glucose in urine in diabetes mellitus

[2] Ensuring adequate conditions for collection can be difficult in outpatient conditions for an active person leading a normal life.

[3] This is no longer an acceptable way of testing; formerly used to diagnose diabetes mellitus, for example (sweet urine).

Dark yellow	Concentrated sample, use of riboflavin (vitamin B_2)	Normal in first morning sample or after strenuous exercise
Amber	Dehydration during fever or burns	
Orange	Bilirubin, urobilin (yellow-orange colour, formed by photo-oxidation of urobilinogen), drugs (nitrofurantoin), carotenes from food	Bilirubin and urobilinogen can be detected using a test strip (not after photo-oxidation)
Yellow-green, yellow-brown	Biliverdin (bilirubin oxidation)	Test for bilirubin is then false negative
Pink/red/brown	Presence of blood, methaemoglobin is brown (haemoglobin oxidation with acidic pH of urine); haemoproteins (myoglobin, haemoglobin porphyrins (chemical blood test is negative) drugs (e.g. rifampicin) food (raw beetroot + alkaline urine = red; blackberries + acidic urine = red; blueberries [pink -> red], senna [yellow to reddish brown], rhubarb [brown])	Erythrocytes, haemoglobin and myoglobin react with a blood test strip (positive result). Distinguishable by the appearance of urine (usually turbid in presence of blood) or serum (may be red in intravascular haemolysis, not in myoglobinuria); or by specific tests (e.g. myoglobin in urine, complement fixation reaction)
Brown Black	Bilirubin, urological tea, melanoma, alkaptonuria	Chemical blood test is negative
Blue/Green	Bacterial infection (Pseudomonas), drugs (methylene blue, amitryptilyn …), inborn errors of amino acid metabolism[3], carbolism	

Table 5.2. Commonest pathological urine colours (modified according to [2])

Turbidity of urine need not be a pathological sign (collection error, especially in women; presence of sperm; white turbidity can be caused by urate or phosphate precipitation in a sample stored in the refrigerator), although turbidity is often seen in the presence of leukocytes and bacteria (urinary tract infection) or blood.

An unusual **urine odour** may warn of ketoacidosis (sweet, fruity odour) or an inherited metabolic disorder (e.g. the "musty" or „mousy" smell in phenylketonuria, or maple syrup smell in maple syrup urine disease).

5.4. Chemical Examination of Urine Using Test Strip (Dipstick)

This is a very common preliminary examination, which takes us closer to or directs us towards the definitive diagnosis and therapeutic action. Urine chemistry is typically indicated in the following diseases and conditions:

- Urinary tract infection
- Diabetes mellitus

- Renal function tests + suspected renal disease (screening)
- Jaundices (hyperbilirubinaemia)

In outpatient conditions, the test result is commonly read by eye; the process is automated and more objective in hospital labs (reflective photometry-based measurement). Instructions for correct use of the test strip:

1. Mix the urine sample thoroughly.
2. Allow samples stored in the refrigerator to reach room temperature.
3. Completely immerse the strip in the sample for a short time.
4. Run the edge of the strip against the rim of the specimen container to remove excess urine. Put the strip on disposable absorbent material (e.g. pulp).
5. Compare the colour change of reagent pads to the manufacturer's colour chart (usually on the package label). Read results according to the chart's time frame (see manufacturer's instructions) and under good lighting conditions.
6. Do not forget about possible interferences and pre-analytical errors during interpretation.

5.4.1. Individual Determinations

Test strips contain either individual tests or a combination thereof (e.g. for urinary tract infection or diabetes mellitus compensation check), but they often contain all of the 10 most frequent parameters. The tests are outlined below – the principle of determination is mentioned only if it is important for interpreting results (relatively frequent false positive or false negative results).

5.4.1.1. Specific Gravity (Density)

Specific gravity expresses the density ratio of the sample to distilled water, and so is a dimensionless quantity. The specific gravity of plasma is 1.010, and typically ranges from 1.015 to 1.025 in the definitive urine of a healthy person ([4]). It is measured using physical (e.g. densitometer or refractometer) or chemical means ([5]). Individual principles of measurement do not provide comparable results; osmometer measurement is most accurate ([6]).

The specific gravity of urine reflects renal tubule functions (the secretion and absorption of ions and water, in particular) and may be one of the first signs of renal damage (in particular a loss of reaction to changes in fluid intake).

Indications: Distinguishing between pre-renal and renal cause of renal failure, test for renal concentrating ability (this indication requires more accurate measurement for urine osmolality, see also Kidney Function Tests).

For a pre-renal cause of renal failure (e.g. dehydration), a high specific gravity of urine can be expected (maximum concentration); with a renal cause, tubule functions are damaged and the kidneys produce less concentrated urine.

5.4.1.2. pH

Hydrogen ion concentration in urine reflects the balance of H^+ production, metabolism and excretion; however, it may also be a sign of a kidney or urinary tract disease. First morning urine pH ranges from 5 to 6, but a random sample from a healthy person may also have a pH of 4.5 or 8 (more alkaline after a meal, more acidic after physical strain, dietary influences…) ([7]). The reference range for urine pH is rather misleading, and the result should be evaluated along with other data (e.g. acid-base balance status). Acidic urine is found in acidosis and alkaline urine in alkalosis (compensation or correction) of the disorder by the kidneys; unless of course the kidneys are the cause of the acid-base balance disorder). Vegetable diets (low protein content, e.g. a vegetarian diet) usually lead to a lower basic production of H^+ and more alkaline urine.

Indications: Primarily the diagnosis of urinary tract infection, but also the evaluation of acid-base balance disor-

[4] It may also range from 1.003 to 1.035 (samples with specific gravity < 1.003 are probably not urine, > 1.035 might occur following i.v. administration of radiopaque substances). The very wide range is caused by the influence of hydration (fluid intake) on this parameter.

[5] Or as osmolality, typically using a cryoscopic method (see Kidney Function Tests).

[6] Osmolality only depends on the particle count in the solution. Density is also affected by particle size. Therefore, Na^+ will increase specimen density less than a higher urea or glucose content. The chemical principle of specific gravity measurement (on a test strip) detects only ions and so it is not affected by the presence of urea, glucose or a contrast agent; however, the presence of protein (anion) leads to a false increase in the results. In clinical terms, these limitations are usually negligible due to the preliminary character of the examination.

[7] Specimens of urine with pH < 4.5 and > 9 are an indication for a new sawmple collection (probably contaminated or artificial proliferation of bacteria with urease).

ders or the diagnosis and monitoring of urolithiasis treatment.

Urinary tract infections caused by bacteria with urease (e.g. Klebsiella) cause alkaline urine pH. The creation of renal calculi also depends on urine pH. Renal calculi (calcium oxalate) are typically formed in acidic urine and it is advisable to maintain a more alkaline urine pH (mainly by dietary measures) to prevent recurrence.

Falsely high values can be expected if the sample is supplied late (bacterial activity, see above); falsely low values may occur if the sample is contaminated by reagents from the adjacent reagent pad (typically in protein determination where the reaction takes place in a strongly acidic environment).

5.4.1.3. Leukocytes, Nitrites

The main purpose of both tests is to rule out/confirm suspected urinary tract infection ([8]) and, if needed, subsequent indication of microbiological urine culture, and identification of the pathogen and pathogen sensitivity to antibiotics. This test is not therefore used for the precise diagnosis and therapy of urinary tract infection (this is done based on microbiological examination). The leukocyte test detects one of the granulocytic leukocyte enzymes (esterase). The test for nitrites utilizes the ability of the bacteria to reduce nitrates to nitrites, which is an ability possessed, for example, by enterobacteria such as E. coli or Proteus. The true positivity (sensitivity) of the test is conditional upon urine being in the urinary bladder for at least 4 hours (required to reduce a sufficient amount of nitrates) and the patient having a sufficient amount of nitrates in the diet ([9]).

Indications: Diagnostics, urinary tract infection screening

5.4.1.4. Protein

The determination of protein (albumin) using a test strip is one of the most important basic urine examinations, since it can reveal a developing renal pathology at an early stage. A small amount of albumin (relative molecular weight is about 70 kDa) and all small proteins (microproteins) penetrate through a healthy glomerulus, although most of these physiologically filtered proteins are then reabsorbed by proximal tubule cells. In a renal disease, either the glomerular membrane (glomerular proteinuria) or tubular cells (tubular proteinuria) are damaged. The limit for proteinuria has been set to 150 mg/24 hrs.

Principle of determination: Acid-base indicator that changes colour in the presence of protein (particularly albumin that has many binding sites for protons and may remove them from the indicator). To avoid colour changes due to changes in urine pH, the strip contains a buffer, ensuring a constant (acidic) pH of around 3 (see Figure 5.1). This method of determination is most sensitive to albumin (less so to other proteins). The strip usually captures albumin concentrations over 150 mg/l.

[8] For the early detection of cystitis, especially in cases of inapparent clinical problems.
[9] Vegetables are the main source of nitrates in the diet. This is why the urine nitrite test may be falsely negative in some inpatients (e.g. fasting due to surgery, on parenteral nutrition).

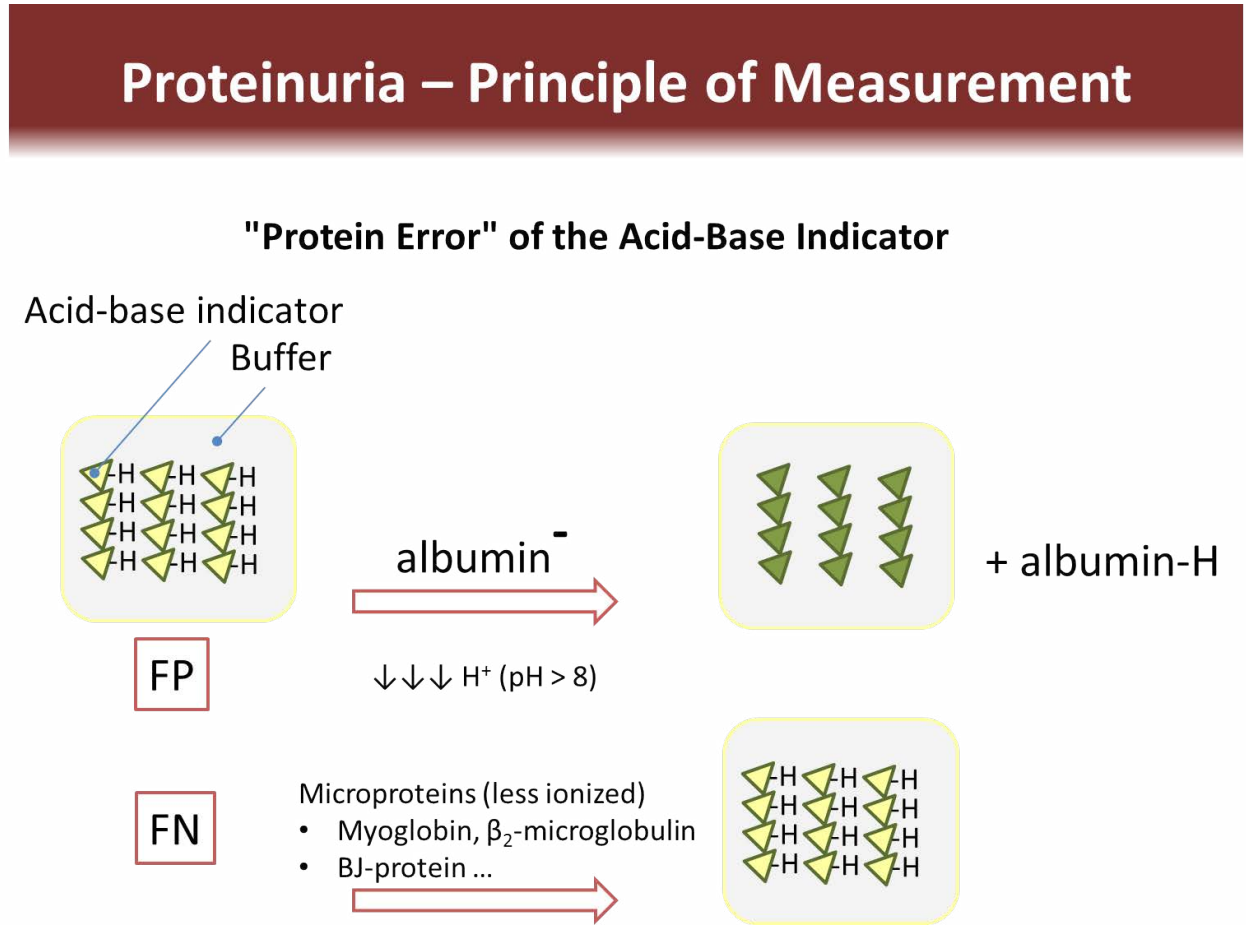


Figure 5.1. The principle of urine protein determinat ion using a diagnostic strip and potential false positive and false negative results

False positive results may be caused by strongly alkaline (pH > 8) and buffered urines. These can remove protons from the indicator (and change its colour) even in the absence of albumin. This can be solved by using another test (e.g. sulfosalicylic acid precipitation test) for strongly alkaline urines, or repeating the examination at a later time (e.g. in non-complicated urinary tract infections with a high pH of urine, it is reasonable to repeat the determination after antibiotic treatment of the uroinfection).

False negative results can be expected in most pre-renal and tubular proteinurias (a relatively low protein concentration in urine and a low indicator sensitivity), including the Bence Jones protein (free light chains in the urine in multiple myeloma). In addition, albumin in the microalbuminuria zone (30-150 mg/l) is usually not detected by standard test strips (refer to Diabetes for more information about microalbuminuria Diabetes/microalbuminuria).

There are also other ways of urine protein (albumin) determination using a test strip that do not suffer so much from pH interference (e.g. sulfosalicylic acid precipitation reaction or the newer chromogenic and immunochemical strip methods).

Depending on the location of the cause, proteinuria is usually divided ([10]) into the following types (see Figure 5.2):

- Pre-renal proteinuria ([11]) – caused by an increased supply of microprotein (which also penetrates a healthy glomerulus) to tubules that do not manage to take it up. This usually involves a low amount of protein (under 1 g per day). Examples of such proteins are β2-microglobulin, α1-microglobulin, acute phase reactants, Bence Jones protein, myoglobin and haemoglobin. The presence of these proteins does not primarily mean damaged renal function, and often it is not detected by the test strip at all. If Bence Jones protein is suspected, serum protein electrophoresis and urine immunofixation must be carried out;
- Renal proteinuria – the cause is kidney damage at the level of glomeruli or tubules. Glomerular proteinuria is caused by a higher permeation of the glomerulus ([12]) and occurs in many glomerulonephritides, diabetic nephropathy or renal amyloidosis;

[10] A classical laboratory test to determine the type of proteinuria is SDS-PAGE electrophoresis, where proteins are sorted out by size (thereby easily determining their origin).
[11] Sometimes also referred to as "overflow" proteinuria.
[12] Can be further subdivided into selective (albumin and/or transferrin penetrate the glomerular membrane) and non-selective (even large proteins such as immunoglobulins penetrate the glomerular memberane) glomerular proteinuria. The prognosis and response to therapy are usually better in diseases with selective glomerular proteinuria.

Tubular proteinuria is subject to damage to tubules (proximal tubules in particular), for example, during poisoning by heavy metals (Hg, Cd), as an adverse effect of some drugs (gentamicin, cyclosporin, cisplatin, lithium …) or in some viral infections ([13]);

- Post-renal proteinuria – presence of proteins from efferent urinary tract (e.g. α2-macroglobulin or IgM), the cause being inflammation or bleeding, or a pre-analytical error (e.g. menstrual blood, prostatic secretion, sperm).

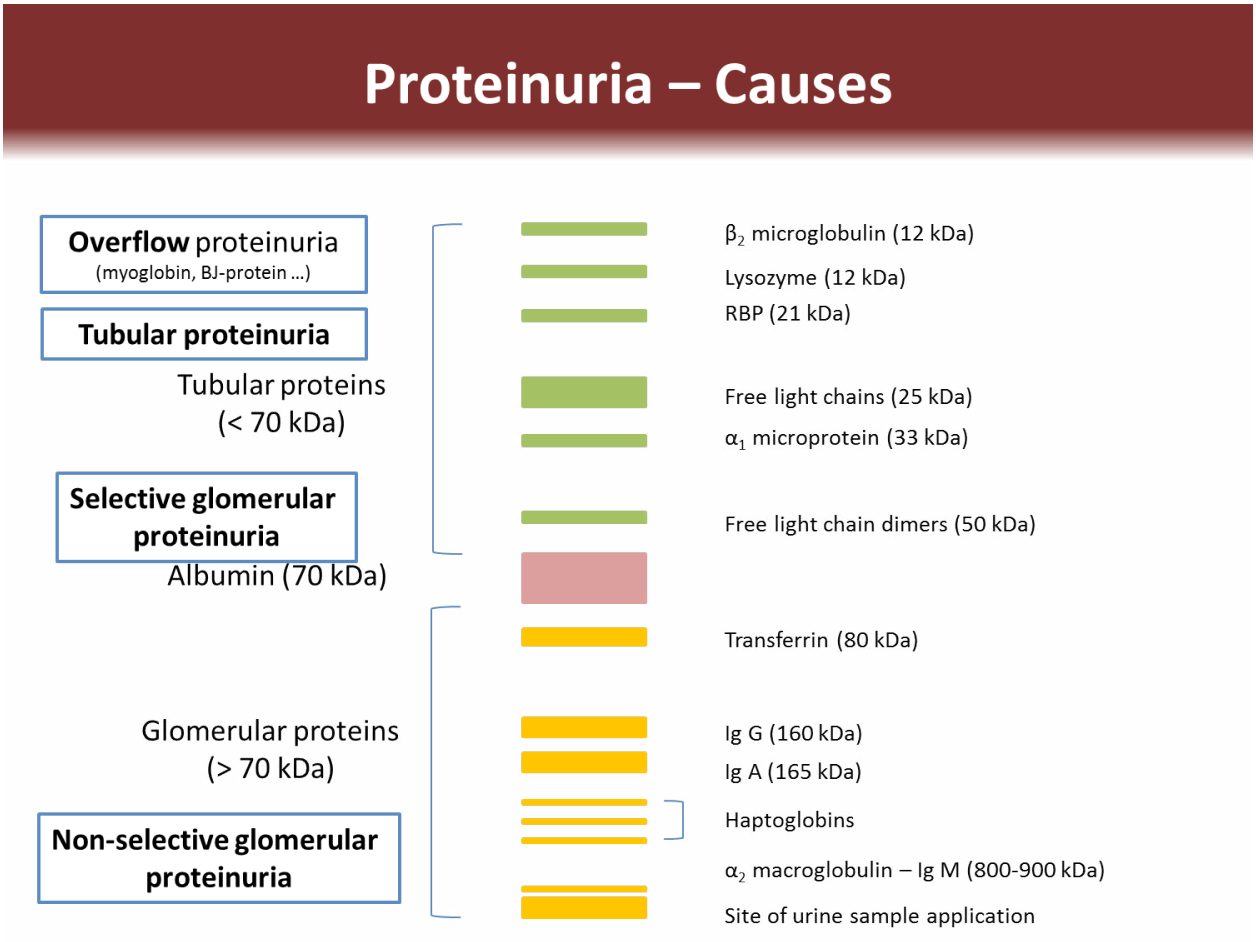


Figure 5.2. The distribution of proteins in urine by size during electrophoresis (takes place in polyacrylamide gel and proteins are usually sodium dodecyl sulfate-coated – SDS-PAGE)

Nephrotic syndrome

Nephrotic syndrome can be defined as a proteinuria capable of causing hypoalbuminaemia and oedemas. The amount of proteinuria may vary; usually it is > 3.5 g/24 hrs. Nephrotic syndrome may be caused by glomerulonephritis with minimal changes, proliferative glomerulonephritis or systemic lupus erythematosus.

U_protein/U_creatinine ratio

This ratio is evidently very useful and practical and can replace urine collection to quantify proteinuria. The ratio is usually expressed in mg (protein)/mmol (creatinine); numerically, the limit is 15 mg/mmol for proteinuria and 350 mg/mmol for nephrotic syndrome.

5.4.1.5. Blood

Principle of determination: Oxidation of chromogen (detection pigment) by haem (pseudoperoxidase activity of haem). The *haem* is detected, so the strip is sensitive both to **erythrocytes** (lysed in contact with the reagent pad) and **haemoglobin** or **myoglobin** (also contains haem). *False positive* results can be measured if the sample contains oxidizers (e.g. disinfectant residues from test tube decontamination); false negative results can be seen in classical tests in the presence of a high concentration of vitamin C ([14]).

[13] The determination of relatively stable microproteins (cystatin C or α_1-microglobulin) in the urine is currently used to diagnose this type of proteinuria.

[14] Note the resemblance of interferences with the determination of glucose. The basic principle in both is chro-

The following haematuria types are distinguished by intensity:

- Macroscopic haematuria (visible to the eye, urine is pink to red in colour, turbidity present – see Appearance of Urine for more details);
- Microscopic haematuria (detectable microscopically or chemically).
- The following haematuria types are distinguished by cause:
- Pre-renal haematuria – haemoglobin gets into the urine (due to the intravascular haemolysis of erythrocytes, e.g. in haemolytic anaemia, during incompatible transfusion), or myoglobin enters the urine (e.g. in extensive muscular trauma, burns, as a rare consequence of hypolipidaemic treatment by statins and fibrates), from the blood. The massive presence of each of the aforementioned proteins can cause acute renal failure (obstruction of tubules by precipitated protein) ([15]). Erythrocytes are not found in microscopic examination. Making the distinction between myoglobin or haemoglobin is based on patient history, the appearance of urine and serum (see Table 5.3) or other laboratory tests (high LD activity and immeasurable haptoglobin in haemolytic anaemia; high CK and AST in muscular damage; alternatively, both proteins can be distinguished in an immunoassay by determining their concentration);
- Renal sometimes also called glomerular haematuria, usually caused by glomerulonephritis. It can be distinguished from non-glomerular haematuria using phase contrast microscopy. This method visually highlights the edges of erythrocytes in the urine – dysmorphic erythrocytes ([16]) (erythrocytes with "thorny" projections, also called acanthocytes are typical – a consequence of passage through the glomerular sieve, see Figure 3) can be found in glomerular erythrocyturia. This type of haematuria is often accompanied by proteinuria and the presence of erythrocyte cylinders (see below);

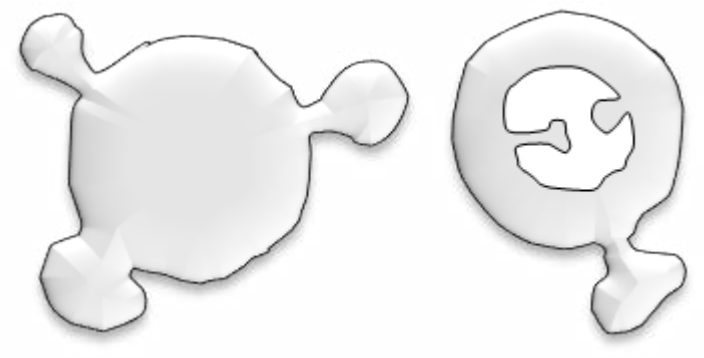


Figure 5.3. Dysmorphic erythrocytes (acanthocytes) – diagram. Acanthocytes are typical of the glomerular origin of erythrocytes

- Subrenal sometimes also called non-glomerular haematuria, usually caused by bleeding in the urinary tract during urinary tract inflammation, urolithiasis, or urinary tract or renal tumour. Traumatic bleeding after catheterization (especially with concurrent anticoagulation therapy) is quite common. A relatively minimal proteinuria is typical of this type of haematuria. Normally shaped (biconcave, biscuit-shaped) erythrocytes can be found in the phase contrast image.

Sometimes even exercise-induced haematuria (temporary, following strenuous exercise, after becoming chilled) or artificial haematuria (the patient intentionally adds blood into the urine sample) may occur.

5.4.1.6. Glucose

Principle of determination: Chromogen (detection pigment) oxidation by hydrogen peroxide produced by enzymatic decomposition of glucose (glucose is oxidized by glucose oxidase to form gluconolactone and H_2O_2). The determination is specific for glucose (other reducing carbohydrates such as galactose or fructose do not react) ([17]).

False positive results can be expected in the presence of oxidizers (e.g. some disinfectants used for test tube decontamination), false negative results are found in the presence of ascorbic acid (vitamin C) ([18]) or, most often, in samples

mogen oxidation (oxidizing agents will therefore cause false positive results and reducing agents false negative ones). Newer strips prevent vitamin C interference by an added agent that oxidizes the vitamin C, thereby rendering it ineffective.

[15] The basic precaution (acute renal failure prevention) in massive myoglobinuria and haemoglobinuria is increased fluid intake.

[16] Acanthocytes disintegrate very easily, so the sample has to be delivered to the lab very early – ideally, the sample should be taken there directly.w

[17] The reaction with the Benedict's reagent is still in use (CuSO4 in alkaline environment + heating -> Cu2O [precipitate colour from blue-green to brick red]). It is of major importance in galactosemia screening and this test is recommended in some countries in all children under 2 years of age.

[18] Some manufacturers add a substance able to oxidize ascorbate into the reaction mix, which prevents interfe-

delivered late (bacterial decomposition).

Indications: In particular, primary diagnosis of diabetes mellitus.

Glucose filtered by the glomerulus of a healthy person is almost completely absorbed in the proximal tubule. If concentration in blood exceeds about 10 mmol/l (known as the renal threshold for glucose), tubular cells are no longer able to take it up and glucose will appear in the urine. Therefore, the causes of glycosuria can be as follows:

- Most commonly, exceeding the renal threshold for glucose (as is the case in diabetic patients);
- Lower renal threshold (Fanconi syndrome, also in healthy individuals) – referred to as renal glycosuria;
- During pregnancy when glomerular filtration is increased (therefore also an increased amount of glucose passing through tubules) and the renal threshold for glucose may be lowered.

Glycosuria may appear even in a healthy person after a meal rich in carbohydrates, so it is advisable to use the second morning urine for (preliminary) diagnostic purposes ([19]) and the patient should be fasting until the sample collection (the first morning urine may contain postprandial glucose from the evening meal). It should be recalled that the sensitivity of this test is not ideal (glucose appears in the urine when concentration in the blood exceeds 10 mmol/l); the renal threshold for glucose can be highly individual and may vary in the course of diabetes mellitus. Glycosuria is therefore just a simple preliminary diagnostic test and is not suitable for the monitoring of diabetes compensation.

5.4.1.7. Ketone bodies

Principle of determination: Reaction of a keto group with nitroprusside (in an alkaline environment).

Ketone bodies are primarily produced by the incomplete oxidation of free fatty acids (from fats) in situations with a lack of glucose (e.g. fasting, diabetes mellitus – type 1 in particular, intensive physical strain, vomiting).Ketone bodies include β-hydroxybutyrate (BHB) and its oxidized form, *acetoacetate* (AcAc), the spontaneous decarboxylation of which generates acetone ([20]) - see Figure 5.4.

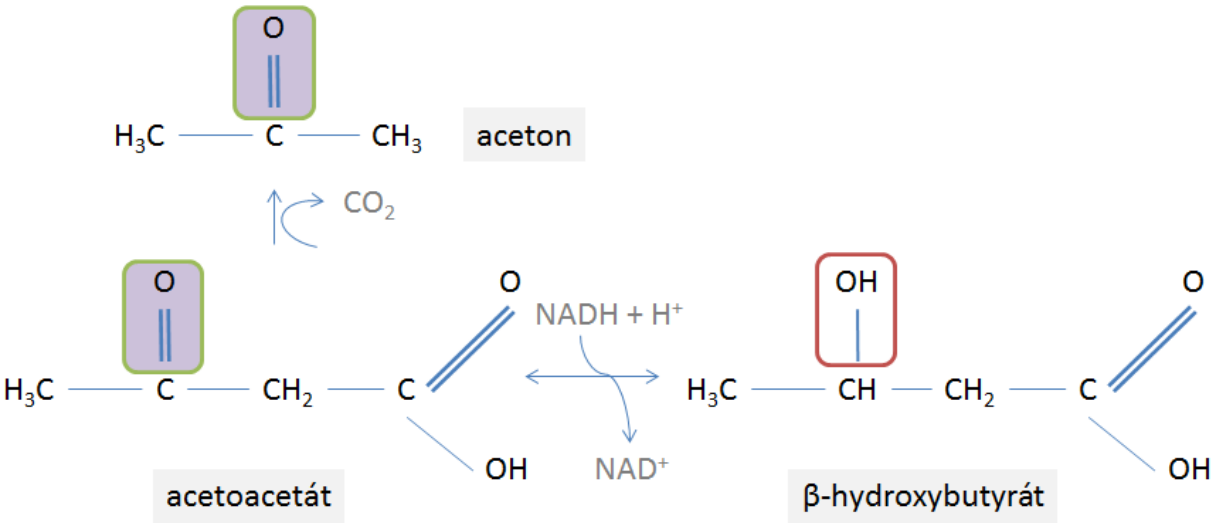


Figure 5.4. Ketone bodies. Note the lilac-shaded acetoacetate and acetone keto group (reacting with nitroprusside). In β-hydroxybutyrate the keto group is replaced by the hydroxy group, so it does not react with nitroprusside

In quantitative terms, β-hydroxybutyrate is predominant in urine (about 80%), while *acetoacetate* forms the rest (the acetone amount is negligible). Only the acetoacetate and acetone keto group reacts with nitroprusside. For example, a significant increase in the BHB/AcAc ratio may occur in severe diabetic ketoacidosis accompanied by hypoxia and shock, or in alcoholic ketoacidosis. As BHB does not react with nitroprusside, a quantitative underestimation of results may occur in this (exceptional) situation.

Indications: Primarily the monitoring of diabetes mellitus compensation, insulinotherapy in type 1 diabetes patients in particular. If ketone bodies occur in a type 1 diabetes patient's urine (on insulinotherapy), it means they are lacking insulin and the insulin dose has to be adjusted.

---

rence with glucose determination.

[19] The definitive diagnosis of diabetes mellitus is established from the plasma (see Diabetes Mellitus).

[20] More information about ketone body pathophysiology can be found in Acid-Base_Balance_Preparation -> production of H^+

5.4.1.8. Bilirubin, urobilinogen

Conjugated bilirubin in the blood (as opposed to non-conjugated bilirubin bound on albumin) freely penetrates into the urine, where it is detected in obstructive jaundice or liver damage, for example. Urobilinogen appears in the urine in liver damage or haemolytic anaemia, for example. Refer to the chapter Liver for more information.

5.5. Microscopic Examination of Urine

The examination of urine sediment supplements the information obtained from the chemical examination of urine using a test strip. Every laboratory has its own algorithm for sample processing and urine sediment examination. The chemical examination is usually for 'screening' purposes, while the urine sediment examination follows only if the result is pathological (mostly in the case of positive leukocytes, nitrites, blood or proteinuria). In most cases, the examination takes place in automatic analyzers on mixed, non-condensed urine specimens. The principle of determination is either a software analysis of the digital microscopic image taken by a digital camera, or flow cytometry (identification of elements is based on the measurement of impedance, diffusion of light and fluorescence). When in doubt however, decisions are taken based on a classical microscopic examination ([21]) made following standard condensation by centrifuging (e.g. centrifuge 12 ml of urine for 5 minutes at 400 RCF, no braking; then pipette 11 ml, resuspend and apply to the slide) ([22]). The examination can be made on the native sample or following staining (to highlight nuclei, cell membranes and other structures). Results are issued as the number of elements per µl.

Correct sample collection and timely processing are very important – cells and cylinders fall apart very quickly, especially in hypotonic and alkaline urine. Urine sediment should be processed *within 1 hour of collection*.

5.5.1. Composition of Urine Sediment

The urine sediment of a healthy person can contain many shaped structures. These also include a small amount of erythrocytes (< 5/µl), leukocytes (<10/µl), some epithelial cells (mainly squamous epithelium), hyaline cylinders or various crystals. In most cases, just a few squamous epithelium cells and phlegm can be found in normal urine sediment.

5.5.1.1. Erythrocytes

Above all, a urine sediment examination confirms whether the positive result of a chemical blood test strip is accompanied by erythrocyturia or not, or whether it is myoglobinuria or haemoglobinuria. In addition, in the urine sediment we can distinguish erythrocytes of glomerular and non-glomerular origin (*Figure 5.5*, also see above).

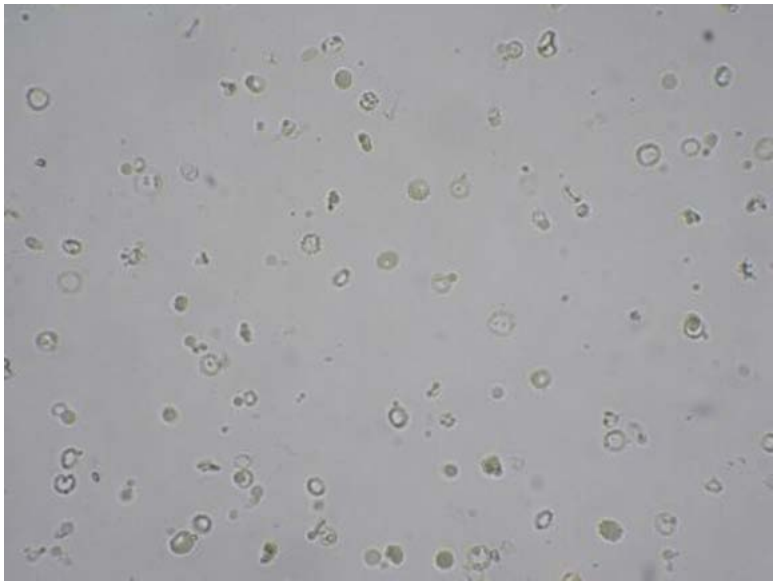


Figure 5.5. Dysmorphic erythrocytes

[21] Alternatively, use phase contrast microscopy (dysmorphic erythrocytes, identification of cylinders and some cells).

[22] There are commercially available slides that standardize this phase (volume of sample applied).

5.5.1.2. Leukocytes

In general, leukocytes in the urine sediment signal urinary tract inflammation. The etiology can be infectious (e.g. pyelonephritis, cystitis) or non-infectious (e.g. autoimmune damage). *Neutrophils* are usually most abundant; the diagnosis of drug-induced interstitial nephritis may be supported by the finding of *eosinophils* (special staining is required).

5.5.1.3. Epithelia

The following types can be distinguished, depending on the origin of epithelia:

- Squamous (from the vagina and the lower 1/3 of the urethra) – large cells with a centrally located nucleus (about the size of an erythrocyte); a physiological part of urine;
- Transitional (renal pelvis lining through calyces, ureters, urinary bladder to the proximal part of the urethra) – smaller than squamous epithelia, centrally located nucleus, different shapes (e.g. spherical, "tailed"); a small quantity is physiological, a large quantity after catheterization and other invasive interventions; atypical cells in tumours or viral infections;
- Tubular cells (come from tubules) – different shape depending on the origin (triangular, spherical [unlike the transitional epithelia, the nucleus has an eccentric position], cuboidal) – always pathological, a sign of tubular necrosis (e.g. heavy metal poisoning, haemoglobin and myoglobinuria, nephrotoxical drugs, viral infections [HBV], acute transplant rejection).

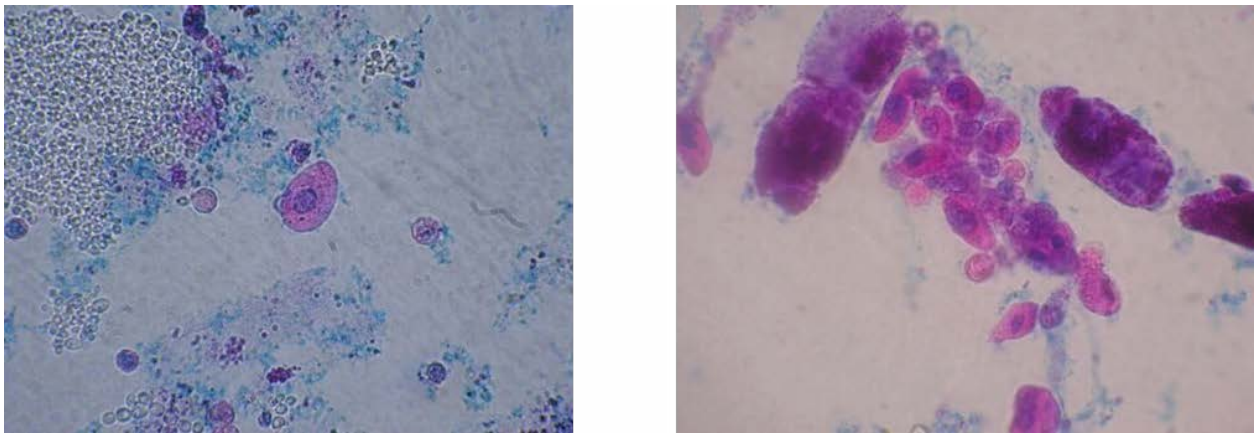


Figure 5.6. A transitional epithelial cell (on the left) and tubular cells (on the right)

5.5.1.4. Bacteria and Yeast

The presence of bacteria along with leukocyturia is usually a sign of urinary tract infection (*Figure 2*). The occurrence of yeast (*Figure 5.7*) is more common in diabetic patients (glycosuria + acidic urine), immunosuppressed patients and patients with vaginal moniliasis. Isolated bacteriuria and the occurrence of yeast (particularly with non-sterile collection and a late sample delivery) do not have diagnostic significance.

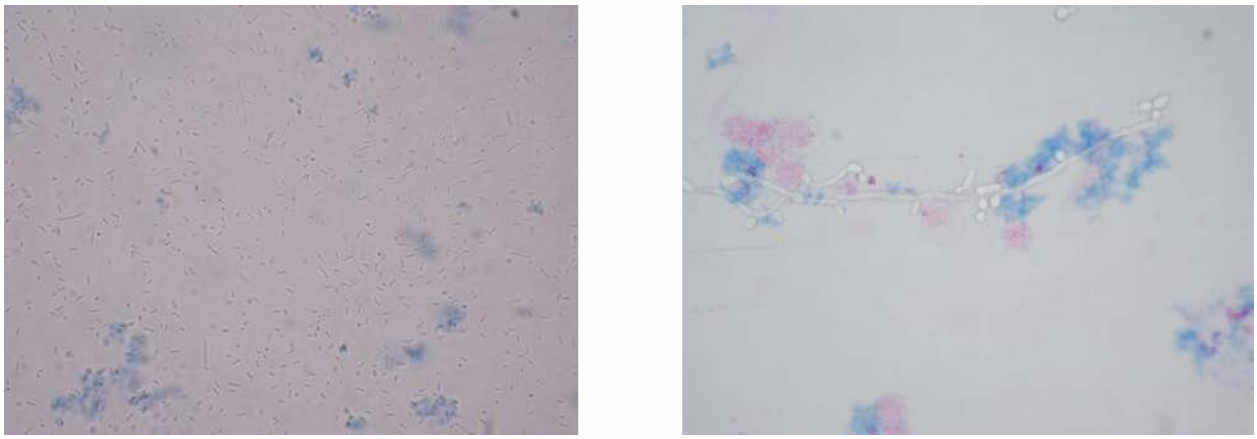


Figure 5.7. Flood of bacteria (on the left – resembling "scattered tea leaves") and yeast (on the right). Notice the budding yeast and mycelia that help distinguish them from erythrocytes

5.5.1.5. Parasites

The most common parasite is *Trichomonas vaginalis*, characterized by 3 flagella and an undulating membrane (*Figure 5.8*). This parasite causes vaginal inflammations.

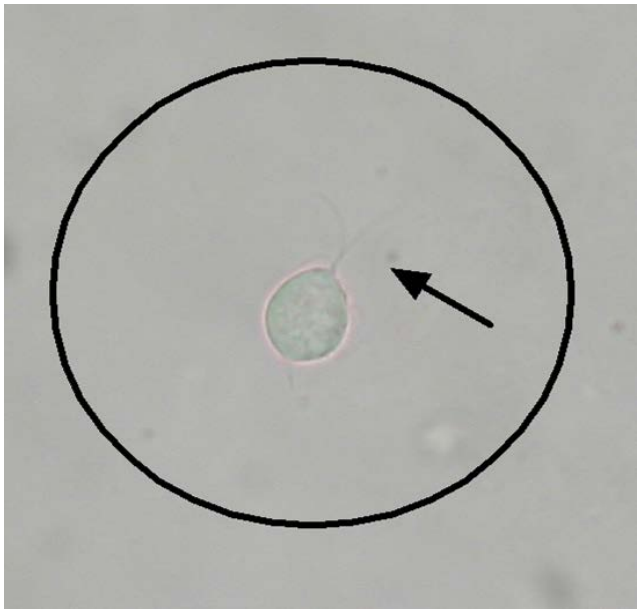


Figure 5.8. Trichomonas vaginalis in an unstained preparation. The arrow points to 3 flagella

5.5.1.6. Mucus

Mucus is a protein from genito-urinary tract glands or tubules (Tamm-Horsfall protein). It is common in physiological urine and has no diagnostic significance (although it can be confused for hyaline cylinders – see below).

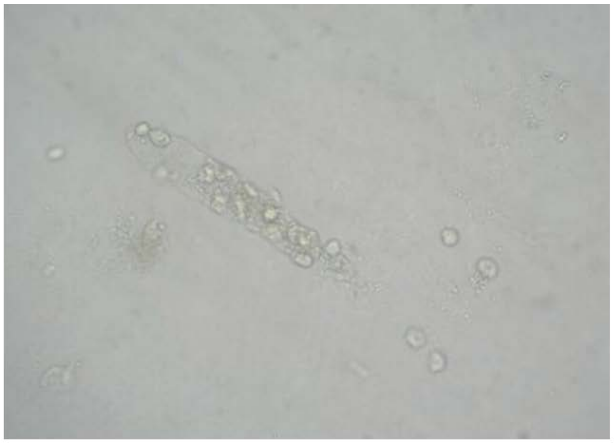
5.5.1.7. Cylinders

Cylinders form in distal parts of the nephron. They are composed of the Tamm-Horsfall protein secreted by tubular cells and other parts which differ according to the specific pathology (e.g. erythrocytes, leukocytes). Their shape (width) copies the size of tubules – which is why wide cylinders always come from pathologically dilated tubules.

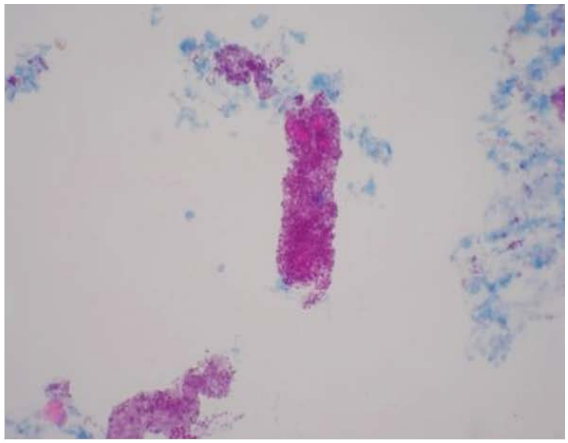
Some cylinders contain *cell* structures:

- Erythrocytic (in erythrocyturia of a glomerular origin);
- Leukocytic (in pyelonephritis);
- Epithelial (containing tubular cells – e.g. in toxic damage by heavy metals, poisons or drugs; transplant rejection), it is sometimes difficult to distinguish them from leukocytes;
- Bacterial cylinders, sometimes also contain leukocytes, can be found in pyelonephritis.

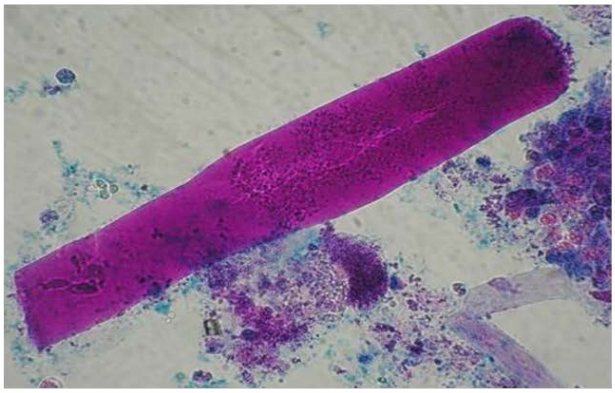
We can also distinguish between **hyaline** cylinders (most common; small amounts can be physiological, large amounts point non-specifically to renal pathology – e.g. acute glomerulonephritis, pyelonephritis, chronic renal insufficiency), **granulated** cylinders (come from renal tubular cell lysosomes or from disintegrated leukocytes; the diagnostic significance is disputable – depending on further findings they can indicate tubular damage, leukocyturia or no pathology at all), and **waxy** cylinders ("chronic renal failure cylinders", usually found with other cylinders). Simply put, a hyaline cylinder in the tubule becomes granulated over time (with a progressive urinary stasis in the tubule) and finally waxy.



Leukocytic cylinder (unstained preparation)



Granulocytic cylinder (... staining)

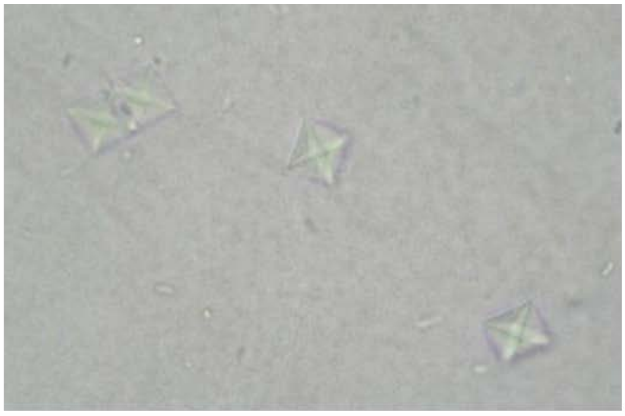


Waxy cylinder

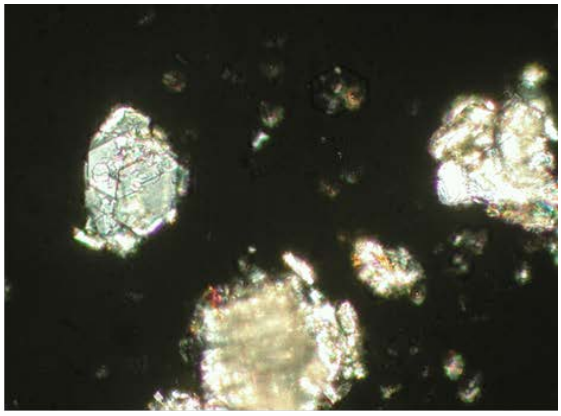
Figure 5.9. Cylinders

5.5.1.8. Crystals

A finding of crystals in the urine is not usually pathological, or indicative of the type of urolothiasis ([23]) (if present). An exception are cystine crystals that cannot occur in a healthy person's urine. A flood of oxalate crystals may be associated with ethylene glycol poisoning (along with metabolic acidosis, see Toxicology).



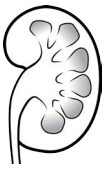
Oxalate crystals ("envelopes")



Hexagonal cystine crystals

Figure 5.10. Crystals.

[23] The most common stone is calcium oxalate; nevertheless, oxalate crystals are common in the urine of completely healthy people (dietary sources of oxalates – e.g. spinach, rhubarb).



CHAPTER 6

6. Kidney Function Tests

Authors: MUDr. Daniel Rajdl, Ph.D.

Reviewer: MUDr. Věra Ploticová, Mgr. Radka Šigutová

Summary:

Kidneys are metabolically very active organs that play a major role in maintaining an optimum internal environment. Some kidney functions such as glomerular filtration and certain tubule functions are also tested in regular clinical practice.

Glomerular filtration „cleans out waste substances from the blood“. Blood is constantly pressed through filter pores of the glomerular membrane (in simple terms similar to a sieve) and only small molecule substances make it to the other („urine”) side[1]. The filtered fluid („plasma ultrafiltrate” or primary urine) is then gradually exposed to the effect of renal tubules, which return necessary substances (e.g. water and minerals) to the blood, and excrete waste substances (e.g. hydrogen ions) into the urine.

6.1. Glomerular Filtration Tests (GF)

Whenever „kidney tests” are discussed, what is often being referred to (inaccurately), are methods to determine glomerular filtration. The level of glomerular filtration is used to adjust the dose of some drugs excreted by the kidneys (e.g. the antibiotic gentamicin or low-molecular heparins), the administration of some substances may be also contraindicated (usually nephrotoxic substances such as X-ray radiocontrast agents or the antitumour drug cisplatin), and it is used to evaluate the development of renal insufficiency, or indicate an extracorporeal cleaning method („dialysis”).

Glomerular filtration usually decreases with age (after maturity is reached), increases by about 50% in pregnancy, and the reference range also depends on the gender.

6.1.1. Serum Creatinine (S-Crea) and Derived Calculations

Creatinine is generated in muscles from creatine and creatine phosphate (molecules providing the muscle with energy). Creatinine has a small molecule, so it passes through the glomerular „sieve” completely. It is partially secreted, nonetheless, in tubules (< 10% in a healthy person) – this „error”[2] is usually disregarded, but must be taken into account at more advanced stages of renal insufficiency (the percentage excreted by tubular secretion into the urine grows with the growing serum creatinine concentration).

The greatest limit for using creatinine as a GF marker is the **dependence on muscle mass**. Creatinine production is different in a „bodybuilder” and a cachectic patient; muscle mass physiologically decreases with age, women have smaller muscle mass than men, etc. Therefore, GF estimates based on serum creatinine in patients not having an „average” muscle mass are subject to great error[3]. The following methods can be used to estimate GF using serum creatinine.

Directly by *serum creatinine concentration* – the problem of this approach consists in the impracticality of its evaluation against the reference range, which depends on age and gender. It is relatively difficult for a physician to imagine

1 A guidance limit for the molecule size is 70 kDa – if the molecules are larger, they do not pass through a healthy glomerular membrane. Nevertheless, filtration also depends on the charge of the filtered molecule and other factors (see Introduction to the Topic).

2 An ideal marker for glomerular filtration is fully filtered by the glomerulus and is not re-absorbed or secreted in the tubules. Besides which, of course, it is inexpensive, easily and reliably measurable…

3 Proportionate to age and gender. Disproportionate (small) muscle mass can be found in the undernourished, people confined to bed (e.g. following ictus), or people with serious conditions (operation, infection, sepsis, burns). In such cases, GF estimated from serum creatinine will be greater than the actual value (false overestimation). In contrast, sportspeople (strength-based sports in particular) have disproportionately large muscle mass.

quantitatively the relationship between the GF and current serum creatinine level in the given patient.

Using different *calculations*, usually considering age and gender (or race – African Americans often have a larger muscle mass) in addition to serum creatinine; the result is expressed in regular units for GF (mL/s). Quantitative interpretation is then easier. The **MDRD** (Modification of Diet in Renal Disease) formula is an example of such calculation. Currently the most widespread MDRD is the four-variable version that includes serum creatinine, age, gender and race (only for African Americans). The MDRD formula is recommended as a population screening method to capture decreased GF. The formula cannot be used for children and pregnant women[4].

Another problem (in addition to the dependence on muscle mass) of creatinine is the method of determination. The principle of determination is the reaction with picric acid in a strongly alkaline environment[5]. Relatively many interfering substances such as acetoacetate, protein or glucose are known, and usually lead to falsely higher results. These interferences can be minimized (at least partially) using different procedures. There is an effort to establish an *enzymatic method* for determining creatinine, which does not suffer from these interferences. It is certain that the results of this method are systematically lower (compared with the Jaffé method), but it is unclear whether this is important for clinical practice[6].

6.1.2. Serum Cystatin C

Cystatin C is a low-molecular protein (13 kD) and a physiological inhibitor of cysteine proteases. It is therefore an important part of the protease - antiprotease balance, which is why a constant amount of cystatin C is usually expressed by all nucleated cells. It is also a microprotein, which freely penetrates the glomerular membrane and is fully re-absorbed and degraded by proximal tubule cells. If GF decreases, it is retained (like the other microproteins[7]) and its serum concentration increases in correspondence with the decrease in GF. There are also formulas expressing the serum cystatin C concentration in GF units (mL/s).

Advantages of using serum cystatin C:

- Earlier detection of the initial forms of glomerular injury (the reaction in acute renal failure is 1 - 2 days earlier than S_Crea);
- No need to collect urine, no gender differences;
- No analytical interferences.

However, the assumption that it is produced in constant amounts is not absolutely true: it is known that corticotherapy, increased thyroid function and some tumours increase cystatin C production; similarly, diabetes mellitus and gender have some effect on cystatin C production. It is also certain, though, that this effect is much lower than S_Crea.

A GF estimate based on serum cystatin C is favoured mainly for children and pregnant women. In hospitalized patients, some recommend measuring the GF estimate based on both MDRD and serum cystatin C[8].

6.1.3. Creatinine Clearance

Sometimes it is also called endogenous creatinine clearance[9]. All methods based on measuring the clearance[10] of a particular substance try to approximate the „ideal", i.e. the substance will be fully filtered in glomeruli, will not be subject to tubular secretion or re-absorption, and will be easy to measure (and the procedure will not be burdenso-

4 There are many other formulas:

the Schwarz formula = 0.6 * height (cm)/ * S_Crea (µmol/L)

is used for children [factor 0.6 is used only as a guide – age and gender must be considered for a more accurate calculation], Lund-Malmö, CKD-EPI and other equations have been used recently for children from 1 year as wells as adults, etc. All of them are based on S_Crea and usually only differ in terms of the populations for which they have been validated (and are valid for). Relationships based on S_Crea are not used for pregnant women, but other suitable options are available: serum cystatin C and creatinine clearance.

5 Known as the Jaffé reaction. Substances reacting with the agent are also called „Jaffé-positive chromogens".

6 It is not clear what percentage of people would be incorrectly classified as having reduced GFR. The positive bias of creatinine assay using the Jaffé method, which can be found particularly in serum concentrations < 140 µmol/L, can be minimized by recalibration to the ID-MS reference method. Muscle mass and tubular excretion interferences are still probably the biggest source of errors.

7 For example, β_2-microglobulin, amylase or myoglobin.

8 One can judge which result is closer to reality based on the discrepancy between the estimates and the patient's history. For example, the MDRD-based estimate will be significantly overestimated in a cachectic person, and the GF estimate from serum cystatin C underestimated in a person on corticotherapy (see also below).

9 To emphasize that it is a substance generated by the body, not supplied externally (such as inulin).

10 Clearance is the virtual volume of plasma cleared of the substance per unit of time (with units such as mL/s).

me for the patient). Creatinine is fully filtered, although it is also partially secreted in tubules (see the chapter Serum Creatinine (S_Crea) and Derived Calculations). Theoretically, the creatinine clearance formula can be derived from[11]:

$$GF * S\_Crea = V * U\_Crea$$
$$\downarrow$$
$$\mathbf{GF = V * U\_Crea/S\_Crea}$$

[values are converted to a standard surface area of 1.73 m^2]

Creatinine clearance is used mainly for early diagnosis of reduced GF or GF tests in pregnancy. Serum creatinine and serum creatinine-based calculations should be used at the later phases of reduced GF. The greatest **disadvantage** of creatinine clearance is that collecting urine is required, usually for 24 hours. Collection-induced errors (incomplete collection) are very common and degrade the result of the test[12]. This is the reason why methods requiring urine collection are generally being abandoned.

6.1.4. Practical Approach to GF Evaluation

Procedures based on S_Crea as well as estimates from serum cystatin C are suitable for use in hospital practice. Two independent estimates are then available, presenting 3 possible situations:

Both estimates are (roughly) the same – the estimate is considered a real value;

The cystatin C-based GF estimate << MDRD

The discrepancy is caused by *reduced muscle mass* in patients confined to bed, suffering from chronic diseases or catabolic conditions. MDRD estimates muscle tissue only by age and gender (or race), and is unable to take into account the muscle loss mentioned above. Reduced creatinine production thus leads to a relatively lower S_Crea level, from which MDRD derives a falsely high GF level.

S_Crea-based estimates are considerably overestimated; cystatin C-based estimates should be used.

The patient is taking *corticosteroids or has hyperthyroidism* – cystatin C production is thus increased and the serum concentration is higher than the real GF. The cystatin C-based GF estimate is therefore underestimated and the S_Crea-based estimate should be used.

The cystatin C-based GF estimate >> MDRD – this is caused either by the patient having an unusually large muscle mass (higher than expected creatinine production → higher S_Crea → falsely lower GF estimate from MDRD), or GF is in the zone over 1.5 mL/s. S_Crea-based estimates are not validated for this zone and their accuracy decreases.

6.1.4.1. Case Report 1

An 86-year-old woman was taken by ambulance from the nursing home, where she had been confined to bed for a long time due to septic fevers (the 3rd day repeatedly over 40 °C). Chronic medication: antihypertensive (perindopril), furosemide; acute medication: Augmentin (empirically). Following admission, the following serum values were measured (selection of pathological results): S_Crea 160 µmol/L, urea 26 mmol/L, Na 131 mmol/L, K 5.8 mmol/L. Urine values were: U_ery 96/µL, U_Leuco 670/µL, nitrites +++, pH 8.6.

Formulate a hypothesis as to the diagnosis, explain the probable cause(s) of the measured pathological results, and suggest additional tests.

It is probable that the patient has considerably reduced glomerular filtration (MDRD 0.45 mL/s). The dispropor-

11 "What enters the glomerule equals what is excreted." We know that creatinine is excreted by tubules in the urine, so its concentration in the urine is higher than ideal. The creatinine clearance-based GF estimate is therefore slightly overestimated. This overestimation grows with a growing S_Crea concentration, and creatinine clearances over 180 µmol/L should not be used (it is also pointless).

12 The patient should be instructed to urinate as usual in the morning after waking up, and then to collect urine in the collection container(s) the rest of the day, before collecting the last sample, also in the container, the following morning. The entire volume of collected urine should be mixed (in one container), the volume should be measured (accurate to 10 mL) and a representative sample should be taken. It is clear, and has been shown in practice, that this procedure leads to frequent errors.

tionally high urea is indicative of a prerenal (and acute) component. Signs of urinary tract infection (UTI) in the urine sediment are evident (high pH, leucocytes, positive nitrites). Given the history, it is obvious that the patient has a considerably reduced muscle mass volume (prolonged immobilization, current catabolic condition). Therefore, a cystatin C test should be indicated. The cystatin-based GF estimate was 0.25 mL/s (the estimate from MDRD was probably considerably overestimated). This condition was caused by a UTI connected with loss of appetite and insufficient fluid supply, in combination with a persistent use of furosemide. This led to acute prerenal kidney failure, water retention (hyponatraemia) and insufficient potassium excretion (hyperkalaemia)[13]. Following antibiotic therapy of the UTI (based on the microbiology assay for the pathogen and pathogen sensitivity to antibiotics) and appropriate hydration, the condition gradually improved and the patient was discharged with a cystatin C-based GF of 0.75 mL/s.

6.2. Tubule Function Test

The main function of tubules is to absorb and secrete various substances such as Na^+, K^+, Cl^-, H^+, HCO_3^- and water – refer to the introduction to clinical biochemistry in nephrology for details. Metabolically, the most active is the proximal tubule, but the other parts of the nephron are also important in this regard.

6.2.1. Secretion and Reabsorption: Fractional Excretion of Low-Molecular Substances

Fractional excretion (FE) is that part of the total amount of filtered substance, which can be found in the definitive urine. For example, the kidneys filter about 180 L of primary urine (plasma ultrafiltrate) a day, while about 1.8 L a day is urinated. Considering that urine (primary and definitive) is mainly water, then the FE of water is 1.8 of 180 = 1%.

The derivation of the FE formula can be demonstrated in the example of the fractional excretion of sodium:

Filtered amount of sodium in millilitres per second is

glomerular filtration [GF; mL/s] * plasma sodium concentration [P_{Na}; mmol/L]

Excreted amount of sodium into the urine in millilitres per second is

urine volume [V; mL/s] * sodium concentration in urine [U_{Na}; mmol/L]

$$FE = \frac{excreted\ amount}{filtred\ amount} = \frac{V * U_{Na}}{GF * P_{Na}}$$

if creatinine clearance is substituted for GF (U_{crea} * V / P_{crea}).

The formula shows that 24-hour collected urine is not necessary for the FE calculation (volumes V cancel each other out); a short collection period (e.g. 4-hr) is optimal to make up for short-term effects (fluid intake, etc.). Any low-molecular substance the FE of which should be calculated can be substituted for the serum and urine sodium concentration (e.g. potassium, see below).

Therefore, the **FE of water** is calculated as a plasma-to-urine creatinine concentration ratio. (U_x and P_x are not substituted from the formula since the proportion of water [water "concentration"] in the primary and definitive urines is roughly the same).

It is evident that although FE is a sign of tubule functions, it is not quite independent of the GF. To avoid significant change in the internal environment, a decrease in GF must lead to a compensatory increase in FE[14].

Examples of using FE:

- FE of water

Reference range (RR): 1 - 2%; max.: 35%

Increase: increased fluid intake, diuretics, diabetes insipidus, tubulopathy, tubular compensation for decreased GF

13 The ACEI therapy could also have contributed to the hyponatraemia and hyperkalaemia.

14 Otherwise a 50% decrease in the GF (from 180 L/day to 90 L/day) would lead (with a maintained FE of water = 1%) to the excretion of half the amount of water in the definitive urine (a drop in diuresis from 1.8 L to 0.9 L) and water would accumulate in the body. The pathophysiological basis of a decrease in the GE is a drop in the number of functional nephrons; these remaining (residual) nephrons then have to make a greater effort to excrete the same amount of waste matter from a lower number of nephrons.

Decrease: dehydration, etc.

- FE_{Na}

RR: 0.4 - 1.2%; max.: 30 - 40%

Increase: diuretics, increased intake, tubulopathy, tubular compensation of decreased GF, hypoaldosteronism, some drugs (renin axis-blocking drugs – angiotensin-converting-enzyme inhibitors [ACEI], spironolacton, etc.)

Decrease: extrarenal losses, lack of Na, hyperaldosteronism

- FE_K

RR: 4 - 19%; max.: 150 - 200%[15]

Increase: diuretics, increased intake, hyperaldosteronism, catabolism, tubulopathy, tubular compensation for decreased GF

Decrease: extrarenal losses, lack of K, hypoaldosteronism, drugs (potassium-sparing diuretics, ACEI), anabolism

- FE of bicarbonate – increased in the diagnosis of proximal renal tubular acidosis (type II).
- FE of amylase – if < 1%, macroamylasemia suspected see chapter Laboratory diagnostics in gastroenterology.

6.2.2. Renal Concentrating Ability

The main physiological regulation of water absorption is due to the action of the antidiuretic hormone (ADH). It affects receptors opening aquaporins (water channels) mostly in the distal tubule and collecting ducts. The main stimulus for ADH secretion in the hypothalamus is increased serum osmolality (but it also reacts to a decrease in pressure and other stimuli such as vomiting; see preparatory materials). In simplified terms, the following ADH secretion disorders are distinguished:

- Diabetes insipidus (DI)
- Insufficient ADH secretion (*central* DI – hypothalamus or neurohypophysis disorder)
- Insufficient ADH effect (*nephrogenic* DI – tubule function disorder; congenital or acquired tubulopathies such as toxicity of metals Cd, Hg, Li, Pt + other substances – furosemide, gentamicin, RTG contrast agents, and the effect of severe ion imbalances – hypokalaemia and hypercalcaemia)
- SIADH – syndrome of inappropriate ADH secretion: hypothalamus produces more ADH than the amount corresponding to the current serum osmolality. This leads to an increased absorption of water in the kidneys and "dilution" manifested as hyponatraemia. This may be due to a brain disease or ectopic production in small-cell lung cancer, TBC or pancreatic cancer.

6.2.2.1. Electrolyte-Free Water Clearance (EWC)

Electrolyte-free water clearance (EWC) is a parameter sensitively evaluating ADH-renal axis disorders (whether or not the volume of urine corresponds to effective osmolality and expected ADH secretion). It is the difference between the volume of urine [V; mL/s] and clearance[16] of the major osmotically active substances (referred to as electrolyte clearance, Cl_{El}; mL/s):

$$EWC = V - Cl_{El} = V - \frac{(U_{Na} + U_K) * V}{S_{Na} + S_K}$$

If other effective solutes (causing the transfer of water between intracellular and extracellular spaces, e.g. glucose or mannitol[17]) are also present in the plasma and urine, this should be considered in the calculation.

Interpretation of EWC:

15 FE greater than 100% means that more substance is excreted than filtered in the primary urine (caused by active secretion in the tubules).

16 According to definition, clearance is the virtual volume of the given substance (Na and K in this case) cleared in a unit of time.

17 In contrast, urea or ethanol pass freely between the ICF and the ECF, do not cause transfers between these spaces and therefore are not regarded as effective solutes (they do not affect effective osmolality).

Increased effective osmolality (e.g. hypernatraemia) leads to an increase in ADH secretion in the hypothalamus and an increased absorption of water in the kidneys (the volume of urine V decreases). EWC should be < 0.005 mL/s. If it is higher, the hypothalamus excretes less ADH than it should with the given effective osmolality (e.g. diabetes insipidus).

Decreased effective osmolality (e.g. hyponatraemia) leads to a decrease in ADH secretion in the hypothalamus and increased losses of water in the kidneys (the volume of urine V increases). EWC should be > 0.116 mL/s. If it is lower, the hypothalamus excretes more ADH than it should with the given effective osmolality (e.g. SIADH).

6.2.2.2. Concentration Test

Renal concentrating ability is usually determined for the differential diagnosis of unexplained *polyuria*[18]. If renal concentrating ability is to be tested, the following options are available:

Deprive the patient of fluids;

Administer a synthetic ADH analogue (1-deamino-8-D-arginine vasopressin = DDAVP).

Always monitor urine osmolality (U_Osm) as a reflection of the kidney's capacity to absorb water. A combination of the two approaches is usually chosen, e.g. depriving the patient of fluids overnight (normal situation: patient sleeps = does not drink) and U_Osm is measured in the morning. If it is higher than 800 mmol/kg, no further testing is required (the renal concentrating ability is good). Where it is lower, DDAVP can be administered (intranasally) and urine osmolality monitored (4 samples every hour). IF U_Osm reaches the relevant value (in tables for the specific age), the reaction of kidneys to the ADH is good, and the cause of polyuria is then insufficient ADH secretion (disorder in the hypothalamus or the pituitary). If the reaction of the kidneys is insufficient, it is probably due to nephrogenic DI.

6.2.3. Acidification and Alkalinization of Urine

This renal ability is tested only in exceptional circumstances see also Introduction to Acid-Base Balance Disorders; the diagnosis of renal tubular acidosis (RTA) is practically the only indication. In general, if hyperchloraemic acidosis not due to bicarbonate losses from the digestive tract is found, RTA should be considered. However, one should first make sure that the patient does not have renal failure, hyperparathyroidism, and is not receiving saline (frequent causes of hyperchloraemic acidosis). Kalaemia may then be useful in determining the type of RTA (*Table 1*). **The Acidification test** is used for the diagnosis of type I renal tubular acidosis (*Table 1*). The principle of the test is the administration of an acidifying substance (usually ammonium chloride; this cannot be administered to patients with severe hepatic function disorder or reduced GF) and subsequent urine pH monitoring (8 times every hour). If urine pH in any urine sample falls below 5.5 (even before the test), the acidification ability of the kidneys is intact (and the diagnosis of RTA is improbable).

RTA I and II are usually treated by administering bicarbonate and potassium, but RTA II requires significantly higher doses of bicarbonate.

Type	Cause	Comments	K^{+1}
I (distal)	Poor excretion of H^+ in distal tubule	Diagnosis by acidification test, phosphate urolithiasis	↓
II (proximal)	Poor absorption of HCO_3^- in proximal tubule	Diagnosis of the FE of bicarbonate (alkalinization test[2]); Often combined in Fanconi syndrome (glucose, amino acids and phosphates are not absorbed in the proximal tubule)	↓
III	Combination of I and II (rare; in premature newborn infants)		↓
IV	Poor excretion of H^+ and K^+ in distal tubule	Also in hypoaldosteronism	↑

Table 6.1. Renal tubular acidoses

18 For example, trying to find the cause of polyuria in a decompensated diabetic is pointless.

6.3. Acute Renal Failure

Acute renal failure is characterized by the fast progress (hours to several days) of the worsening, or loss of, renal function. Since the transition between the worsening and loss of function (failure) is continuous, modern literature uses the term **acute kidney injury (AKI)**. AKI is very common[19], particularly in patients in intensive care; it often occurs therefore with other diseases (cardiac failure, sepsis, etc.) and frequently accompanies chronic renal insufficiency. An important difference (compared to chronic renal failure) is that AKI is potentially reversible (although certainly not in every case).

Traditionally, acute renal failure is divided into:

- *Prerenal* – most common; usually caused by cardiac failure, hypovolaemia (loss of blood and fluids) or hypotension (sepsis, drugs), less often with renal artery occlusion → low perfusion pressure (often potentiated by the medication of ACEI or non-steroid anti-inflammatory drugs [NSAID]);
- *Renal* – may be caused, for example, by glomerulonephritis or nephrotoxins (aminoglycosides, heavy metals, cisplatin, Li, NSAID);
- *Postrenal* – bilateral obstruction (unless there is only one functioning kidney) of urine drainage, most commonly due to prostate, cervical, urinary bladder or retroperitoneal tumours; or bilateral nephrolithiasis see Case report 2.

In prerenal and postrenal failure, kidney injury (renal failure) will also always gradually develop, but tubule functions are maintained at the beginning (concentrating ability, tubular reabsorption), which helps to distinguish it from a renal cause (*Table 26.2*).

Parameter	Prerenal	Renal
U Na (mmol/L)	< 20	> 40
FE Na (%)	< 1	> 2
U osm (mmol/kg)	> 400	< 350
U osm / P osm	> 1.5	< 1.1
U urea / P urea	> 10	< 5

Table 6.2. Laboratory differences between prerenal and renal failure

Therefore, AKI is always accompanied by a decrease in GF, and usually also decreased urine production. Hyperkalaemia in particular may be life-threatening (accompanied by metabolic acidosis). Nevertheless, if tubule functions are impaired, loss of ions and water (more common in postrenal causes of AKI or damage by nephrotoxins such as heavy metals, or in kidney repair following AKI) may predominate, which leads to polyuria and, for example, hyponatraemia, hypokalaemia, hypomagnesaemia and hypophosphataemia.

Current classifications of AKI (*Table 3*) are based on changes in serum creatinine concentration and urine production. The relevance of both parameters is considerably limited (the S_Crea rise is late, muscle mass and tubular secretion bias; patients without AKI also show decreased urine production, but normal or increased urine production does itself not rule out AKI); this is why researchers are trying to find new biomarkers for early and specific detection of a developing kidney injury. The most promising markers available for clinical practice include NGAL and cystatin C.

RIFLE Stage	S_Crea	Urine Production
Risk	> 1.5 x	< 0.5 mL/kg/hr for 6 hrs
Injury	> 2 x	< 0.5 mL/kg/hr for 12 hrs
Failure	> 3 x	< 0.3 mL/kg/hr for 24 hrs
Loss	AKI or complete loss of renal functions persists for more than 4 weeks	
End-stage renal disease	Complete loss of renal functions for more than 3 months	

Table 6.3. RIFLE classification of acute kidney injury[20].

19 It is reported in about 7% of inpatients; the mortality of patients with AKI in intensive care units is > 50%. The earlier AKI is detected, the better the patient's prognosis.

20 The newer Acute Kidney Injury Network (AKIN) classification is based on parameters similar to RIFLE (S_Crea and urine production), but the combination of criteria is different (addition of the criterion for an absolute rise in S_Crea).

6.3.3.1. Case Report 2. Postrenal kidney failure

A 69-year-old man, treated for chronic arterial hypertension (calcium channel blocker monotherapy), admitted for epigastric pain and urination problems (inability to urinate), worsening for about 2 days and visible addition of blood in the urine. Following admission (selected results): S_Urea 16 mmol/L, S_Crea 192 µmol/L (MDRD-based GF 0.5 mL/s), S_Na 123 mmol/L, S_Cl 99 mmol/L, S_K 7.6 mmol/L, a flood of red blood cells in the urine. According to ultrasonic examination, the probable cause of the problems is prostatic hyperplasia combined with a clot in the bladder; bilateral hydronephrosis is present. Urine chemistry or urine sediment tests were not made at any time during hospitalization. Following prostate removal, spontaneous urination re-established; the patient had polyuria (diuresis of about 4 L), and Na had to be substituted (significant losses in the urine due to tubule compression during obstruction), and there was a partial improvement of renal functions (MDRD-based GF 0.8 mL/s when discharged from hospital).

Benign prostatic hyperplasia may lead to bilateral hydronephrosis and renal failure, which is often accompanied by signs of damaged tubule functions (hyponatraemia, polyuria – impaired renal concentrating ability).

6.3.1. NGAL

NGAL (neutrophil gelatinase associated lipocalin, lipocalin-2) is a protein produced in the nephron's distal tubule, and is considered an early marker of kidney injury. NGAL grows much earlier than creatinine (2-3 days earlier; an increase is noticeable as early as 2 hours after kidney injury), it is therefore able to predict the development of AKI and allow timely therapeutic intervention. The usefulness of NGAL has been proved in the AKI diagnosis, e.g. after cardiac surgeries or the administration of a contrast agent. Clinical algorithms and decision limits have not been sufficiently validated for different groups of patients, however. NGAL test sets are commercially available for immediate assay. Serum and urine NGAL can be measured. Urine NGAL assays seem to be more beneficial for AKI diagnosis.

Notice: Due to its expected benefit, NGAL is sometimes compared to troponin. However, unlike troponin, NGAL is not organ-specific (NGAL is produced in neutrophils, the liver, lung, bone marrow and other organs, it is also produced by some tumours and corticotherapy enhances NGAL expression).

6.3.2. Cystatin C

Cystatin C see Serum Cystatin C can also be measured in urine as an indicator of tubule functions. All filtered cystatin is normally absorbed and completely decomposed in proximal tubule cells. If proximal tubule cells are damaged, cystatin appears in the urine (tubular type of proteinuria) see also chapter Basic Urine Tests.

Serum cystatin C concentration is reported as a relatively early indicator of AKI (compared to S_Crea).

6.4. Chronic Renal Failure

Chronic renal insufficiency (CRI) is an irreversible, progressive process, usually developing for a couple of years or even tens of years. CRI is most commonly caused by diabetic and hypertensive nephropathy or polycystic kidney disease. Gradual loss of all renal functions is typical. Chronic renal insufficiency is classified into 5 stages, depending on glomerular filtration.

Stage	GF (mL/s)
1	> 1.5 + other kidney pathology[3]
2	1.0 – 1.5
3	0.5 – 0.99
4	0.25 – 0.49
5	< 0.25

Table 6.4. Classification of chronic renal insufficiency (based on KDOQI)

The decrease in GF is compensated by an increased function of residual nephrons increased fractional excretion – see Secretion and Reabsorption for a very long time; after these compensatory mechanisms have been depleted, the internal environment starts to change slowly (hyperkalaemia, acidosis, hyperphosphataemia). Other problems that can be found include anaemia (reduced erythropoetin production in the kidneys), bone disease (reduced activation = vitamin D hydroxylation in position 1 → reduced absorption of Ca^{2+} from the intestine → secondary hyperparathyroidism)

and a significantly increased risk of cardiovascular diseases.

Footnotes

1. All cases involve hyperchloraemic acidosis with normal AG
2. Give sodium bicarbonate until urine pH reaches about 7.8; then measure accurate urine pH and pCO2, calculate HC_{O3-}, followed by the FE of bicarbonate (the value for RTA II will be > 15%). The test is carried out only very rarely.
3. For example, a finding of protein in the urine (including microalbuminuria) [[see chapter 28 Diabetes Mellitus]]

CHAPTER 7

7. The Importance of Plasma Protein Assays

Authors: MUDr. Ing. David Zeman, Ph.D.; RNDr. Pavlína Kušnierová, Ph.D.

Reviewer: MUDr. Milan Dastych

7.1. Methods for Plasma Protein Examination

7.1.1. Determination of Total Protein Concentration

The biuret reaction is used to determine the serum protein concentration, and it is also currently used as a reference method. The principle is the formation of a red-purple copper (II) cation complex with nitrogen atoms involved in peptide bonds in an alkaline environment. The intensity of staining is proportional to the number of peptide bonds, and its dependence on other properties of each of the proteins is relatively small. Falsely elevated values are measured in strongly lipaemic and strongly haemolytic sera and in marked hyperbilirubinaemia; falsely lower values occur in marked hyperammonemia (competing reaction of ammonia with Cu^{2+} ions). This method cannot be used to determine urine protein or cerebrospinal fluid concentrations lower by two to three orders; in such cases turbidimetric reactions (the measurement of turbidity formed by adding trichloroacetic acid or benzethonium chloride to the sample) or various dyes (pyrogallol red-molybdate, Coomassie blue) binding to specimen proteins are used. These methods are more dependent on the properties of the protein being measured: the dye bond depends on the aromatic amino acid content, and turbidimetric reactions on the molecule shape.

7.1.2. Serum Protein Electrophoresis

This method employs the different mobility of proteins in the electric field. In an alkaline buffer environment, proteins migrate from the cathode to the anode. Five to six fractions can be distinguished by protein mobility: albumin, alpha-1, alpha-2, beta (often beta-1 and beta-2) and gamma globulins. Prealbumin migrating even faster than albumin is not normally apparent.

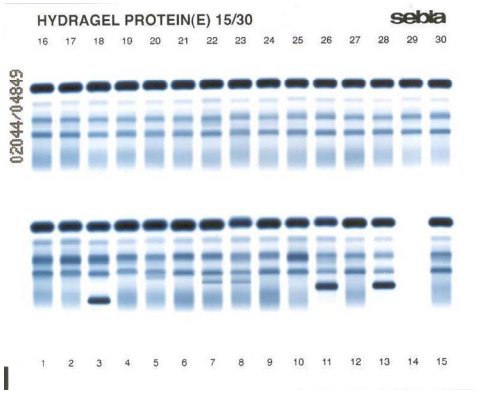


Figure 7.1. Serum protein electrophoresis. A distinct M-gradient is apparent in positions 3, 11 and 13

Electrophoresis is important primarily as a low-cost screening examination for the diagnosis of monoclonal gammopathy. It also provides an overview of the protein spectrum and allows judgements to be made concerning the presence of an acute or chronic inflammation, hepatopathy or conditions connected with a loss of proteins (through the kidneys or gastrointestinal tract).

	Albumin	Alpha-1	Alpha-2	Beta	Gamma	Comment

Acute inflammation	N - -					
Chronic inflammation	-					
Hepatitis	--	-	-	-		
Hepatic cirrhosis	--		-			Beta-gamma bridge
Nephrotic syndrome	--					
Hypogammaglobulinaemia					--	
Hypergammaglobulinaemia	-					
Alpha-1 antitrypsin deficiency		--				
Paraprotein			M-gradient	M-gradient	M-gradient	

Table 7.1. Commonest anomalies on the electropherogram

Some sites today use the technique of capillary electrophoresis, where proteins are separated in very thin silica capillaries from the anode to the cathode depending on their ion mobility using electroosmotic flow, and detected at the cathode using photometry in the UV region. The option of "statim" (immediate) analysis is favourable; however, there is no option of direct visual evaluation. Figure 2 shows an electropherogram waveform with an apparent M-gradient (paraprotein, see below in the text).

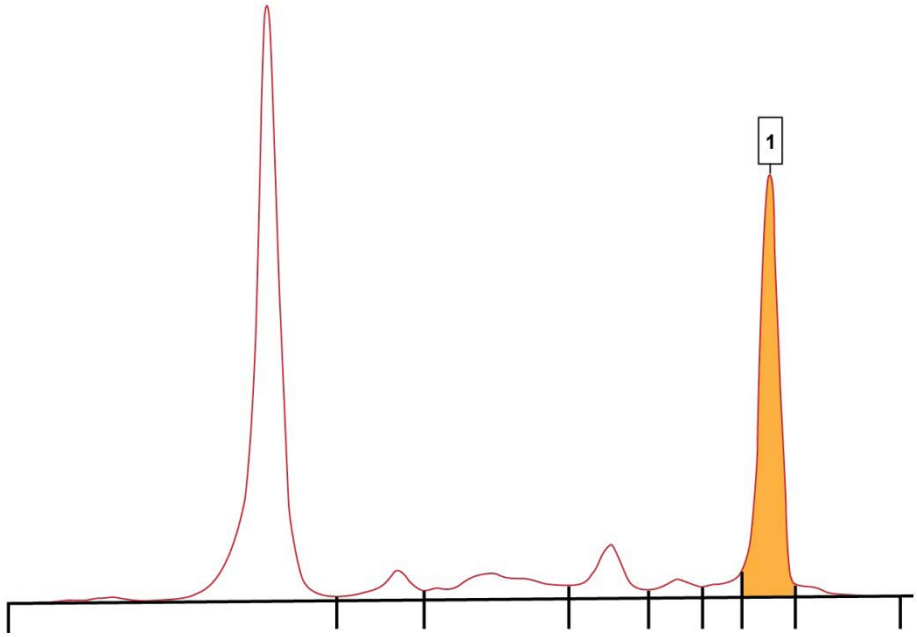


Figure 7.2. Electropherogram. Left to right: albumin, alpha-1, alpha-2, beta-1, beta-2 globulins, and a high, narrow paraprotein peak (M-gradient) in the gamma-globulin zone

Note: The electropherogram is actually plotted from right to left during capillary electrophoresis, i.e. gamma globulins are detected first in the arrangement above, and albumin last. Go back over the principles of protein fragmentation by electrophoresis and capillary electrophoresis (particularly the concept of electroosmotic flow) and work out why this is the case!

7.1.3. Determination of Plasma Protein Concentration

Specific protein concentrations are measured using immunochemical methods, i.e. methods employing reactions of the protein with the specific antibody. The turbidimetric or nephelometric principle of measurement, which has completely superseded radial immunodiffusion previously used in clinical laboratories, can be used for most major plasma proteins. Labelled reactant methods (ELISA, RIA, chemiluminescence, fluorescence) are used for proteins present in very low concentrations. An exception is the determination of plasma albumin, where either a calculation from the total protein concentration and the relative albumin proportion on the electropherogram can be used (it is not possible to determine the other proteins in this way because the fractions in which they migrate in electrophoresis are generally formed by proteins impossible to distinguish from one another), or else a simple photometric reaction is used

based on the stain bond to albumin (bromcresol green) with a subsequent change in their absorption maximum (the BCG-albumin complex gives a marked maximum at 630 nm).

7.1.4. Overview of Major Plasma Proteins

A variety of views exist regarding the sorting out of proteins. We choose to sort proteins by their electrophoretic mobility, from prealbumin and albumin through alpha-1, alpha-2 and beta globulins to gamma globulins or immunoglobulins. Following immunoglobulins, we will examine the issue of monoclonal gammopathies. Protein groups sharing common features will then be described: the group of acute phase reactants, the group of low molecular weight proteins and the group of enzymes. The issue of protein hormones is dealt with in another chapter.

Protein	Mol. Weight (Da)	Reference Limits (Adults)	Electroph. Mobility
Prealbumin	55,000	0.1 – 0.4 g/l	Prealbumin
Albumin	66,000	35 – 55 g/l	Albumin
Alpha-1-acid glycoprotein	44,100	0.55 – 1.20 g/l	Alpha-1
Alpha-1-antitrypsin	54,000	2 – 4 g/l	Alpha-1
Alpha-1-antichymotrypsin	68,000	0.3 – 0.6 g/l	Alpha-1
Haptoglobin	Hp 1-1: 85,000 Hp 2-1: 120,000 Hp 2-2: 160,000	1.0 – 2.4 g/l 1.5 – 3.2 g/l 1.1 – 2.8 g/l	Alpha-2
Gc-globulin (group specific component; vitamin-D binding protein)	51,000	0.2 – 0.55 g/l	Alpha-2
Alpha-2-macroglobulin	725,000	1.5-4 g/l	Alpha-2
Transferrin	76,500	2.0 – 4.0 g/l	Beta-1
Haemopexin	57,000	0.5 – 1.2 g/l	Beta-1
Caeruloplasmin	151,000	0.1 – 0.6 g/l	Beta-1
Beta-2-microglobulin	12,000	0.9 – 2.3 mg/l	Beta-2
C3	180,000	0.8 – 1.6 g/l	Beta-2
C4	206,000	0.20 – 0.65 g/l	Beta-2
CRP	135,000	0 – 10 mg/l	Beta-2
Fibrinogen	341,000	2.8 – 4.5 g/l	Beta-gamma interzone
IgM	900,000	0.6 – 2.5 g/l	Beta-gamma
IgA	160,000	0.9 – 4.5 g/l	Beta-gamma
IgG	150,000	8.0 – 18.0 g/l	Gamma

Table 7.2. Blood serum proteins – molecular weight, reference limits and electrophoretic mobility

Protein	Reduced Concentration	Elevated Concentration
Prealbumin	Malnutrition (best indicator – short half-life!) Hepatic disease Acute phase reaction Newborns (about half the concentration compared to adults) Slight drop after 50 years of age	Effect of glucocorticoids and androgens

Albumin	Reduced synthesis Congenital (analbuminaemia) Acquired (e.g. inflammation, hepatic disease, inhibition of synthesis by alcohol) Increased catabolism Losses through urine (nephrotic syndrome), gastrointestinal tract Malnutrition (slow response)	(Acute dehydration)
AAG	Reduced synthesis (effect of oestrogens) Increased losses (nephrotic syndrome, losses through gastrointestinal tract)	Acute phase reaction (3 – 4x, peak 3 – 5 days from disease onset) Ulcerative colitis (activity marker) Effect of glucocorticoids (Cushing's syndrome, corticotherapy)
AAT	Genetic defect (phenotypes Pi Z, 00) Increased utilization (neonatal respiratory distress syndrome, neonatal hepatitis, severe pancreas diseases) Losses through urine or gastrointestinal tract	Acute phase reaction (from day 2, peak in 3 – 4 days) Stimulation of synthesis by oestrogens (pregnancy, oestrogen therapy)
Caeruloplasmin	Genetic defect (rare) Secondary in Cu^{2+} deficiency in food, Menkes disease, Wilson's disease	Acute phase reaction (slow and small increase) Stimulation of synthesis by oestrogens (oestrogen therapy, pregnancy), cholestasis
CRP		Acute phase reaction (fast and marked increase particularly in bacterial infections)
Alpha-1-foetoprotein	In maternal serum in Down syndrome	Tumour marker: hepatocellular carcinoma, germ cell tumours In maternal serum in neural tube defects
Haptoglobin	Genetic defect (rare) In vivo haemolysis Effect of oestrogens Hepatic disease Newborns	Acute phase reaction (4 – 6 days from disease onset) Protein losses in connection with Hp 2-1 or 2-2 phenotypes Effect of androgens or glucocorticoids
Alpha-2-macroglobulin	Acute phase reaction Pancreatitis Prostate carcinoma (AMG binds PSA and complex is very quickly taken up by liver, unlike PSA complex with alpha-1-antichymotrypsin)	Effect of oestrogens In childhood (two to three times higher than concentration in adults) Nephrotic syndrome (compensation of renal losses of lower molecular weight proteins)
Transferrin	Iron deficiency anaemia (low iron saturation) Anaemia due to defect of iron incorporation into red blood cells (high iron saturation) Acute phase reaction Chronic hepatic diseases Malnutrition Protein losses through urine or gastrointestinal tract Congenital defect (atransferrinemia)	Effect of oestrogens (pregnancy, oestrogen therapy)

C3	Genetic defect Increased C3 activation (systemic connective tissue disease, severe infections, septicaemia) Concentration in newborns 1/3 lower than adults	Acute phase reaction (slight increase) Biliary tract obstruction Focal glomerulosclerosis (30% of patients, correlates with favourable prognosis)
C4	Genetic defect (associated with high prevalence of autoimmune diseases, SLE in particular) Consumption (SLE, hereditary angioedema, autoimmune haemolytic anaemia, autoimmune nephritis) Newborns (1/2 to 1/4 lower concentration than adults)	Acute phase reaction (slight increase)
Immunoglobulins	Primary humoral immunodeficiencies (rare) Secondary humoral immunodeficiencies: Synthesis disorder (IgM drops first, then IgA, and IgG last): lymphoproliferative diseases, effect of drugs (phenytoin, penicillamine) Loss of proteins (nephrotic syndrome, burns, losses through gastrointestinal tract)	Polyclonal increase: response to infection, chronic inflammations, autoimmune disease (primary biliary cirrhosis: increase in IgM in particular) Monoclonal: paraprotein

Table 7.3. Changes in plasma protein concentration in some pathological conditions

7.1.4.1. Prealbumin and Retinol Binding Protein

Prealbumin (transthyretin – can bind two thyroxine or triiodothyronine molecules) is formed in the nuclei and choroid plexus of brain ventricles. It is partially associated with retinol-binding protein in a 1:1 ratio. It is a non-glycosylated tetrameric protein consisting of four identical subunits. Due to its short biological half-life (48 hours; retinol-binding protein 12 hours) it is a sensitive marker of acute malnutrition and proteosynthesis.

Note: Prealbumin determination is sometimes used in calculating the Prognostic Inflammatory and Nutrition Index (PINI):

$$PINI = \frac{AAG\ (^{mg}/_{l}) \cdot CRP\ (^{mg}/_{l})}{albumin\ (^{g}/_{l}) \cdot PA\ (^{mg}/_{l})}$$

Normal PINI values are <1; low-risk patients have values of 1 – 10, medium-risk patients 11 – 20 and high-risk patients 21 – 30.

Prealbumin is one of the negative acute phase reactants (see below). Prealbumin concentration drops in inflammatory and tumour diseases, hepatic cirrhosis and losses of proteins through the kidneys or gastrointestinal tract. The retinol-binding protein is a monomeric all-trans retinol-carrying protein (a form of vitamin A). It is synthesized in the liver and its secretion from hepatocytes is stimulated by the retinol bond. The production of the complex with prealbumin prevents filtration of the retinol-binding protein in glomeruli. The retinol release is followed by prealbumin and retinol-binding protein complex dissociation, which is then eliminated in the kidneys. Elevated retinol-binding protein concentrations can be found in patients with chronic renal disease, reduced concentrations in hepatic diseases and protein malnutrition.

7.1.4.2. Albumin

Albumin is synthesized solely in the liver, which has a high reserve in case of increased need or loss of albumin. Albumin is one of the few plasma proteins that does not contain carbohydrates. The charge is negative with a physiological pH. Due to its high plasma concentration, albumin plays a substantial role in maintaining the colloid osmotic pressure. In quantitative terms, it is the most important transport protein in the body. Many substances in the blood

bind to albumin; this is conditioned by the presence of numerous charged groups, ion- and specific hydrophobic binding sites. In particular, these are insoluble substances (unconjugated bilirubin, fatty acids, hormones, drugs). Hydrophilic substances can bind to albumin by adsorption (e.g. $Ca^{2+,}$ Zn^{2+}, uric acid). The transporting function of albumin is non-specific, unlike the specific transport proteins mentioned below. In addition, albumin takes part in antioxidation activity, acts as a buffer in pathological conditions, and inhibits the production of leucotrienes and actin. Albumin is a negative acute phase reactant; albumin concentration drops in the presence of proinflammatory cytokines (TNF-alpha, IL-1, IL-6).

7.1.4.3. Alpha-1-Globulins

Alpha-1 Lipoprotein (HDL) – refer to relevant chapter

Orosomucoid (Alpha-1-Acid Glycoprotein, AAG)

For the most part, orosomucoid is synthesized in the liver, but the synthesis in granulocytes and monocytes contributes considerably to the increase in plasma concentration in septic conditions. Catabolism takes place predominantly through the uptake of desialylated AAG by hepatic receptors for asialoglycoproteins; while the biological half-life of an intact AAG is about 3 days, the half-life of desialylated AAG is just a few minutes. AAG belongs to lipocalins, i.e. proteins binding lipophilic substances. Its function is not yet fully known; it is supposed to have an immunomodulatory effect, probably conditioned by the structure of oligosaccharide chains of the AAG molecule.

Alpha-1-Foetoprotein

Alpha-1-foetoprotein has the function of albumin in the foetus, and is replaced by albumin after birth. AFP is a major marker in prenatal diagnosis (elevated amniotic fluid and also maternal serum concentrations in neural tube defects, but also in multiple pregnancy; reduced concentrations in Down syndrome – only a combined determination with beta-HCG and/or unconjugated oestriol in maternal serum, the “triple test”, is currently used in this indication for screening for congenital developmental defects [“CDD screening” in the 2nd trimester]). In non-pregnant adults, AFP is elevated in hepatocellular carcinoma and germ cell tumours; it can also be elevated in hepatic cirrhosis and hepatitis.

Alpha-1-Antitrypsin (Alpha-1-Proteinase Inhibitor, AAT, Alpha-1-Pi)

In quantitative terms, alpha-1-antitrypsin is the most important proteinase inhibitor in the plasma. In the case of a significant genetic reduction or absence (PiZZ and Pi00 genotypes) there is insufficient inactivation of proteinases (elastases in alveoli) released by granulocytes and macrophages during phagocytosis, and proteolytic tissue degradation occurs. This leads to pulmonary emphysema with a ventilation disorder and a tendency to severe airway infections, and often liver damage.

Alpha-1-Antichymotrypsin

Alpha-1-antichymotrypsin is another proteinase inhibitor; it inhibits cathepsin G, mast cell chymase and PSA (increased concentrations of PSA bound to alpha-1-antichymotrypsin can be found in patients with prostate carcinoma).

7.1.4.4. Alpha-2-Globulins

Alpha-2-Macroglobulin

Alpha-2-macroglobulin (AMG) is another proteinase inhibitor. It is composed of four polypeptide chains; the active unit is a dimer (one AMG molecule can bind two proteinase molecules). For the most part, it is synthesized by hepatocytes. The biological half-life of an intact protein is a few days, but after cleaving the intrachain thioester bond or cleaving off sialic acid, it is quickly (within minutes) taken up by receptors in hepatocytes. Apart from its main function – proteinase inhibition – AMG participates in the transport of many peptides (cytokines, growth factors) and bivalent cations (Zn^{2+} in particular), and modulates inflammatory reactions.

Increased AMG concentrations can be found in nephrotic syndrome; children have higher concentrations than adults, and women of reproductive age have higher concentrations than men of the same age (oestrogens increase

AMG synthesis). Reduced AMG concentrations are found in pancreatitis and advanced prostate carcinomas (AMG binds to PSA and this complex is very quickly taken up by the liver).

Haptoglobin

Haptoglobin binds haemoglobin generated during intravasal haemolysis; the formed complex is quickly taken up from the bloodstream by RES cells in the liver. This prevents the loss of haemoglobin (Fe^{2+}) in urine as well as the creation of the dangerous hydroxy radical involving free haemoglobin.

Haptoglobin is a sensitive marker of haemolysis, which causes a marked drop in haptoglobin concentration down to zero. Reduced concentrations are also associated with disorders of proteosynthesis in the liver; the haptoglobin level correlates with the severity of hepatitis. Increased concentrations are found in the acute phase reaction.

Caeruloplasmin

The main function of caeruloplasmin is ferroxidase activity (oxidizes Fe^{2+} to Fe^{3+}, thus allowing iron to bind to transferrin). One caeruloplasmin molecule binds 6 – 8 atoms of copper, but the main copper-carrying protein is albumin.

Markedly decreased concentrations of caeruloplasmin are found in Wilson's disease (mostly < 0.1 g/l) and Menkes syndrome. Caeruloplasmin concentrations can be reduced secondarily in the case of protein losses and protein synthesis disorders (hepatic cirrhosis). Elevated concentrations do not have any diagnostic significance; they can be found in inflammations and tumours, during pregnancy or the use of oral contraceptives (effects of oestrogens) and in cholestasis.

Gc-Globulin (Group Specific Component, Vitamin-D Binding Protein)

This is sometimes referred to as "inter-alpha globulin" as it can migrate in the alpha-1-alpha-2 interzone. Gc-globulin is the main vitamin D-carrying protein; another function is the uptake of actin released from damaged cells. Women have slightly higher Gc-globulin concentrations than men; the levels rise in pregnancy and fall in acute hepatic disease or following serious trauma.

7.1.4.5. Beta Globulins

Transferrin

Each transferrin molecule has two binding sites for the ferric cation (Fe^{3+}). Apo-transferrin is synthesized primarily in the liver; synthesis increases with a lack of iron in the body. Transferrin is the major iron-carrying element from the sites of absorption (small intestine) and takes part in haemoglobin degradation in haematogenous organs, where it binds to erythroblasts through the receptor. In situations where there is a lack of iron, transferrin receptor synthesis is stimulated and soluble transferring receptor (sTfR – the released extracellular domain of the membrane receptor) concentration increases. The sTfR concentration is not affected by the acute phase reaction; sTfR (or the sTfR concentration to ferritin concentration logarithm ratio) is therefore a valuable marker helping to distinguish iron-deficiency anaemia from chronic anaemias.

The percentage saturation of transferrin with iron can be calculated from the following formula:

$$Transferrin\ saturation = \frac{3.98 \cdot serum\ iron\ (^{\mu mol}/_{l})}{serum\ transferrin\ (^{g}/_{l})}$$

Physiologically, about 1/3 of transferrin is saturated with iron; vacant sites for the iron bond represent the free binding capacity. (The total iron-binding capacity of transferrin is given by the number of transferrin molecules available as well as the fact that one transferrin molecule has two binding sites for Fe^{3+}.)

Transferrin contains 2 N-linked glycans that may further branch out; each branch is capped by a negatively charged molecule of sialic acid. Theoretically, a transferrin molecule can contain 0 – 8 sialic acid molecules; most molecules contain 4 (64 – 80 %) or 5 (12 – 18 %) sialic acid molecules. The isoelectric point of transferrin increases with the decreasing

number of sialic acid molecules. Alcoholics show an elevated concentration of transferrin isoforms containing a lower number of sialic acid residues (CDT – carbohydrate-deficient transferrin, i.e. disialo-, monosialo- and asialotransferrin) and therefore a higher pI, which can be used to monitor these patients. However, the isoelectric point of transferrin decreases with the Fe^{3+} bond, so transferrin should be saturated with iron before the CDT assay.

Haemopexin

Haemopexin binds haem. A decrease in haemopexin is a less sensitive (delayed) indicator of haemolysis than haptoglobin. When the haptoglobin capacity to bind haemoglobin is exhausted, the remaining haemoglobin circulates in the plasma; it is then partly taken up by liver cells, partly excreted in urine (haemoglobinuria) and partly oxidized to methaemoglobin, which readily dissociates haem. Haem can bind to haemopexin and is transported to liver cells in this bond. Haem can also bind non-specifically to albumin to form methaemalbumin, from which it can be "given back" to haemopexin as soon as haemopexin is available.

The haemopexin concentration is usually decreased in haemorrhagic pancreatitis and bleeding into body cavities with concurrent normal concentrations of haptoglobin.

Beta-Lipoprotein

Lipoprotein in the beta-2-zone corresponds to low density lipoprotein (LDL). *LDLs (low density lipoproteins)* are the main lipoprotein particles that carry cholesterol in the blood plasma. Their density is 1,019 to 1,063 kg/m^3, i.e. 1.019 to 1.063 g/ml; they are produced in IDL and VLDL catabolism, and their major apoprotein is apo-B-100. The major part comes from VLDL transformation, but some are synthesized directly (especially in patients with familial hypercholesterolemia). LDL can be catabolized by various cell types (75 % in the liver), using both the LDL-receptor dependent mechanism and mechanisms independent of these receptors (e.g. using "scavenger" receptors).

C3 and C4 Complement Components

The complement system consists of about 30 proteins that can be sorted into several groups by their function:

- Classical pathway proteins (C1, C4, C2, C3);
- Alternative pathway proteins (C3, factors B and D, properdin);
- Membrane attack complex (C5 – C9);
- Aforementioned pathway inhibitors;
- Cell receptors.

The main functions of the complement system are:

- Opsonization (a fragment of the C3 component referred to as C3b that firmly binds to the microbial surface);
- Chemotaxis (fragments C3a and C5a);
- Osmotic lysis (C5b-C9 complex penetrating into the plasma membrane of the attacked cell and making pores in it; most microorganisms are protected from this lytic action by the cell wall).

The C3 and C4 components are determined most often in clinical practice. Their concentration rises slightly in the acute phase reaction, while it decreases with their activation or consumption in some systemic autoimmune diseases. The C3 component has the highest plasma concentration of all complement components. It is synthesized primarily in the liver; bacterial endotoxin also induces synthesis in monocytes and fibroblasts. The C4 component is likewise synthesized primarily by liver cells, a small amount in other tissues.

Fibrinogen

Fibrinogen is a glycoprotein with a molecular weight of 340 kDa. It migrates between beta and gamma globulins in electrophoresis. It is a symmetric dimer consisting of three pairs of alpha, beta and gamma chains.

During haemocoagulation, fibrinopeptide A is released by the action of the proteolytic enzyme, thrombin, to form fibrin monomer. Thrombin then cleaves off fibrinopeptide B and fibrin monomer polymerization occurs as activated coagulation factor XIII cross-links fibrin to form a stable thrombus.

Fibrinogen is one of the positive acute inflammation phase reactants. Within a few hours of the onset of inflammation the fibrinogen level increases by the induction of increased synthesis in the liver, and in a few days may reach many times higher levels than the baseline. High concentrations can be found in inflammatory connective tissue diseases, for example. Fibrinogen is a risk factor for aterosclerosis.

A decrease in concentration can be caused by insufficient synthesis in severe hepatopathies or by increased consumption, for instance in disseminated intravascular coagulopathy (DIC).

Fibrinogen can be found in the beta-gamma interzone on the electropherogram. It can be confused for the M-gradient (and vice versa), which is the main reason why serum is preferred over plasma for electrophoresis.

7.1.4.6. Gamma Globulins

Immunoglobulins

Immunoglobulins (antibodies) detect an antigen and initiate mechanisms to destroy it. All immunoglobulin molecules consist of two identical H-chains (H for heavy) (alpha for IgA, delta for IgD, gamma for IgG, mu for IgM, epsilon for IgE) and two identical L-chains (L for light) (kappa or lambda). Each of the chains has a constant part and a variable part, the latter participating in the distinguishing and binding of antigens. The variable part of the heavy and light chain consists of a single domain; the constant part of the light chain also consists of a single domain, but the constant part of the heavy chain consists of multiple domains. The light chain together with the variable part and the first domain of the constant part of the heavy chain form the Fab fragment ("ab" stands for "antigen binding" – antigen-binding fragment); the other domains of the constant parts of heavy chains form the Fc fragment ("c" stands for crystallizable). Therefore, every immunoglobulin molecule has two Fab fragments and one Fc fragment. Unlike most plasma proteins synthesized in the liver, immunoglobulins are synthesized and secreted by plasma cells.

Immunoglobulins are glycoproteins; oligosaccharide is always bound on the constant part of the heavy chain (one oligosaccharide binds to the gamma chain, multiple oligosaccharide molecules bind to the other H-chain types).

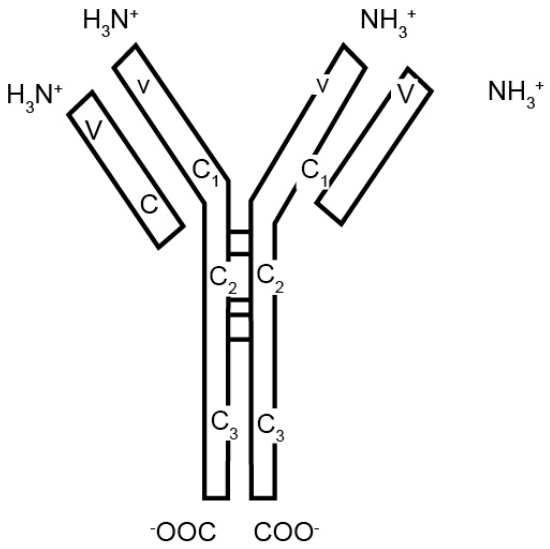


Figure 7.3. The simplest schematic representation of an immunoglobulin molecule consisting of two heavy (H) chains and two light (L) chains. The light chains have a variable (V) part and a constant (C) part of about the same size, while the variable part of heavy chains is just about one quarter, and the constant part about three quarters (C1 – C3) of its size. The lines between the chains represent disulphide bonds. Intra-chain disulphide bonds are not shown. The number of disulphide bonds in the hinge region (connecting both heavy chains) varies in different classes and subclasses of immunoglobulins

Immunoglobulin G (IgG)

Quantitatively this is the most common (70 – 75 % of total immunoglobulins). 65 % of IgG is present extravascularly. The main function of IgG is to neutralize toxins in tissues. IgG contains one molecule of oligosaccharide capped by

glucosamine (GlcN) bound by the N-glycosidic bond to asparagine in the constant part of the heavy chain. IgG is sorted into several subclasses (IgG1 – IgG4), distinguished from each other primarily by the hinge region arrangement.

Immunoglobulin M (IgM)

IgM is the first immunoglobulin synthesized during the antibody response, and also the only immunoglobulin normally synthesized in newborns. It exists as a monomer bound to the membrane, while most IgM present in the serum is a pentamer consisting of five monomer immunoglobulin units linked by a small glycopeptide known as the J-chain. IgM contains about 10 % carbohydrates. The large molecular weight of IgM (MW 900,000 per pentamer) prevents it from escaping into the extravascular space. IgM effectively activates the complement; this function is provided by the Fc part of the molecule.

Immunoglobulin A (IgA)

Immunoglobulin A represents 10 to 15 % of serum immunoglobulins. IgA contains 10 % carbohydrates, and the biological half-life is about 6 days. The majority of IgA is present as monomers but 10 to 15 % of serum IgA is dimeric. The dimer is more resistant to destruction by pathogenic microorganisms. Another form of IgA is secretory IgA, which is present in tears, saliva, sweat, milk, colostrum, and airway and digestive tract secretions. Secretory IgA MW is 380,000 and consists of two monomeric IgA units, the secretory component (MW 70,000) and the J-chain (MW 15,600). It is synthesized primarily by plasmocytes in the intestine and bronchi mucosa, and in mammary gland outlets. The secretory component provides secretory IgA with an increased resistance to enzymes, thereby protecting mucosas against microorganisms. IgA presence in the colostrum and milk protects newborns against intestinal infections.

Immunoglobulin D (IgD)

Immunoglobulin D represents less than 1 % of serum immunoglobulins. It is a monomer containing 12 % carbohydrates, with a MW of 184,000. It is present as a membrane receptor for antigen on B lymphocytes.

Immunoglobulin E (IgE)

IgE contains 15 % carbohydrates and has a MW of 188,000. IgE molecules bind to the surface of mast cells (with their Fc region). If antigen (allergen) is bound and the two linked IgE molecules connected, mast cells are stimulated to release histamine and other vasoactive substances responsible for allergic reaction manifestations.

Physiological, temporary hypogammaglobulinaemia can be found in newborns and infants (see Figure 4). Antibody immunodeficiencies lead to permanently low to zero immunoglobulin levels; immunoglobulins of all isotypes (e.g. Bruton's X-linked agammaglobulinaemia), or just some isotypes (e.g. common variable immunodeficiency with IgG and IgA antibody disorder, or selective IgA deficiency, locally the commonest antibody immunodeficiency) can be affected. Secondary (acquired) antibody immunodeficiencies are more frequent than these primary immunodeficiencies. The former occur in intestinal inflammations, nephrotic syndrome, long-term immunosuppresive or cytostatic therapy and B-cell malignancies.

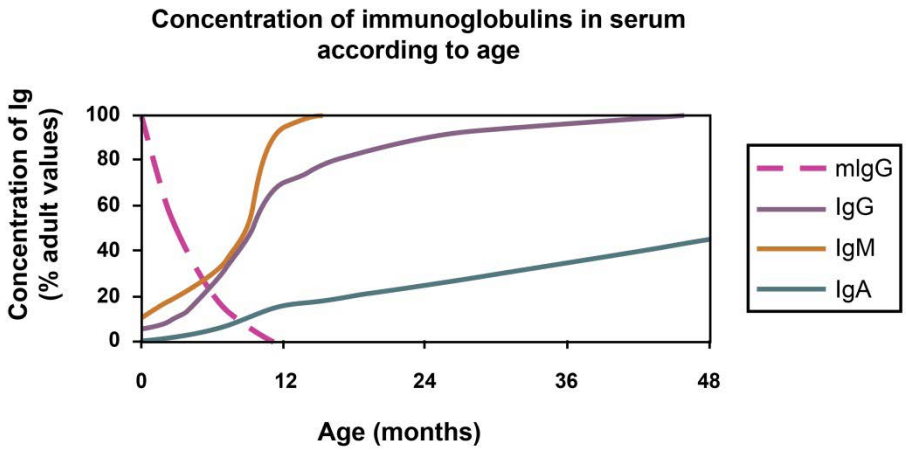


Figure 7.4. Schematic representation of changes in immunoglobulin concentrations in the first months of life. Transplacental IgG antibodies (mIgG) decrease to zero in the first year of life; the lowest concentration of total IgG is between the 3rd and the 6th month. The IgA concentration reaches adult values around 10 years of age

Kappa and Lambda Free Light Chains

Light chains are part of the immunoglobulin molecule; their physiological production is slightly higher than the production of heavy chains, which is why they can be also found in the blood serum. It seems that they have a certain importance in immune response regulation. Markedly elevated (even by several orders) concentrations of free light chains of one type can be found in the blood serum in tumorous plasma cell proliferations; besides absolute concentrations, the ratio (free kappa/free lambda), or more recently the difference (free kappa – free lambda, or vice versa) of their concentrations are also monitored. Their determination requires the use of a specific antibody against hidden, "invisible" determinants, if the light chain is bound to the heavy chain in the complete immunoglobulin molecule. The issue of free light chain immunochemistry is therefore highly specific and has not been satisfactorily resolved up to now; nevertheless, they are a valuable marker in the diagnosis and monitoring of monoclonal gammopathy patients, as well as a promising marker of some autoimmune disease activity (systemic lupus erythematosus, rheumatoid arthritis, Sjögren's syndrome, etc.).

7.1.4.7. Monoclonal Gammopathies

This term refers to the presence of immunoglobulins (or fragments thereof) produced by one clone of plasma cells. In accordance with the 2003 international classification, they include the following conditions:

- Monoclonal gammopathy of undetermined significance (MGUS)
- Asymptomatic (smouldering) multiple myeloma (ASMM)
- Symptomatic multiple myeloma
- Solitary bone plasmacytoma
- Extramedullary plasmacytoma
- Multiple solitary plasmacytoma
- Plasma cell leukaemia

Monoclonal immunoglobulins can also be rarely found in non-tumour diseases such as AIDS, chronic hepatitis, cirrhosis or rheumatoid arthritis; a part of such cases probably coincides with MGUS.

Serum protein electrophoresis is used as a screening test for monoclonal gammopathy diagnosis. If a monoclonal (M) gradient is found, immunofixation electrophoresis is added to determine the M-gradient type (IgG, IgA, IgM, kappa, lambda, and/or free kappa, free lambda, IgD and IgE). IgM monoclonal gammopathies are accompanied by a disease known as Waldenström's macroglobulinaemia. Other immunoglobulins may bind to monoclonal IgM molecules during immunofixation and "false" positive results may occur in the other classes. This situation can be elegantly solved by adding a small amount of mercaptoethanol or dithiothreitol (see Figures 7.5a-c).

Figure 7.5. a.

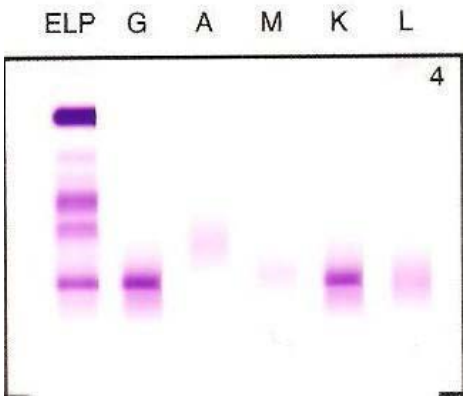


Figure 7.6. b.

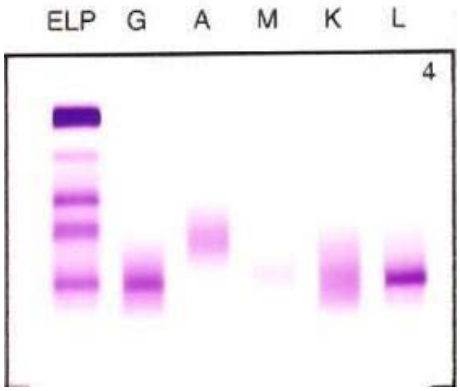


Figure 7.7. c.

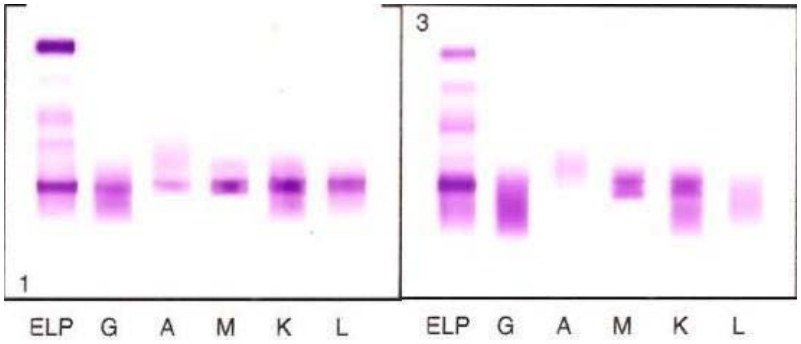


Figure 7.8. Immunofixation electrophoresis

ELP: electrophoresis

G, A, M, K, L: Immunofixation using antisera against gamma, alpha, mu, kappa and lambda chains.

IgG-kappa paraprotein

IgG-lambda paraprotein

IgM-kappa paraprotein apparent after sample incubation with dithiothreitol; “false” positive results occur in all positions of the sample without dithiothreitol

If paraprotein is found, its quantity is determined using densitometry. Immunofixation should be added at the first examination to determine the heavy and the light chain type; the lab report for the attending physician then looks like this: “IgG-kappa paraprotein, 30 g/l by densitometry”. As long as paraprotein is quantifiable by densitometry, immunofixation is not repeated during further patient follow-up. Paraprotein often disappears when modern therapeutic methods are used; as soon as it is no longer quantifiable, immunofixation is indicated again. Stem cell transplantation is sometimes followed by an oligoclonal response (some bands are only detectable by electrophoresis or immunofixation; they may belong to different light or heavy chain classes); this does not mean a relapse of the disease but a temporary dysregulation of immunoglobulin production after transplantation.

Cryoglobulinaemia

Cryoglobulins are immunoglobulins reversibly precipitating at temperatures lower than human body temperature. There are 3 types of cryoglobulins:

- Type I: Isolated monoclonal cryoglobulins (IgM paraproteins, rarely IgG, IgA paraproteins or monoclonal light chains);
- Type II: Mixed cryoglobulins (a combination of paraprotein with polyclonal immunoglobulin, usually IgM paraprotein with antibody activity against polyclonal IgG);
- Type III: Polyclonal immunoglobulins, where cryoglobulin consists of immunoglobulins of one or multiple classes forming an antigen-antibody complex.

Cryoglobulinaemia affects many laboratory examinations (e.g. it slows down sedimentation); the serum or plasma

specimen should be preheated for all assays; incubation of the serum with mercaptoethanol or dithiothreitol is sometimes recommended before electrophoresis.

Even rarer than cryoglobulins are pyroglobulins; these are monoclonal immunoglobulins irreversibly precipitating at 56°C.

Polyclonal Gammopathy

Elevated concentrations of polyclonal gammaglobulins (i.e. the homogeneous increase in gammaglobulins during electrophoresis) can be found in connective tissue diseases (systemic lupus erythematosus, Sjögren's syndrome, rheumatoid arthritis, etc.), liver diseases (autoimmune and viral hepatitis, primary biliary cirrhosis, primary sclerosing cholangitis, haemochromatosis, alcoholic liver damage, α1-antitrypsin deficiency), chronic infections, lymphoproliferative diseases and non-haematological malignancies. Skin, lung and intestine diseases and also some liver diseases are often accompanied by an elevated IgA concentration, while systemic infections lead to elevated concentrations of all immunoglobulins, IgG in particular.

7.1.4.8. Acute Phase Reactants

This term refers to proteins whose concentration in the plasma alters in an acute inflammatory involvement. Acute phase reactants are divided into positive reactants (CRP, SAA and others), whose concentration increases in such circumstances, and negative reactants (prealbumin, albumin, transferrin), whose serum concentration conversely decreases.

C-Reactive Protein (CRP)

The name of this protein is derived from its ability to bind the cell wall C-polysaccharide of *Streptococcus pneumoniae*. However, not only does it bind a variety of polysaccharides on the surface of bacteria, fungi and protozoa, but also phosphatidylcholines and polyanions. After binding, the complement is activated in the classical way. The main function of the CRP is to detect potentially toxic substances released from damaged tissues.

The CRP is the most important positive acute phase reactant as the CRP assay is easily available and its concentration rises quickly (within 24 – 48 hours). CRP levels increase in bacterial infections, but also following myocardial infarction, traumas, surgeries and tumour diseases.

Serum Amyloid A (SAA)

Serum Amyloid A is a protein with a total MW of 220,000 – 235,000, consisting of monomers with a MW of 12,500. For the most part, it is bound to HDL lipoproteins. This is a group of small lipoproteins that take up lipid residues from cell membranes during acute infection, thereby having a positive effect; during chronic inflammations, on the other hand, the precursor protein SAA may form insoluble deposits in organs and participate in secondary amyloidosis (unlike secondary amyloidosis, primary amyloidosis is typical of light immunoglobulin chains in malignant myeloma and related diseases). The serum SAA concentration is also increased in some malignant tumours and rises as the disease progresses.

Procalcitonin (PCT)

Procalcitonin increases in the plasma in bacterial infections, sepsis and multiple organ failure. The main stimulus is bacterial endotoxins. An increase in the plasma occurs after 2 hours; concentrations are clearly elevated in 6 – 8 hours and remain elevated for 72 hours.

Ferritin

Above all, ferritin is a reserve protein for iron; one molecule of apoferritin can bind up to 4500 iron atoms (Fe^{3+}). Since it is a positive acute phase reactant at the same time, it is mentioned here.

7.1.4.9. Some Low-Molecular Weight Proteins

Some of the major low-molecular weight proteins will now be described. A general feature of low-molecular weight proteins is a rise in serum concentration with a decrease in glomerular filtration; they are therefore used as renal disease markers, and are discussed in the relevant chapter as such. Only the most important ones are mentioned here:

Beta-2-Microglobulin

Beta-2-microglobulin is a simple non-glycosylated polypeptide consisting of 100 amino acids with one disulphide bridge and a molecular weight of 11,800 Da. It is a part of the HLA system on the cell surface. The plasma beta-2--microglobulin is mainly a product of myeloid and lymphoid cells, and can be used as a tumour marker reflecting a proliferation of cells of this origin.

Beta-2-microglobulin freely passes through the fully functional glomerulus. Its level in the urine rises in tubular cell lesion; the serum level increases with a decrease in glomerular filtration. A long-term increase of serum beta-2-microglobulin concentration, in dialysed patients for example, may be the cause of secondary amyloidosis.

Cystatin C

Cystatin C is a protein from the cysteine protease inhibitor family, filtered by glomeruli and fully reabsorbed by renal tubules, and suitable for evaluating glomerular filtration. Cystatins have an evolutionary and functional relationship to serine protease inhibitors. Cystatin C, which is an eminent subject of interest for clinical biochemists, belongs to cystatin Family 2, consisting of a group of extracellular cysteine protease inhibitors. Each representative of this group has a serial number, although the third cystatin continues to be referred to as cystatin C.

Beta-Trace Protein

Beta-Trace Protein is an enzyme converting PGH_2 to PGD_2 (prostaglandin-H_2 D-isomerase), with a MW of 25 kDa. It originates solely in the brain; the main point of synthesis is the choroid plexus and leptomeninges. From there, BTP is released into the cerebrospinal fluid (CSF). BTP is present in traces also in other body fluids, including serum. Its CSF concentration is 32 times higher than its serum concentration. Therefore, it is an excellent marker of liquorrhea. Apart from being used in CSF diagnosis, it is sometimes used as a marker of glomerular filtration (like cystatin C). One drawback to the assay is its high cost.

Cytokines

The term “cytokine” refers to a signal substance of cellular origin. Cytokines are low-molecular weight (often glycosylated) proteins produced by immune system cells (lymphocytes, monocytes, macrophages, neutrophil granulocytes) as well as classical “non-immune” cells (endothelial cells, fibroblasts, etc.). They are distinguished by their major action on the cells that produce them, and on cells in their close vicinity (autocrine and paracrine action). They act in very low concentrations and have a very short half-life (minutes).

Interleukin 6 (IL-6)

The IL-6 level rises in 2 – 4 hours in acute inflammation, independently of liver function. It is therefore especially suitable for an early diagnosis. The IL-6 concentration in the plasma or serum of healthy people is <11.3 ng/l; concentrations >400 ng/l give rise to a strong suspicion of sepsis. The sensitivity of the assay is almost 100 %, while specificity depends on the underlying disease. Elevated concentrations may be found in autoimmune diseases, glomerulonephritis, tumours, and in the first days after a surgery.

Interleukin 8 (IL-8)

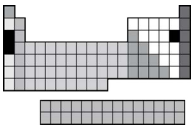
Interleukin 8 has a similar significance as IL-6. Elevated concentrations can be found in acute inflammations, but also tissue hypoxias and traumas.

Tumour Necrosis Factor (TNF-α)

Tumour Necrosis Factor is considered a marker of Systemic Inflammatory Response Syndrome (SIRS). The determination is a problematic one. Following stimulation, TNF-α is secreted for just a few hours and its half-life is <5 minutes. It is necessary to distinguish between tests determining biologically active TNF-α (trimer), and tests determining "total" TNF-α (which also detects monomers and splitting products).

A concurrent increase in IL-8 and TNF-α concentrations is indicative of monocyte activation (SIRS). If only IL-8 is elevated, this is indicative of non-immune cell activation, such as that following the attachment of bacteria or their lipopolysaccharides to endothelial cells.

A drawback is the high cost of cytokine assays, of which only the IL-6 assay as an early acute inflammation marker has been widely used.



CHAPTER 8

8. The Importance of Na, K, Cl Assays in Clinical Practice

Authors: MUDr. Milan Dastych; MUDr. Michaela Králíková, Ph.D.

Reviewer: prof. MUDr. Antonín Kazda, Ph.D.

8.1. Natrium

8.1.1. Basic Properties, Absorption

Sodium is the major extracellular cation. Sodium metabolism and transfers between the body compartments are closely associated with the transfer of water. Changes in water volume in the body are expressed by a deviation of serum Na^+ concentration from the standard level. Any loss of water in extracellular fluid (ECF) while retaining Na^+ reserves is associated with an increase in Na^+ concentration (hypernatraemia). Conversely, an increase in water volume in ECF leads to a drop in Na^+ concentration (hyponatraemia). Water and sodium metabolisms are closely related, and the rule "where sodium goes, water follows" generally applies. Sodium is absorbed in the small intestine, often in association with the absorption of other substances (amino acids, glucose), where sodium, or the sodium concentration gradient, drives the transport of other substances (co-transport). Sodium is pumped from cells to the extracellular space through the action of ubiquitary Na^+K^+-ATPase. The energy from this transport comes from ATP produced in oxidative processes in mitochondria during the metabolism of energy substrates.

Water and Na^+ absorption is completed in the ileum, where tight junctions are present which are minimally permeable to Na^+ intraluminally from the paracellular space of red blood cells. This is why any loss of ileum and the creation of a jejunostomy or jejuno-colon anastomosis leads to significant water and sodium losses and the development of dehydration of the organism. The absorption of amino acids and energy substrates (carbohydrates, fats) is usually not disturbed if the jejunum is kept.

8.1.2. Daily Requirement in Sickness and in Health

The minimum estimated daily Na^+ requirement is 50 – 85 mmol, which corresponds to 35 g of NaCl. Since sodium is excreted from the body primarily through the kidneys, the 24-hr urine Na^+ output is indicative of Na^+ intake in food and fluids. Common amounts of Na^+ output in 24-hr urine are 80 – 250 mmol, which correspond to 5 – 15 g of NaCl (1 g NaCl contains 17 mmol Na^+). Intake depends on the needs of the body and relates to sweating, the presence of diarrhoea, polyuria, drugs, etc.

The NaCl intake in this population exceeds the minimum requirement several times, being 10 – 15(20) g a day on average. This excessive intake is caused by modern food processing, preservation, and a high consumption of meat products and cheeses. Most people are able to cope with this increased Na^+ intake; only one third of "salt sensitive" people consequently develop hypertension.

A low sodium intake can lead to the development of hypo-osmolar dehydration, or, if pure water intake is maintained, to a "water poisoning" condition with clinical signs of cerebral oedema. Low sodium intake is rare under these conditions; most cases concern a relatively higher pure water intake with the potential development of hyponatraemia. Seniors with a decreased sensation of thirst and patients with anorexia nervosa are an exception.

An increased sodium requirement is typical for people with increased losses and limited absorption that cannot be made up for by increased oral intake of Na^+. These patients require infusion treatment. This is the case with the following conditions:

- Resection of small bowel, the ileum in particular;
- High-output small bowel fistulas, pancreatic fistulas;

- Intestinal malabsorptions with exudative enteropathy (inflammatory bowel diseases (IBD), coeliac disease, Whipple disease, postradiation enteritis, vasculitis, bowel ischaemia, drug-induced intestinal injury);
- Polyuric phase of renal failure with increased sodium excretion.

The NaCl requirement in these conditions increases to 15 – 25 g/day. The optimal supply should be determined based on the measurement of daily urine output; losses in sweat and stool should be estimated, and osmolality and serum Na^+ concentration determined. An important approach in estimating the Na^+ requirement in these conditions is observing the dynamics of change over time. Any reliable determination of the requirement from a single assay is impossible. It is advisable to monitor output trends and the development of natraemia, osmolality and diuresis for at least seven days, or on Day 1, 3 and 7. This monitoring is used to determine the long-term daily requirement and a decision should be made regarding i.v. substitution. I.v. hydration should be considered especially in conditions with a loss of weight and a loss of diuresis under 800 ml/24 hrs. The urine Na^+ output in these conditions drops under 20 mmol/24 hrs and the urine potassium output increases, with the potential development of hypokalaemia. The above-mentioned alterations are a sign of secondary hyperaldosteronism. If this condition develops even in the case of maximum tolerated increase in the daily sodium intake (salt addition, NaCl cps [3x2g] rehydration solutions), patients are candidates for long-term i.v. substitution.

8.1.3. Sodium Balance

Daily patient weight monitoring as a global indicator of the total water and sodium balance is an important method for determining sodium balance. Daily changes in body weight correspond to changes in body hydration, or sodium retention/mobilization. A retention or loss of 140 mmol Na^+ (8 g NaCl) corresponds to a retention or loss of 1 litre of water.

Weight loss over 3 days corresponds to water (Na^+) losses and may warn of a changing clinical situation:

- Originally normovolemic patients (healthy patients with developing diarrhoeas of varied etiology) are at risk of dehydration (lassitude, faintness, reduced blood pressure, increased heart rate, increased albumin, increased haemoglobin), and should be treated by infusion;
- In originally hypervolemic patients (oedema conditions – cardiac failure, hepatic cirrhosis, sepsis, acute pancreatitis, polytrauma), it is a favourable situation signalling the mobilization of fluids.

The clinical difference of these situations is considerable and must not be confused because treatment is exactly the opposite, infusion therapy for rehydration in the former case, and not increasing daily fluid intake, or a diuretic therapy, in the latter.

Weight increase within days should be evaluated as sodium and water retention, which may correspond to a worsening of the disease (cardiac failure, increased portal hypertension) with the occurrence of oedemas, or the development of catabolism (metabolic stress) with sodium retention.

24-hour urine sodium output is an indicator of similar importance. If the 24-hr Na^+ outputs fall under 20 mmol, two situations may have occurred:

- Sodium deficiency in extrarenal sodium losses (diarrhoea, sweating, fistulas) even with the diuresis maintained. Na^+ is reabsorbed in the kidneys due to an increased aldosterone level. K^+ and H^+ outputs increase under the influence of this hormone. Developing hypokalaemia and alkalosis are then a sign of Na^+ depletion and dehydration of the body. Hypokalaemia cannot be adjusted by simple substitution of K^+; sodium and water must also be replenished.
- The situation in oedema conditions is different in that the Na^+ content in the body is increased, even in the presence of hyponatraemia. The reason is a systemic disease that leads to circulatory alterations with renin-angiotensin-aldosterone system (RAAS) activation and Na^+ and water retention. The principle of treating this condition is quite different. It consists in limiting water and sodium intake and mobilizing fluids and Na^+ using diuretics. When choosing diuretics, it is best to combine aldosterone antagonists (spironolactone), which interfere directly with the activated RAAS axis, together with loop diuretics (furosemide).

High sodium outputs (more than 250 mmol/24 hrs) indicate sodium and fluid mobilization during transition to an anabolic condition after overcoming metabolic stress (sepsis, acute pancreatitis, polytrauma), the positive effect of diuretics, or a renal disease with sodium reabsorption disorder (tubulointerstitial nephritis).

Plasma Na concentration	135 – 145 mmol/l
Na^+ reserve in ECF	1400 mmol
Na^+ reserve in ICF	1000 mmol

Na^+ intake	140 – 260 mmol/24 hrs
Na^+ output in urine	120 – 240 mmol/24 hrs
Na^+ output in stool	10 mmol/24 hrs
Na^+ output in sweat	10 – 80 mmol/24 hrs
Total Na^+ output	140 – 260 mmol/24 hrs

Table 8.1. Sodium balance and distribution

8.1.4. Hyponatraemia

Hyponatraemia is a condition where the serum concentration of sodium drops under 135 mmol/l. It is the commonest electrolyte disorder in clinical practice. Hyponatraemia develops more often due to Na^+ dilution than as a result of Na^+ depletion. Hyponatraemia is clinically significant when Na^+ drops under 130 mmol/l, and becomes serious or potentially life-threatening at levels under 120 mmol/l. The development rate is also important for clinical seriousness; the faster the development, the more dangerous are the consequences. In the terms of therapy, it is important to distinguish whether it is acute hyponatraemia (a Na^+ decrease under 125 mmol/l within 48 hours), or chronic hyponatraemia, where a drop in Na^+ occurs within more than 48 hours. Treatment for acute hyponatraemia should be aggressive and fast, in particular if an Na^+ decrease under 120 mmol/l has clinical symptoms such as confusion, faintness, muscle twitches, headache, vomiting, or coma as a clinical impact of cerebral oedema. On the other hand, aggressive treatment of chronic hyponatraemia may have fatal consequences in the form of developing CNS demyelination (demyelinating syndrome) due to the fast transfer of water from the ICF to the ECF in the brain as a result of the osmotic gradient. Following a temporary improvement of the condition, clinical signs include disorientation, pyramidal tract symptoms with spasm of the extremities and pseudobulbar palsy developing at different rates. The final result may be vigil coma or exitus. Chronic hyponatraemia must, therefore, be treated slowly, with a maximum increase in natraemia of 8 – 10 mmol/l/24 hrs. Clinical symptoms of chronic hyponatraemia such as faintness, somnolence and muscle twitches are slighter and insidious. If properly treated, patients can survive without consequences even at Na^+ levels of around 90 mmol/l.

8.1.4.1. Pseudohyponatraemia

This condition occurs with high serum protein (more than 100 g/l) or lipoprotein (chylous serum) levels. These large molecules take up considerable space in the plasma, and the water and Na^+ content in the plasma decreases. The amount of sodium per ml of plasma is lower. Serum osmolality related to 1 kg of water is normal in pseudohyponatraemia.

Hyponatraemia may be associated with an increase in osmolality, which is due to a higher concentration of pathologically elevated osmotic substances (glucose, urea). Water is transported to the extracellular space and sodium concentration decreases during hyperglycaemia. An increase in blood sugar by 5.5 mmol/l above the upper limit of the normal range is connected with a decrease in Na^+ by about 1.6 mmol/l. In other cases, hyponatraemia is always associated with hypoosmolality.

8.1.4.2. Hyponatraemia during Dehydration (Hypovolaemia)

The principle of this condition is the fact that Na^+ losses are higher than pure water losses. Increased sodium losses have extrarenal and renal causes.

- **Extrarenal losses** are due to diarrhoea, sweating, vomiting and high output to stomata or fistulas (intestinal, biliary, pancreatic). In these situations, the Na^+ concentration in the urine is low and decreases under 20 mmol/l, while urine osmolality is high and increases over 400 mOsmol/kg H_2O;
- **Renal losses** are induced by interference with the tubular reabsorption of Na^+ due to different causes:
  - Diuretics (thiazides, loop diuretics);
  - Renal diseases leading to tubular reabsorption disorder (tubulointerstitial nephritis, polycystic kidney disease, the polyuric phase of acute renal failure);
  - Transplanted kidney;
  - Non-sodium osmotic diuresis where polyuria and increased fractional excretion of Na^+ occur due to pathologically increased excretion of osmotically active substances (glucose, ketone bodies, urea);

- Mineralocorticoid and cortisol deficiency – Addison's disease.

In these situations, the concentration of Na^+ in the urine is high, or higher than 20 mmol/l, Na^+ losses are higher than 20 mmol/d and urine osmolality is low – lower than 400 mOsmol/kg H_2O.

Gastric juices	10 – 115 mmol/l
Jejunal secretion	85 – 150 mmol/l
Ileal secretion	85 – 118 mmol/l
Pancreatic fistula	115 – 150 mmol/l
Biliary fistula	130 – 160 mmol/l
Diarrhoeic stool	30 – 60 mmol/l

Table 8.2. Na^+ concentration in intestinal secretions

8.1.4.3. Euvolaemic Hyponatraemia

This condition is due to slight solute-free water retention (undetectable using common clinical methods) without any clinical development of oedema, with ECF dilution and the occurrence of hyponatraemia. In this situation, solute--free water excretion is reduced (the fractional excretion of pure water is decreased) and sodium excretion is maintained. The commonest cause of this condition is the syndrome of inappropriate antidiuretic hormone secretion (SIADH). Due to the secretion of substances with an ADH effect, there is a lower excretion of solute-free water in this condition. However, there is no clinically manifested occurrence of oedemas, most probably because slight water retention with low intravascular volume expansion is a sufficiently strong volumoregulation stimulus for increased renal Na^+ excretion through natriuretic peptides. The diagnosis of this condition is based on serum osmolality and hyponatraemia determination (under 130 mmol/l) without clinically manifested oedemas and maintained renal Na^+ excretion with urine osmolality over 100 mOsmol/kg H_2O. It is sometimes difficult to distinguish this syndrome from the Cerebral Salt Wasting Syndrome (CSWS). In CSWS, the increased Na^+ losses in the urine are conditioned by the action of pathologically produced natriuretic peptide. Patients show clinical signs of dehydration. The treatment of SIADH consists in limiting the oral intake of fluids to under 1 litre in 24 hrs. Severe symptomatic hyponatraemia is treated by loop diuretics with Na^+ loss compensation to achieve pure water excretion with maintained sodium balance. CSWS should be treated by Na^+ substitution together with solute-free water. The causes of SIADH include pulmonary diseases (pneumonia, TBC, aspergillosis), tumours (lungs, pancreas or duodenum tumours, thymomas, lymphomas), CNS diseases (encephalitis, meningitis, cerebrovascular disease, tumours, abscesses, Guillain-Barré syndrome, psychoses), drugs stimulating ADH output (narcotics, barbiturates, carbamazepine, nicotine), or potentiating the effect of ADH (NSAID, cyclophosphamide, clofibrate).

Another cause of euvolemic hyponatraemia can be psychogenic polydipsia, accompanying disease conditions such as anorexia nervosa, neurosis or schizophrenia. If hyponatraemia is found with absent oedemas (euvolaemia), potential hypocorticalism (Addison's disease) or hypothyreosis should be taken into account.

8.1.4.4. Hypervolaemic Hyponatraemia

ECF is increased in this case, and oedemas are clinically evident in patients. The pathophysiological background for this condition is insufficient renal excretion of Na^+ and water due to circulatory alterations and organ disorders:

- Hepatic failure (RAAS activation);
- Cardiac failure (RAAS activation);
- Nephrotic syndrome with the development of hypoalbuminaemic oedemas;
- Anuric renal failure with hyperhydration;
- Severe malnutrition.

Normonatraemic hypervolaemia is frequent at the beginning of these conditions. If organ failure continues, general alterations in the body occur with the development of catabolism due to lack of energy substrastes in fasting. The manifestation of total energy depletion in the body is a decrease in Na^+K^+-ATPase activity with intracellular Na^+ and extracellular K^+ accumulation, and the development of terminal hyponatraemia.

8.1.5. Hypernatraemia

Hypernatraemia is characterized by increased Na^+ concentrations over 145 mmol/l. Clinical manifestations occur at levels over 150 mmol/l; levels over 160 mmol/l are often lethal. The clinical presentation is dominated by thirst, decreased skin turgor, CNS disorders, cephalalgia, vomiting, spasms and coma. This condition is caused most often by predominant pure water losses with clinical symptoms of dehydration. More rarely, hypernatraemia occurs in euvolaemia or hypervolaemia, usually due to iatrogenic overloading by infusions containing sodium.

8.1.5.1. Hypernatraemia Due to Lack of Water – Hypernatraemic Dehydration

This condition may be caused by **insufficient intake** or **increased losses** of pure water. Seniors with a decreased sensation of thirst are predisposed to pure water deficiency due to insufficient water intake. Even a slight water deficiency may lead to mental state alterations in elderly patients. If the situation is not assessed and treated correctly with an i.v. supply of pure water (or 5% glucose), or, conversely, if the patient is sedated by antipsychotic drugs, then dehydration and disorientation worsens, and coma with fatal consequences will occur. Fluid intake disorder can also be caused by diseases featuring swallowing problems such as dysphagia of various origin, pseudobulbar syndrome following cerebrovascular disease (CVD), head and neck injuries, or orofacial tumours.

A laboratory finding in pure water deficiency reports hypernatraemia with Na^+ over 145 mmol/l and hyperosmolality of more than 300 mOsmol/kg H_2O. The kidneys reabsorb water and sodium to a maximum extent; diuresis decreases and urine osmolality increases. Increased water losses are due to extrarenal and renal losses.

Extrarenal pure water losses with the development of hypernatraemia are caused by feverish conditions that increase perspiratio insensibilis (insensible perspiration) as well as sweat output. The sodium concentration in sweat is individual (9 – 77 mmol/l) and the increased sodium output may affect Na^+ and H_2O losses to different degrees. In the vast majority of cases, water loss from sweating is higher than sodium loss. In addition, water loss may be caused by vomiting, resulting in metabolic alkalosis, and diarrhoea, resulting in hypokalaemia, and metabolic acidosis due to bicarbonate losses.

Increased **renal water losses** occur in the polyuric phase of acute renal failure as a result of tubule function disorder in tubulointerstitial nephritis (TIN) or septic kidneys, where the renal medulla is unable to produce concentrated urine. An important cause of renal water losses is insufficient ADH production – central (neurogenic) diabetes insipidus. An insensitivity of the renal tubules to ADH is referred to as nephrogenic diabetes insipidus. The above-mentioned increased renal water losses are manifested in laboratory results by a fall in urine osmolality below the level of serum osmolality, 250 mOsmol/kg H_2O being considered the limit; the values drop beneath 100 mOsmol/kg H_2O in maximum water polyuria, fractional pure water excretion increases over 2 % and polyuria occurs (daily diuresis over 2500 ml, diuresis reaches 8000 – 10000 ml/24 hrs in extreme cases of diabetes insipidus).

Body temperature	Perspiratio insensibilis (ml/24 hrs)
37.2	550
37.8	600
38.3	800
38.9	900
39.4	1000

Table 8.3. Dependence of perspiratio insensibilis (insensible perspiration) on body temperature

8.1.5.2. Hypernatraemia with Euvolaemia or Hypervolaemia

Hypernatraemia without signs of dehydration occurs in rare cases. Most cases involve an increase in sodium intake in the form of infusion solutions, where therapeutic use is excessive (e.g. saline = 0.9% NaCl solution contains 154 mmol/l Na^+ = 9 g NaCl). In addition, this condition may be caused by primary hyperaldosteronism or Cushing's syndrome.

8.2. Kalium

8.2.1. Basic Properties

K^+ is the major intracellular cation. Potassium concentrations in the ICF range from 100 to 160 mmol/l. A person weighing 70 kg has about 4000 mmol K^+. The high K^+ concentration in the ICF is provided by Na^+/K^+-ATPase. That is why ATP availability and the energy balance of the body are of crucial importance for the proper distribution of K^+. Na^+/K^+-ATPase transports 3 Na^+ ions from the cell and transports 2 K^+ ions into the cell. It therefore generates the electronegativity of the intracellular space, which is caused by, among other factors, for example differing cell membrane permeability to Na^+ and K^+ (the cell membrane is more permeable to K^+ than Na^+). The result of this ion distribution is the genesis of resting membrane electric potential, which is fundamental for many physiological processes such as neuromuscular transmission, nerve impulse propagation or the function of the heart. Given the fundamental role of K^+ in the generation of resting membrane electric potential, changes in K^+ concentration in body fluids have far-reaching and often fatal consequences.

Hypokalaemia may cause muscular weakness and slower nerve impulse propagation, which may become clinically manifested by reduced ventilation (among other effects), resulting in respiratory insufficiency and the necessity of artificial lung ventilation.

The function of smooth muscles is also reduced at lower potassium levels, resulting in paralytic ileus.

The function of the heart is pathologically affected by both hypokalaemia as well as hyperkalaemia with its potential development of malignant arrhythmias – ventricular fibrillation or asystoles. It is evident from the above that maintaining the proper K^+ level in the body compartments is of major importance.

8.2.2. Potassium Requirement, Balance, Relation to ABB and Nutritional Status

The daily requirement of K^+ ranges from 1 to 2 mmol/kg per day. This requirement is increased during anabolism, when insufficient K^+ substitution results in serious hypokalaemia. In physiological conditions, potassium is excreted via the kidneys and the regulation of excretion is related to the effect of aldosterone.

The anion counterpart of intracellular potassium is phosphate ions in the form of organic macromolecular phosphates. Although macromolecular phosphates do not pass the cell membrane, they may affect the K^+ level. Their intracellular concentration decreases as they disintegrate and deplete during catabolism, K^+ is released into the ECF due to decrease of the negative charge of the ICF. During anabolism, however, when the synthesis of macromolecular phosphates increases, potassium flows to the cells due to an increase of the negative ICF charge. During anabolism, intracellular potassium concentration increases and the potassium pool is replenished. Severely undernourished people have an enormous intracellular potassium deficiency, which may reach 800 – 1000 mmol.

If the patient is re-alimented, a sufficient substitution of potassium of 250 – 300 mmol/24 hrs is required along with the substitution of phosphorus and magnesium. Failure to maintain an increased supply of intracellular ions at the beginning of nutritional intervention in seriously undernourished patients may lead to refeeding syndrome and the potential death of the patient. The cause of death is usually muscular weakness, respiratory failure, coma and cardiac arrhythmia.

The K^+ transport between the ICF and ECF may also take place in exchange for H^+, which points to the relation between the acid-base balance (ABB) and the K^+ concentration in the ICF and ECF. Changes in the ABB lead to transfers of K^+, whereby MAC supports the development of hyperkalaemia, and MAL supports the development of hypokalaemia. A drop in pH of 0.1 increases kalaemia by 0.4 mmol/l on average.

Plasma concentration	3.8 – 5.2 mmol/l
Intracellular concentration	100 – 160 mmol/l
Urine concentration	30 – 80 mmol/l
Intracellular reserve	3200 mmol
Extracellular reserve	60 mmol
Intake	50 – 100 mmol/24 hrs
Output	50 – 100 mmol/24 hrs

Table 8.4. Potassium balance and distribution

8.2.3. Hypokalaemia

Hypokalaemia is defined as a drop in serum K^+ concentration under 3.5 mmol/l, serious hypokalaemia under 3 mmol/l, and life-threatening hypokalaemia under 2.5 mmol/l. The rate of change is important for clinical manifestation. The faster hypokalaemia develops, the more serious is the impact. Hypokalaemia may have serious clinical manifestations such as malignant cardiac arrythmia, respiratory muscle paralysis with respiratory insufficiency, or a decrease in smooth muscle activity with the development of paralytic ileus.

8.2.3.1. Extrarenal Hypokalaemia

The 24-hr urine K^+ output is used to differentiate between extrarenal and renal hypokalaemia. 24-hr urine K^+ outputs under 20 mmol are indicative of extrarenal origin. Extrarenal hypokalaemia is caused by **low potassium intake** in patients with anorexia nervosa, fasting people, insufficient K^+ supply during parenteral (PN) or enteral nutrition (EN) of undernourished people. **Increased extrarenal losses** are caused by vomiting, gastric juice losses with an inserted nasogastric tube, fistulas, diarrhoea and stomata. Losses from the proximal part of the digestive tract lead to the development of MAL with decreasing K^+; losses in the distal part of the digestive tube (ileum, colon) lead to decreased HCO^{3-} and K^+, resulting in hypokalaemia with MAC. Increased potassium losses may also be caused by overuse of laxatives.

8.2.3.2. Renal Hypokalaemia

Renal losses are usually caused by renal tubular acidosis, the polyuric phase of acute renal failure, chronic renal failure, tubulointerstitial nephritis, Gitelman syndrome, Bartter syndrome and magnesium deficiency. Hypokalaemia due to major renal losses may also be caused by chronic treatment by or abuse of saluretics (furosemide, thiazides). Long-term hypokalaemia induces hypokalaemic nephrosis with tubule involvement, which leads to elevated K^+ losses in the urine. Increased renal losses are also caused by a primary or secondary increase in the aldosterone level. Causes of primary hyperaldosteronism include: adrenal adenoma, increased ectopic ACTH production in tumours, increased ACTH production in hypophyseal adenoma. Secondary hyperaldosteronism occurs in cardiac failures and hepatic cirrhosis as a manifestation of decreased intravascular volume or cardiac output. In these cases, the use of aldosterone antagonists (spironolactone) is recommended to treat hypokalaemia.

Clinical signs of hypokalaemia are listed above. Flattened T-waves, a pronounced U-wave, ST depression and premature ventricular contractions (PVC) can be found on the ECG waveform. Ventricular fibrillation and muscular rhabdomyolysis are dangerous complications. Renal tubules become damaged in long-term hypokalaemia.

8.2.3.3. Treatment for Hypokalaemia

The kalaemia level should always be evaluated in relation to the ABB, water balance and energy balance of the body (catabolism vs. anabolism). Potassium output in urine over 12 (or preferably 24) hours should be determined to make the right decision about treatment. Kalaemia determination alone reveals very little about the cause and does not allow the right therapeutic strategy to be chosen. Steps to take for hypokalaemia treatment are as follows:

- Acid-base imbalance treatment;
- Supply of energy substrates through appropriately selected nutritional intervention and energy balance adjustment; intracellular K^+ deficiency cannot be adjusted without concurrent synthesis of intracellular phosphate macromolecules – an anion partner of K^+ in the ICF;
- Removal of dehydration of the body that leads to secondary hyperaldosteronism with increased renal excretion of K^+, H^+ and reabsorption of Na^+ and water. This hypokalaemia cannot be treated by mere K^+ substitution without adequate hydration;
- Magnesium substitution, if a deficiency is found;
- Increased K^+ supply by infusions. Determine the daily dose depending on the urine K^+ output and estimated deficiency under regular monitoring of serum K^+ concentrations. The rate of substitution is usually 50 mmol/6 hrs, and the dose 100 – 200 mmol/24 hrs. In extreme situations, substitutions may be about 300 mmol/24 hrs. For K^+ levels of 1.4 mmol/l with ECG alterations, the rate of supply can be 40 mmol/hr with regular K^+ level monitoring every 2 hours. If administration through a peripheral vein is used, the potassium solution concentration should not exceed 30 mmol/l and the rate 10 mmol/hr;
- Conditions with elevated aldosterone levels should be treated by aldosterone antagonists. Hyperaldosteronism is manifested by a decrease in the U-Na/U-K concentration ratio to under 1. In other words, the urine potassium output exceeds the sodium output;
- Removal of the cause of potassium losses.

8.2.4. Hyperkalaemia

Slight hyperkalaemia is defined by a K^+ concentration of 5.5 – 6.5 mmol/l, medium hyperkalaemia by 6.5 – 7.5 mmol/l, and severe hyperkalaemia by more than 7.5 mmol/l. Like hypokalaemia, hyperkalaemia also has renal and extrarenal causes.

8.2.4.1. Extrarenal Hyperkalaemia

24-hr urine K^+ outputs are used to find the cause of hyperkalaemia. With an extrarenal cause, 24-hr urine K^+ outputs exceed 50 mmol, and the urine K^+ concentration increases to more than 100 mmol/l with severe catabolism. Extrarenal hyperkalaemia occurs as a result of K^+ redistribution between the ICF and the ECF. Potassium discharge from cells is caused by intracellular acidosis from various causes and catabolism during fasting, which results in the loss of phosphate macromolecules. The total K^+ reserve in the body may become decreased in these situations. An increased supply of K^+ in i.v. solutions due an incorrect determination of the balance may lead to artificial hyperkalaemia.

8.2.4.2. Renal Hyperkalaemia

This type of hyperkalaemia is caused by an absolute decrease in glomerular filtration in acute or chronic renal failure with a decreased renal K^+ output and concomitant acidosis. Another cause is diseases and conditions affecting the tubular functions of the kidneys, specifically a decrease in tubular K^+ secretion. These are several endocrinopathies with a low aldosterone level (Addison's disease, hypoaldosteronism, hypopituitarism) and medicines (ACEI, Cyclosporin A, NSAID, spironolactone). In these cases, potassium outputs in the urine decrease under 50 mmol/24 hrs, or under 30 mmol/l, respectively.

Major clinical signs of hyperkalaemia include cardiac arrythmia (bradyarrythmia, asystole) and paresthesias of the muscles. In most cases, the patient's clinical picture is given by the underlying disease that caused the disorder. The ECG shows peaked T waves, widened QRS complexes and bradycardia.

8.2.4.3. Treatment for Hyperkalaemia

In addition to determining serum K^+ concentration, for correct hyperkalaemia treatment S-creatinine, S-urea, ABB and 24-hr urine K^+ outputs also have to be determined. Urine K^+ outputs allow renal and extrarenal causes of the condition to be distinguished. Excessive input or significant catabolism lead to a urine K^+ concentration increase over 100 mmol/l, while renal insufficiency leads to a urine K^+ concentration decrease under 30 mmol/l. An important moment in hyperkalaemia treatment is to discontinue the use of drugs that lead to potassium retention. The intensity of hyperkalaemia treatment depends on serum K^+ concentration, the dynamics of the increase, the clinical condition and ECG alterations. Dialysis or conservative treatment is indicated based on these indicators. Acute haemodialysis is indicated for K^+ levels of more than 7 mmol/l, or K^+ levels of more than 6 mmol/l with clinical signs not responding to the conservative treatment.

The conservative therapy of acute hyperkalaemia includes:

- Administration of Calcium gluconicum 10 – 30 ml i.v.;
- Administration of bicarbonate 8.4% 80 ml. i.v.;
- Administration of 500 ml 10% glucose $^+$ 10 U insulin;
- Ion exchangers (Resonium 100 g in the form of an enema).

The conservative therapy of slight chronic hyperkalaemia includes:

- Decreasing daily potassium supply to under 1.5 g (under 40 mmol);
- Enhanced excretion through the use of loop diuretics (furosemide);
- Oral bicarbonate treatment in MAC;
- Sufficient hydration;
- Ruling out drugs that decrease K^+ excretion;
- Ion exchangers (Resonium 15 g 3x a day).

8.3. Chloride Anion

Chloride anion is a major extracellular space anion and its distribution space corresponds to 20% of body weight. An increase or decrease in the chloride content is accompanied by changes in the Na^+ reserve because chloride is an anion partner of sodium. Cl^- is also in electroneutral balance with bicarbonate, so changes in the chloride concentration are associated with ABB alterations. Hyperchloridaemia causes acidosis and hypochloridaemia alkalosis.

8.3.1. Hyperchloridaemia

Hyperchloridaemia is accompanied by sodium retention and occurs due to inadequate pure water retention and ECF expansion. If the ECF expands together with Na^+ retention, the Cl^- reserve grows but the serum concentration remains unchanged. Metabolic acidosis and hypernatraemia develop due to hyperchloridaemia. This is usually caused by increased chloride supply in the form of infusion solutions, or decreased excretion in hepatic cirrhosis, cardiac failure or renal failure. Hyperchloridaemia is also found in patients with ureteroileostomy following cystectomies, as a result of increased absorption of chlorides excreted in the urine by the neovesica made from the intestine. Oral bicarbonate treatment should be used in this case. In the other cases, treatment consists in reducing the supply of chlorides (most often in infusions) and enhancing their excretion using saluretics.

8.3.2. Hypochloridaemia

As in hyperchloridaemia, if Na^+ and water deficiency are concurrent, thereby leading to a lower ECF volume, hypochloridaemia will not develop but the Cl^- reserve in the body will decrease. The chloride concentration in the ECF does not change in this situation.

If the ECF volume is maintained, hypochloridaemia will develop together with hypochloremic alkalosis, hypokalaemia and, in most cases, hyponatraemia. This condition is most often caused by the loss of intestinal secretions by vomiting, losses by fistulas in the proximal part of the GIT, or withdrawal of gastric content through a NG tube. Furthermore, hypochloridaemia occurs due to increased excretion via the kidneys – the polyuric phase of renal failure or massive saluretic treatment. Hypochloridaemia treatment consists primarily in removing the inciting cause (vomiting, fistula, inadequate diuretic treatment) and sufficient chloride substitution in the form of saline infusion solutions. If there is a marked ECF expansion, arginine chloride solutions are recommended since they do not contain Na^+ and do not cause fluid retention. If hypokalaemia is present, balanced solutions such as Ringer, Ringerfundin or Darrow's solutions containing K^+ are available.

Plasma concentration	100 mmol/l
Reserve in ICF	1400 mmol
Reserve in ECF	1000 mmol
Intake	140 – 260 mmol 24/hrs
Output in urine	120 – 240 mmol/24 hrs
Output in stool	10 mmol/24 hrs
Output in sweat	10 – 80 mmol/24 hrs
Total output	140 – 260 mmol/24 hrs

Table 8.5. Chloride reserve and distribution in the body

CHAPTER 9

9. Metabolism of Calcium, Phosphorus and Magnesium

9.1. Metabolism of Calcium, Phosphorus and Magnesium – Preparation

Author: MUDr. Richard Pikner, Ph.D.

Reviewers: prof. MUDr. Antonín Kazda, DrSc., MUDr. Roman Cibulka, Ph.D.

9.1.1. Introduction

Together with sodium, potassium and chlorides, calcium (Ca), phosphorus i.e. phosphates HPO_4^{2-} (P) and magnesium (Mg) form a substantial part of human body's inorganic component. Regulation of calcium, phosphorus and magnesium is humoral, and calciotropic hormones play a role in this: parathormone (PTH), parathormone-related protein (PTHrP), vitamin D, calcitonin (CT) and fibroblast growth factor 23 (FGF23). The hypophosphaturic effect in the kidneys takes place through the interaction of FGF23 and the Klotho protein as a local co-factor. This chapter provides an overview of the physiology and pathophysiology of Ca, P and Mg metabolism.

In conclusion the key information summarizes essential information from the entire chapter.

9.1.2. Calcium, Ca

A healthy person's body contains approximately 1.0 to 1.3 kg Ca; 99 % is found in the bones and teeth, 1 % in soft tissues, and 0.1 % in extracellular space. The concentration of Ca in intracellular fluid (ICF) is about 10,000 lower than in extracellular fluid (ECF). In plasma, 50 % of Ca exists in a free (ionised) form bound to proteins, mainly albumin (40 %), and 10 % of Ca in complexes with citrate and bicarbonate. Ca contained in complexes and ionised Ca represent a diffusible, biologically active fraction. This biologically active Ca component is strictly regulated by humoral mechanisms – calciotropic hormones.

Ca concentration is greatly affected by 2 factors:

- Albumin concentration: The albumin level affects the overall Ca level in plasma. Patients with hypoalbuminaemia (↓ albumin) have lower overall calcium concentrations but the ionised, biologically active fraction is usually normal. This is relatively common cause of hypocalcaemia in hospitalised patients without clinical symptoms;
- Changes in pH: Acidosis reduces the calcium bond to albumin and increases the ionised calcium fraction, while alkalosis reduces the ionised calcium fraction and invokes clinical symptoms connected to hypocalcaemia. A typical example is hyperventilation tetany, where quickly induced respiratory alkalosis causes hypocalcaemia with tetany and limited activity of respiratory muscles, which deepens dyspnoea. CAVE: For this reason acute treatment involves normalisation of alkalaemia, not calcaemia.

Physiology

Calcium is an important regulator of many biological processes, which is why the ionised calcium level is very strictly controlled and kept within a fairly narrow reference range.

Regulation mechanisms are set up to maintain serum normocalcaemia.

The small intestine (absorption), bone tissue (deposits) and kidneys (excretion) are involved in regulating the Ca level.

Calcium is important for:

- Bone mineralization;

- Blood clotting;
- Neuromuscular transfer (creation and propagation of impulses);
- Muscle contraction;
- Cytoskeleton function (endo and exocytosis, cell division);
- Intracellular signal transfer – 2nd messenger;
- Enzyme activation (protein kinase, phosphorylase).

Daily metabolism of calcium:

An optimum daily intake ranges from 1000 to 1500 g depending on growth, pregnancy or lactation.

Absorption takes place in the small intestine, is greatest in the duodenum in the presence of acid pH, and gradually decreases towards the ileum. The process of intestinal resorption is partly a passive paracellular process following the concentration gradient, but mostly an active process (Ca bond on the specific carrier calbindin) via enterocytes. Active calcium resorption is stimulated by the presence of active vitamin D (1,25 $(OH)_2D$). Some food constituents (fat, oxalates) have an adverse effect on Ca absorption.

Absorbed Ca enters the ECF where it is exchanged for Ca from bones. Ca is deposited in the period of growth or with free bone capacity – positive calcium balance. The exchange is controlled by calcitropic hormones, PTH in particular.

Ca is excreted in the kidneys. Most filtered Ca is absorbed passively in the proximal tubule and the loop of Henle. Active regulation takes place in the distal tubule – about 8 – 10 % of Ca. This is an active process controlled by calcitropic hormones, predominantly by PTH with minimal involvement of calcitonin.

Excretion is also affected by drugs such as diuretics. Loop diuretics (furosemide) reduce the tubular resorption of both Na and Ca, while thiazide diuretics boost Ca resorption with increased secretion of Na and Cl.

Metabolic acidosis (excretion) of hydrogen ions also prevents Ca resorption in the distal tubule.

9.1.3. Phosphorus

The human body contains about 700 – 800 g of phosphorus, 85 % of which is deposited in bones, less than 1 % is in the ECF and the rest (14 %) is deposited in tissues. The phosphate anion is the most important intracellular anion; its ICF concentration is 100 times as high as its ECF concentration. Around 88 % of P contained in the plasma is 53 % of diffusible P in free (ionised) form and 35 % in the form of complexes. The rest (12 %) is non-diffusible P bound to proteins.

Physiology:

Phosphorus not only forms an integral part of the skeleton, but is also essential for key cell structures. It forms part of nucleic acids, energy metabolism (high-energy phosphates), buffer systems and the phospholipid membrane.

Absorption takes place in the small intestine and is the greatest in the jejunum. Intestinal resorption is proportionate to the amount of phosphates in food and is stimulated by the presence of 1,25$(OH)_2D$ and its stimulating effect on the Na-P co-transporter.

Like Ca, absorbed P enters the ECF where it is exchanged for P from bones which are a major site of phosphate deposits in the body. P is excreted in the kidneys. 75 – 85 % is absorbed in the proximal tubule, 10 – 15 % in the distal tubule and 5 – 10 % in the collecting duct. It is an active process involving specific Na-P co-transporters, whose activity is controlled by the presence of FGF23 and its specific renal Klotho co-factor. The presence of FGF23-Klotho subsequently leads to increased phosphaturia.

Although phosphorus is essential for many intracellular processes, its permanently elevated serum concentration is toxic for the body. This accelerates the cell aging process and is, among other things, responsible for vascular calcifications in chronic dialysed patients.

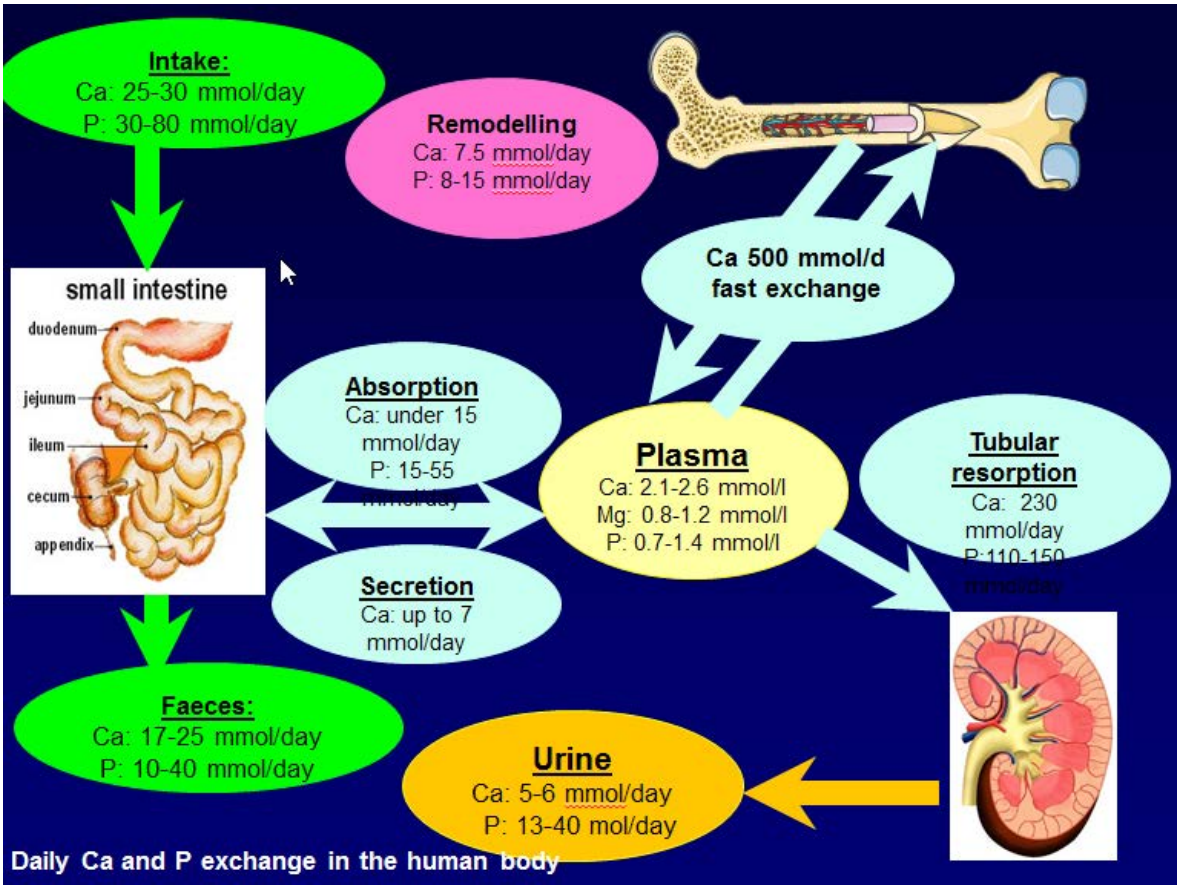


Figure 9.1. Summary of daily calcium and phosphorus metabolism

9.1.4. Magnesium

The human body contains only about 24 g of Mg, 50 % of which is in the skeleton, and the remaining 50 % is in the ICF, mostly in muscles. Less than 1 % is found in the ECF. About 70 – 80 % of Mg in plasma represents diffusible ionised Mg, and 1 – 2 % is bound in complexes. The remaining 20 – 30 % is non-diffusible Mg bound to proteins.

Physiology

Mg is the 2nd most common cation in the ICF, is involved in about 300 enzymatic reactions (energy metabolism, nucleic acid synthesis, cytoskeleton, membrane stability). Mg plays a major role in modulation, maintaining homeostasis and intracellular calcium concentration. In addition, Mg mediates immunological reactions and PTH production and secretion, regulates smooth muscle tonus and skeletal muscle relaxation. The recommended daily allowance is 320 – 420 mg. Absorption takes place actively in the small intestine and is likely affected by the presence of $1,25\ (OH)_2D$.

Mg is excreted in the kidneys. 15 – 20 % is resorbed in the proximal tubule, 60 – 70 % (passively) in the cortical ascending arm of the loop of Henle, and 5 – 10 % actively in the distal tubule, probably under similar regulating effects as Ca.

9.1.5. Parathyroid Hormone (PTH)

Polypeptide – basic regulating hormone of the calcium-phosphate metabolism.

PTH is a polypeptide consisting of 84 amino acids. It is produced in the parathyroid glands as a 115-amino acid pre--pro PTH and then cleaved to the 90-amino acid PTH. The last cleavage in the Golgi apparatus either produces an intact (biologically active) 1-84 PTH, or a 7-84-amino acid (biologically inactive) fragment. Both 1-84 and 7-84 are released into the circulation. The intact 1-84 PTH is around 30 – 40 % of the total PTH released into the circulation. The serum ionised calcium level regulates PTH secretion through negative feedback; the signal is transferred through the specific calcium sensing receptor (CaSR). PTH production in the parathyroid glands is affected by the phosphorus level (stimulation), whereas hypomagnesemia and hypermagnesemia and 1,25 (OH)2D suppress the PTH production. Degradation takes place in the liver and kidneys by cleaving the 33-34-amino acid and 36-37-amino acid molecules to form other

inactive PTH fragments. The intact 1-84 PTH biological half-life is about 5 minutes.

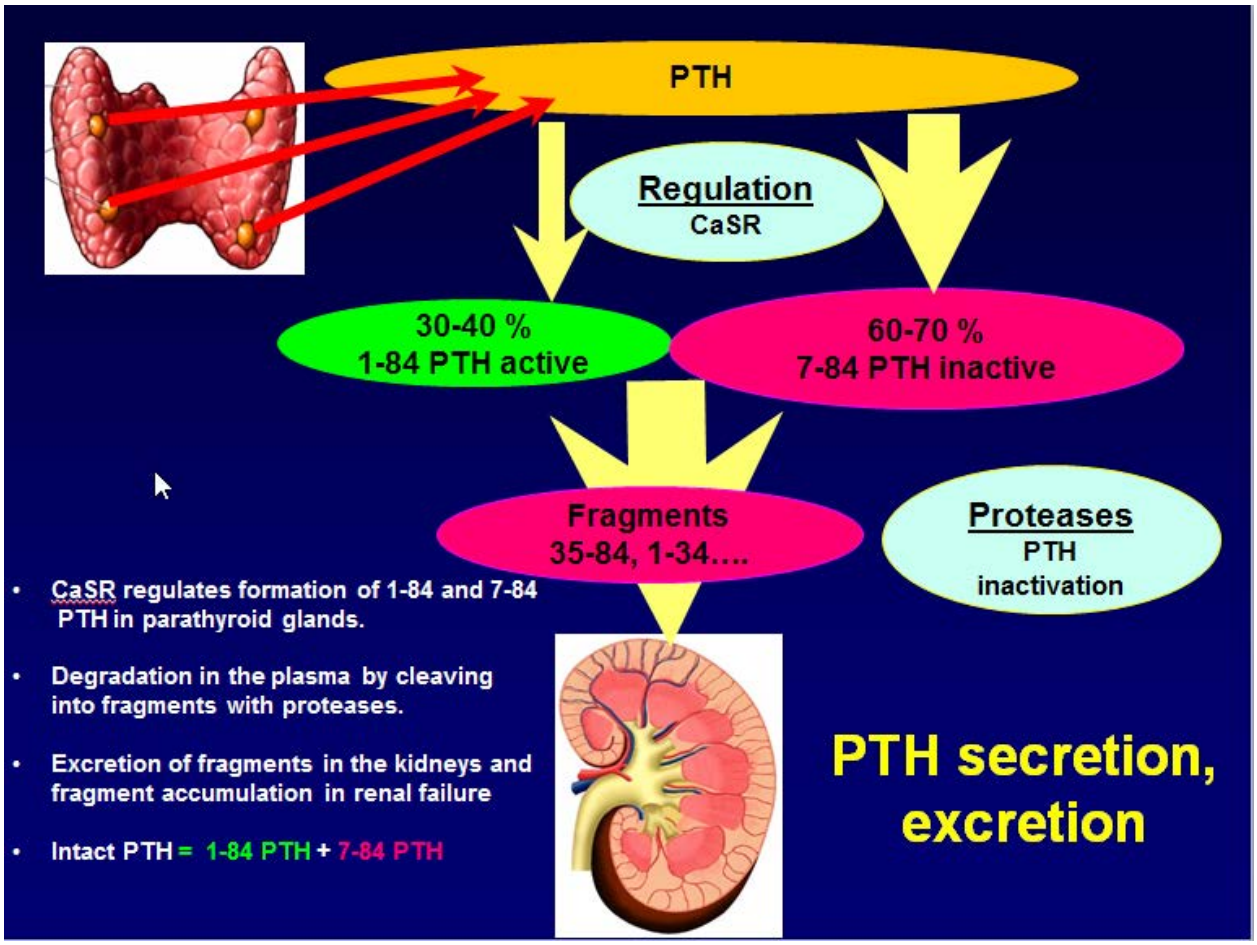


Figure 9.2. PTH metabolism

Physiology

The main effect of the PTH is an increase to plasma Ca concentration.

The effect is takes place through 2 types of receptors: PTH1 and PTH2. PTH 1 is common for the PTH and PTH-related protein (PTHrP – see below) and is found in bones and kidneys.

- In the bone, PTH induces a change in OPG-RANKL production indirectly through osteoblasts with indirect osteoclast stimulation (osteoresorption) and the release of Ca into the ECF. *(For description of local regulation of osteoblasts and osteoclasts see Bone Metabolism.);*
- PTH increases reabsorption of Ca and excretion of P in the kidneys and oosts the production of alpha-1--hydroxylase (enzyme forming active 1,25 $(OH)_2D$).

Circadian rhythm with a maximum at 3:00 a.m. and a minimum at 11:00 a.m.

9.1.6. PTH-Related Protein (PTHrP)

PTHrP is a peptide with a gene comprising 9 exons, and their different expression levels are responsible for the formation of 3 possible isoforms (139 amino acids, 141 amino acids and 173 amino acids). Owing to the homology in the first 13 amino acids, PTHrP binds to the same PTH/PTHrP-1 receptor. Unlike PTH, PTHrP is produced by many cells and its mechanism of action is primarily autocrine or paracrine. Known functions of PTHrP include:

- Influence on enchondral ossification – blocks chondrocyte maturation (bone growth in the prenatal period);
- Breastfeeding – production of PTHrP in the mammary gland to provide an adequate supply of calcium in breast milk. PTHrP systemically increases osteoresorption and is locally involved in transporting Ca to breast milk;
- Transepithelial transport of calcium in the placenta;
- Unlike PTH, PTHrP does not stimulate 1-hydroxylase in the kidneys.

Increased production of PTHrP in adults with malignant diseases is primarily linked to hypercalcaemia. It is a paraneoplastic overproduction of PTHrP with tumour cells (bone metastases) leading to the formation of osteolytic bone metastases.

9.1.7. Vitamin D – 1,25 $(OH)_2$D (Calcitriol)

Ranks among hormones in terms of its mechanism of action since it has its specific receptor.

Vitamin D ranks among fat-soluble hormones. Vitamin D exists in two basic forms: vitamin D2 (ergocalciferol) and vitamin D3 (cholecalciferol). Ergocalciferol arises from UV-B irradiation of fungal steroid ergosterol, and naturally occurs as a rare form of vitamin D, but is a major vitamin D produced in the pharmaceutical industry. Cholecalciferol arises from UV-B irradiation of a 7-dehydrocholesterol precursor, high levels of which are found in the skin. The time for reaching the maximum D3 precursor level in the skin depends on the strength of UV radiation, time of exposure and skin pigmentation. In addition, UV radiation prevents overproduction of cholecalciferol, since it is photo-degraded into inactive metabolites (lumisterol, tachysterol) following maximum saturation of tissue. When transported into the circulation, 85 % of ergosterol and cholecalciferol is bound to vitamin D-binding protein (DBP) and 15 % to albumin. In addition to conversion in the skin, vitamin D is absorbed in the intestine from food. The liver is the site of vitamin D conversion into 25(OH) vitamin D (calcidiol). 1,25 $(OH)_2$D (calcitriol) is an active hormone produced by hydroxylation of calcidiol (25 (OH)D) with α1-hydroxylase. Around 15 % of calcitriol is produced in the kidneys and is supplied to tissues without their own 1-hydroxylase such as enterocytes or pancreas. 85 % of calcitriol is produced by hydroxylation right in the peripheral tissues with their own 1-hydroxylase (epithelial cells, nerve and immune systems, prostate, breast, colorectum). Its role here is primarily autocrine and paracrine, and it is responsible for non-skeletal effects of vitamin D.

24,25-$(OH)_2$D is a degradation product of the vitamin D metabolism without any biological function.

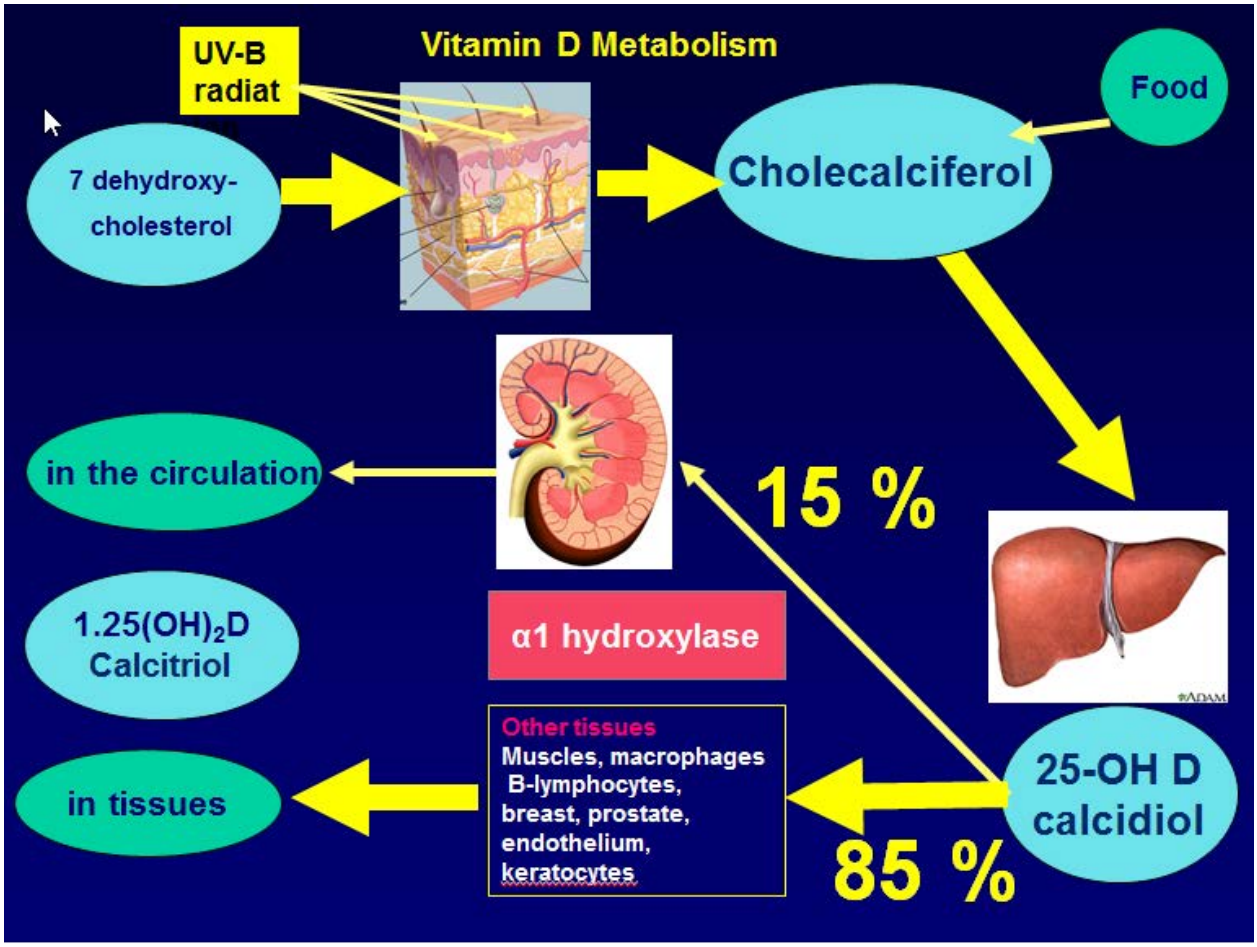


Figure 9.3. Vitamin D production

Physiology

The main function of calcitriol produced by the kidneys is calcium and phosphate absorption regulation in the small intestine.

More broadly, vitamin D acts through the specific intracellular vitamin D-receptor (VDR) and subsequently the retinoid X receptor (RXR). The calcitriol-VDR-RXR complex binds to specific DNA sites known as vitamin D response elements (VDRE). This bond initiates the transcription of specific DNA and RNA with corresponding protein production. Calcitriol is presently assumed to have an influence on the expression of around 2000 genes in the human body, meaning vitamin D has skeletal (calcium-phosphate metabolism) and non-skeletal effects.

Skeletal effects of vitamin D:

- Intestine: Increases calcium and phosphorus absorption;
- Bone: Minimal importance;
- Kidneys: Stimulates PTH receptor and Ca-carrying protein production – increases calcium reabsorption.

Non-skeletal effects of vitamin D:

- Stimulates insulin secretion;
- Immunomodulatory and antibacterial function;
- Affects muscle fibre formation and calcium metabolism in muscles;
- Suppresses the production of renin;
- Improves differentiation and supports cell apoptosis, suppresses proliferation, angiogenesis (prostate, breast, epithelia, colorectum).

Main forms in the circulation:

- 25 OH vitamin D (calcidiol) – active calcitriol precursor and a reserve prohormone for conversion in peripheral tissues. **Indicator of vitamin D supply status;**
- 1,25 $(OH)_2D$ (calcitriol) – active vitamin, its serum level correlates to production in the kidneys.

9.1.8. Calcitonin (CT)

Hormone, 32-amino acid peptide is produced by parafollicular C-cells of the thyroid; involved in controlling the homeostasis of calcium-phosphate metabolism and is a sensitive tumour marker for medullary thyroid cancer.

Physiology

CT primarily inhibits osteoclast-mediated bone resorption. Calcitonin secretion is controlled by the serum ionised calcium level. Calcitonin also has a slight renal effect where it inhibits reabsorption of Ca and P. Despite this physiological effect, calcium-phosphate homeostasis is not greatly dependent on calcitonin. After undergoing a complete thyroidectomy medullary cancer, patients display practically normal calcium levels.

9.1.9. Fibroblast Growth Factor 23 FGF23 and Klotho Protein

These are molecules with a primary effect on the concentration of phosphate. The main source of FGF23 is the bone tissue (1,25 $(OH)_2D$ stimulates its production), whereas Klotho protein is produced in the kidneys. Klotho protein mediates the tissue-specific effect of FGF23 in the kidneys which makes it a co-factor for the FGF23 function. FGF23 blocks protein production in renal cells responsible for reabsorbing phosphates. It has a substantial phosphaturic effect while blocking 1-α- hydroxylase production in the kidneys, thus reducing the 1,25 $(OH)_2D$ level as well as resorption of phosphates and calcium in the small intestine.

9.1.10. Regulation of Calcium and Phosphate Metabolism

The regulation of calcium and phosphate metabolism is based on mutual interaction with PTH-1,25 $(OH)_2D$-CT and FGF23-Klotho.

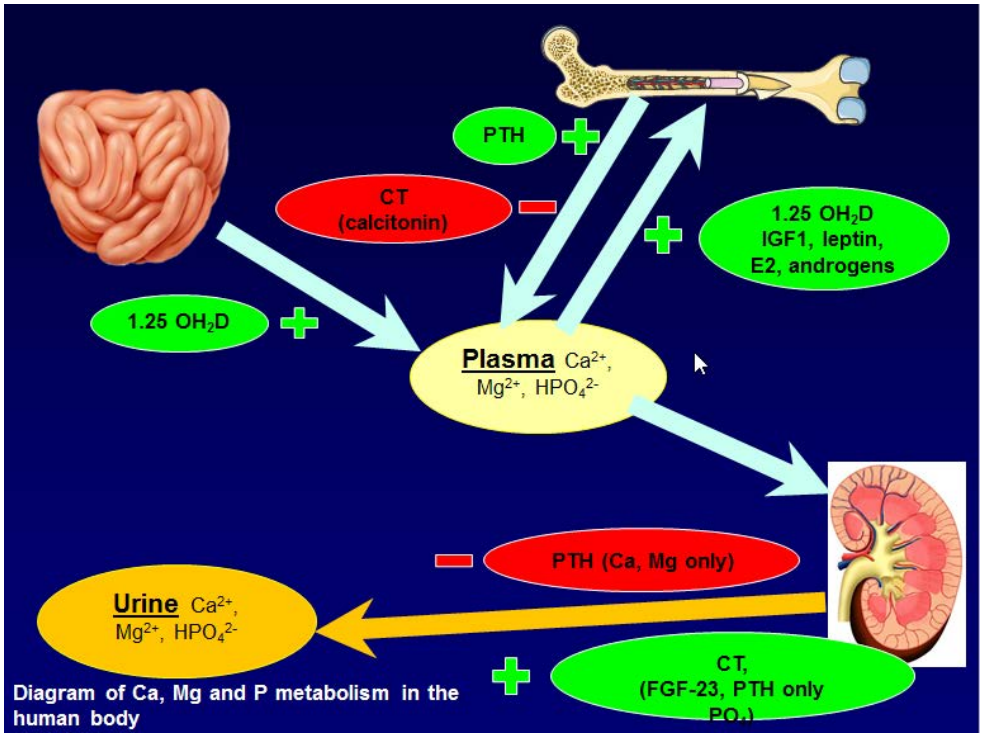


Figure 9.4. Basic mechanisms of Ca, Mg and P concentrations

Mg metabolism regulation is practically identical to calcium regulation, and is also likely controlled by the PTH-1,25 $(OH)_2D$ system. Serum Mg concentration affects PTH secretion. Both a lack and surplus of magnesium block PTH production.

The basic goal of homeostasis is to maintain normocalcaemia and normophosphataemia. Physiologically, it is more difficult for the human organism to provide sufficient calcium levels and excrete excess phosphorus. Magnesium is above all an intracellular cation, so except for severe deficiency or surplus forms, serum Mg concentration does not reflect its current deficiency or surplus in tissues.

PTH, 1,25 $(OH)_2D$ and calcitonin are involved in calcaemia regulation.

Hypercalcaemia is induced by PTH (↓ EF Ca in the kidneys and ↑ osteoresorption). At the same time, magnesium directly stimulates 1,25 $(OH)_2D$ production, which increases calcium resorption from food in the small intestine.

Conversely, 1,25 $(OH)_2D$ partially inhibits PTH production. However, the main regulator of PTH production is the ionised calcium level through the calcium sensing receptor (CaSR).

Hypocalcaemia is induced by calcitonin (CT) which, as the only hormone, directly blocks osteoclast function and thereby resorption in bones. CT slightly increases calciuria and phosphaturia in the kidneys. Nonetheless, its biological effect on the human body is insignificant, because the main causes of hypocalcaemia are a deficient supply of Ca in food or a parathyroid gland disorder (PTH deficiency).

The body retains normocalcaemia in 3 ways:

- Sufficient **supply from food** supported by 1,25 $(OH)_2D$. 1,25 $(OH)_2D$ concurrently stimulates the production of FGF23, which supports phosphate production and blocks 1,25 $(OH)_2D$ production, thus finely balancing the whole system;
- **Release from bones** which makes up for temporary fluctuations of ionised calcium levels by increasing or inhibiting PTH secretion. The body makes up for a long-term calcium deficiency by increasing PTH secretion. This results in raised osteoresorption leading to damaged bone integrity. Concurrently released P from bones is excreted by a direct phosphaturic effect of PTH in the kidneys and owing to increased FGF23 production in the bones. At the same time PTH stimulates 1,25 $(OH)_2D$ production as well as the effort for a physiological increase of calcaemia;
- **Retention of calcium in the kidneys** – indirect effect of PTH: 1,25 $(OH)_2D$ via the stimulation of PTH receptor production.

If calcium is not present in food, increasing 1,25 $(OH)_2D$ production does not help.

(It's not possible to select something from nothing.)

FGF23, PTH and 1,25 $(OH)_2D$ are involved in phosphataemia regulation.

Hypophosphataemia is primarily induced by the FGF23-Klotho complex in the kidneys (phosphaturic effect), which concurrently blocks 1,25 $(OH)_2D$ production, thus also reducing phosphate resorption in the small intestine. On the one hand PTH increases phosphataemia by osteoresorption from bones, while on the other hand also reduces phosphataemia by directly stimulating phosphaturia in the kidneys.

Hyperphosphataemia is primarily induced by an inability to excrete phosphorus from the body in diseases such as chronic renal failure, and by increasing intake from food in the presence of 1,25 $(OH)_2D$.

The effect of 1,25 $(OH)_2D$ in the intestine is limited to the phosphorus amount in food.

Normophosphataemia is primarily retained by the body through FGF23, its direct phosphaturic effect in the kidneys, and indirectly by blocking 1,25 $(OH)_2D$ production. PTH has an auxiliary phosphaturic effect in the kidneys, which is partially balanced by stimulating P release with bone resorption and 1,25 $(OH)_2D$ production stimulation (increased resorption of phosphates in the small intestine).

Thus it can be said that all calcitropic hormones fine-tune the ultimate Ca and P concentration through their mutual positive and negative interactions. Disturbed plasma (serum) P and Ca levels are subsequently a sign of depletion or disorder of these regulatory mechanisms.

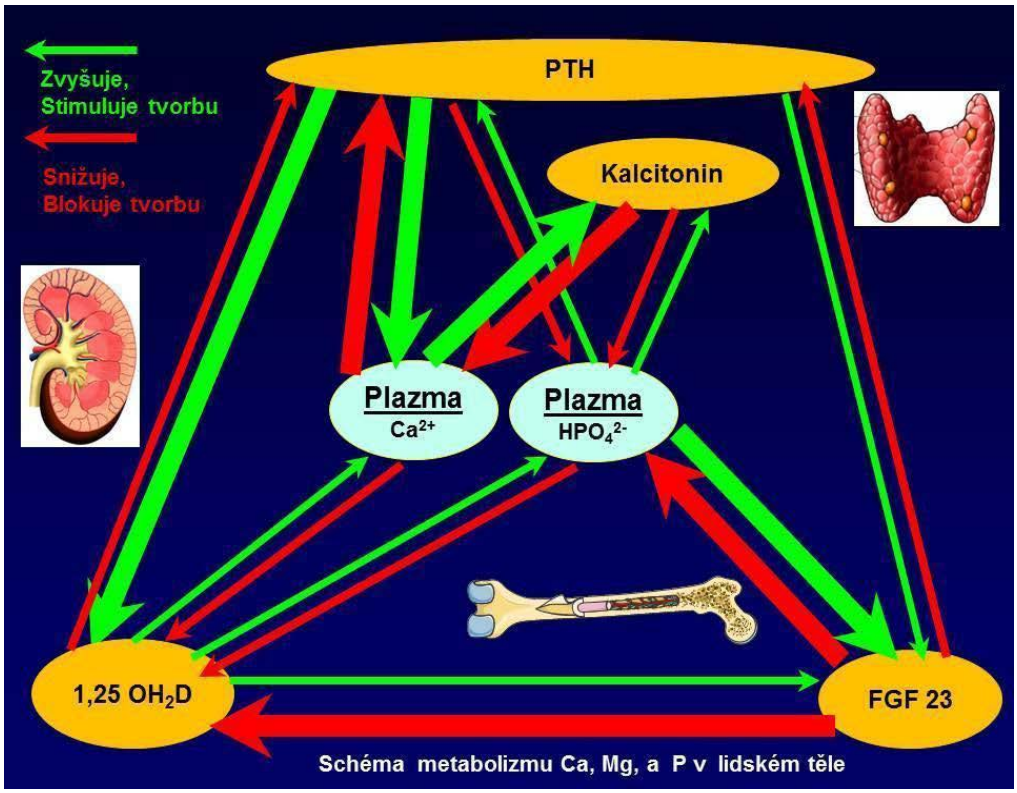


Figure 9.5. Ca, Mg and P regulation: regulating hormones and their interactions (major regulatory mechanisms in bold)

9.1.11. Summary – Key Information

- The basic goal is to maintain normocalcaemia and normophosphataemia. Physiologically, it is more difficult for the human organism to provide sufficient calcium and excrete excess phosphorus;
- Ca concentration in the extracellular fluid (ECF) is 1000 times higher than Ca concentration in the intracellular fluid (ICF);
- Ca and Na are major ECF cations, while Mg and K are major ICF cations. Bones work as a deposit site for all 3 components, which is why the Ca, Mg and P reserve is the highest here;

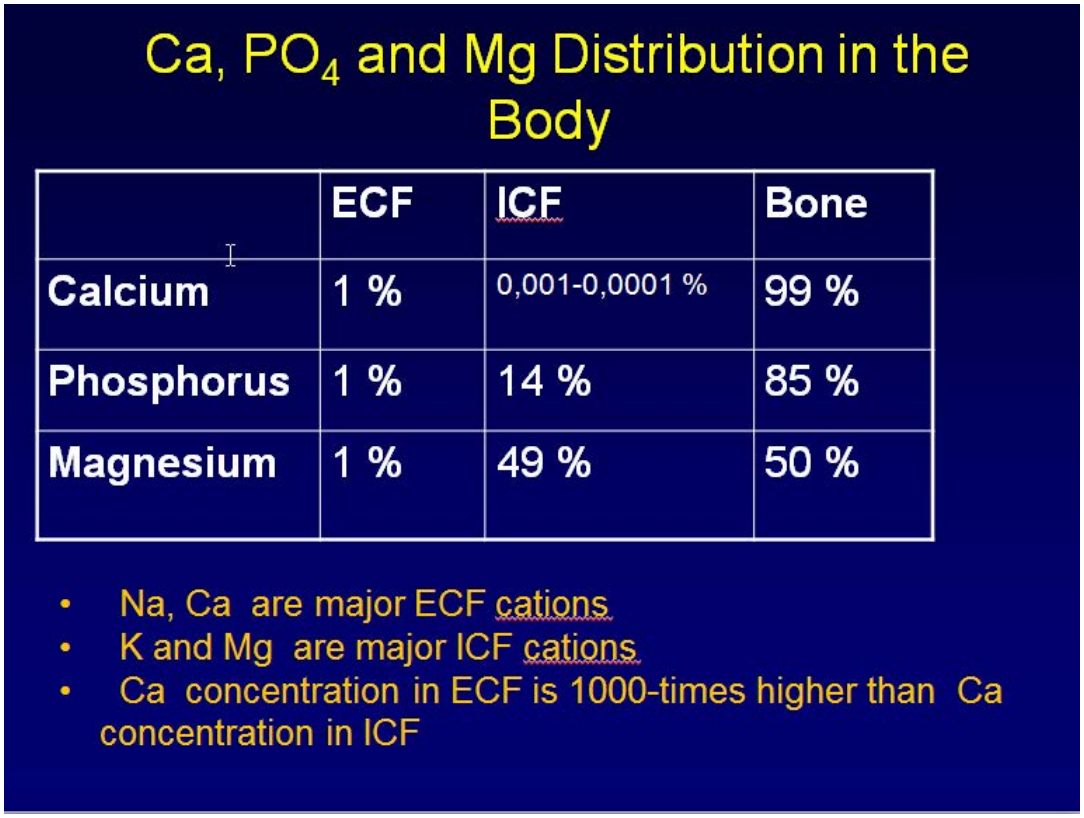


Figure 9.6. Distribution of Ca, P and Mg in the body

- The huge concentration gradient of Ca between the ECF and the ICF is the reason why Ca is used in regulating and activating functions at an intracellular level;
- The regulation and biological effects are based on ionised fractions of Ca, Mg and P, which represent about 50 % of their overall value in the plasma;

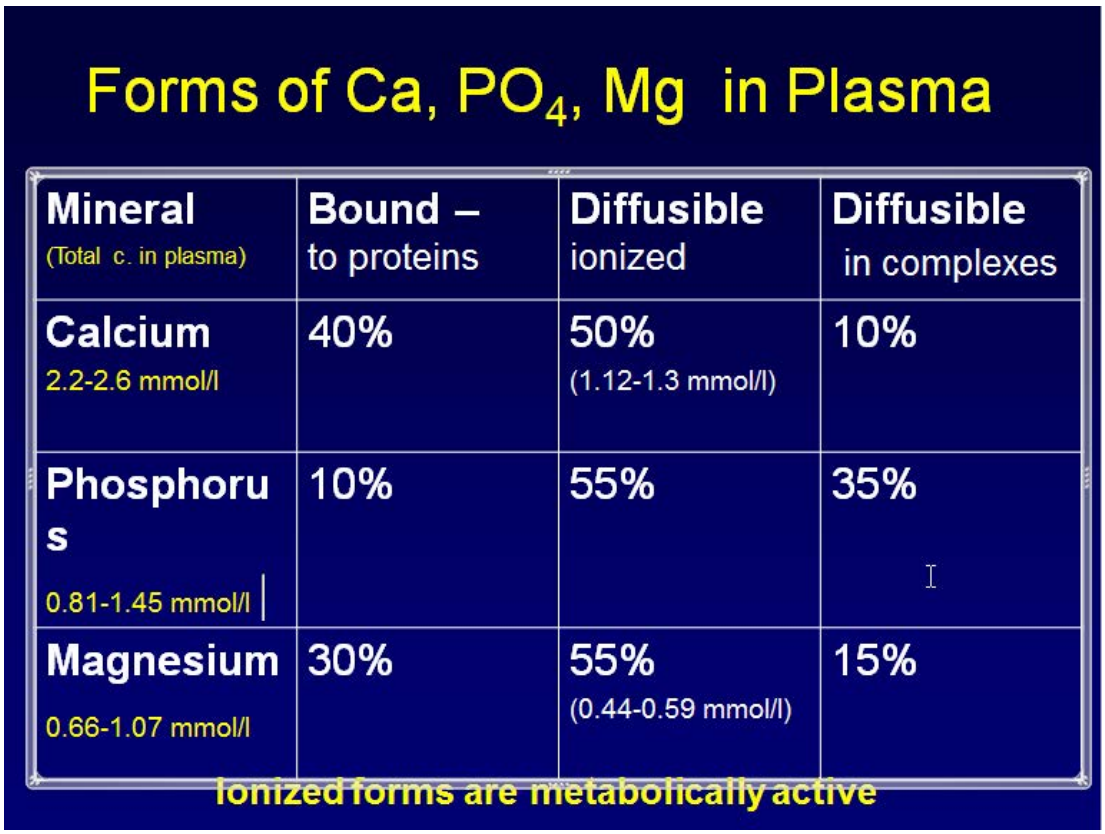


Figure 9.7. Forms of Ca, P and Mg in the plasma

- PTH, 1,25 $(OH)_2D$ and CT are involved in regulating calcaemia;
- FGF23 PTH, 1,25 $(OH)_2D$ are involved in regulating phosphataemia;

- Mg metabolism regulation is practically identical to calcium regulation, and is also likely controlled by the PTH-1,25 $(OH)_2D$ system;
- PTH is produced in the parathyroid glands, has a hypercalcemizing effect while its production is controlled by the Ca^{2+} level;
  - In bones – stimulates osteoresorption and release of Ca and P from the bones into the ECF;
  - In the kidneys – retains Ca and increases P secretion while also stimulates 1,25 $(OH)_2D$ production;
- 1,25 $(OH)_2D$ is produced in the kidneys or locally right in tissues from 25 (OH) D. It is predominantly found in the small intestine where it increases Ca and P resorption, while also regulates bone metabolism and stimulates FGF23 production in bones;
- Calcitonin is produced in parafollicular C-cells of the thyroid; plays a minor role in regulating calcaemia in the human body; has the following functions:
  - In the bones: blocks osteoclasts, thus reducing calcaemia
  - In the kidneys: increases secretion of Ca and P
  - In the CNS: stimulates endorphin production – relieves pain
- FGF23 is produced in the bones and released into the circulation; its production is stimulated by PTH and 1,25 $(OH)_2D$; in the kidneys FGF23 and Klotho protein substantially increase P secretion into urine. In addition to this, it has so far displayed unexplained systemic functions.
- The source of Ca, Mg and P is food; absorption takes place in the small intestine, which is largely passive and largely controlled by 1,25 $(OH)_2D$. Absorbed Ca, P and Mg thus enter the circulation and are deposited in bones or intracellularly (Mg, P). Deposition and resorption from bones are primarily controlled by 1,25 $(OH)_2D$ and PTH. The deposition of Ca, Mg and P is a multihormonal process involving 1,25 $(OH)_2D$, oestrogens, androgens, IGF1, calcitonin, leptin, etc. Conversely, PTH plays a crucial role in bone resorption. Excretion takes place in the kidneys, passively in the proximal tubule, actively in the distal tubule, and the resulting amount of urine is controlled by hormones (PTH, CT, FGF23).

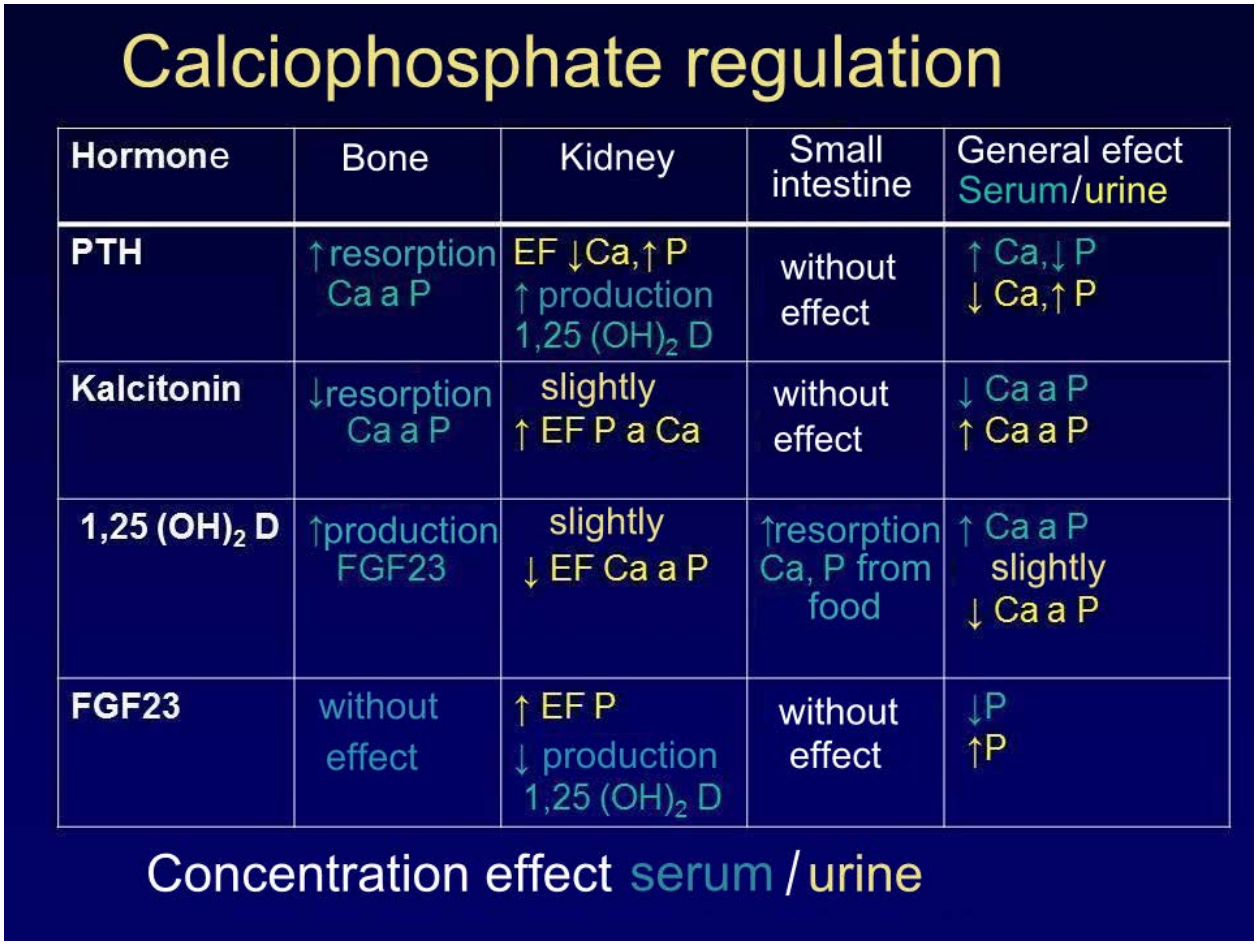


Figure 9.8. Summary of basic hormonal regulatory mechanisms of homeostasis of Ca, Mg, P (EF = excretion fraction)

- The body reaches normocalcaemia in two ways:
  - Sufficient supply from food with support from 1,25 $(OH)_2D$, which concurrently stimulates the production of FGF23, which conversely supports the excretion of phosphates taken from food, and blocks 1,25 $(OH)_2D$ production, thus finely balancing the whole system and avoiding disturbance to bone integrity.

- Release from bones, which makes up for the temporary fluctuations of ionised calcium levels by increasing PTH secretion. Chronically increased PTH secretion leads to disturbed bone integrity. Ca and P released from bones are excreted through direct PTH stimulation in the kidneys and by increasing FGF23 production in the bones. PTH stimulates 1,25 $(OH)_2D$ production and as such an effort for increasing the physiological method of retaining calcaemia.

- The body primarily retains normophosphataemia via FGF23, by direct phosphaturic effect in the kidneys and indirect blocking of 1,25 $(OH)_2D$ production. PTH is an auxiliary hormone, whose phosphaturic effect is partially balanced by the stimulation of P release with bone resorption and 1,25 $(OH)_2D$ production stimulation, which leads to increased resorption of phosphates in the small intestine.
- Disturbed plasma (serum) P and Ca levels are a sign of depletion or disorder of regulatory mechanisms or target organs.
- Calcium is an important regulator of many biological processes, which is why the ionised calcium level is very strictly controlled and kept within a fairly narrow reference range.
- The overall plasma/serum calcium concentration is affected by albumin concentration. The most common cause of pseudo-hypocalcaemia in hospitalised patients is hypoalbuminaemia (the level of ionised, biologically active Ca does not change).
- A change to the pH will considerably alter the ionised Ca level: Acidosis reduces the calcium bond to albumin and increases the ionised calcium fraction, and vice versa.
- Phosphorus is essential for many intracellular processes, but its permanently elevated plasma level is toxic for the body and accelerates the cell aging process. Among other things, it is responsible for vascular calcifications in chronic dialysed patients.

9.2. References:

1. P. Broulík, *Poruchy kalciofosfátového metabolismu*, Grada Publishing,. Praha 2003
2. A. Jabor a kol, Vnitřní Prostředí, GradaPublishing, Praha, 2008
3. I.Sotorník, Š. Kutílek a kol.; Konstní minerály a skelet při chronickém onemocnění ledvin, Galén, Praha 2011
4. C.J. Rosen;, Primer on Bone Metabolic Diseases and Disorfders of Mineral Metabolism 7th Ed, ASBMR, Washington, 2008
5. L. de Groot, J.L.Jameson; *Endocrinology 4th Edition*, Vol2W.B Sounders, 2001
6. R. J Belin, K. He :,Magnesium physiology and pathogenic mechanisms that contribute to the development of the metabolic syndrome Magnesium Research. Volume 20, Number 2, 107-29, June 2007,
7. M.S. Razzaque: Osteo-Renal Regulation of Systemic Phosphate Metabolism
8. *IUBMB Life*. 2011 April ; 63(4): 240 – 247
9. A.Martin, V.David, L. D. Quarles REGULATION AND FUNCTION OF THE FGF23/KLOTHO ENDOCRINE PATHWAYS Physiol Rev. 2012 January; 92(1): 131 – 155.

9.3. Metabolism of Calcium, Phosphorus and Magnesium

Author: MUDr. Richard Pikner, Ph.D

Reviewer: MUDr. Daniel Rajdl, Ph.D.

9.3.1. Learning Material

The preceding Preparation chapter described basic physiology and analytes and provides a basis for the clinical and biochemical part – Learning Material. Again, this chapter is divided into subchapters addressing the use of laboratory methods, indication principles, requirements for sampling (variability and stability) and basics of interpretation. The last section Key Information summarizes the basic important facts of this chapter.

9.3.2. Indications for Assay

Assays for calcium and phosphate metabolism form an integral part of medical practice. Test selection corresponds to the expected disease and pathophysiological mechanisms. It is usually good to gain comprehensive information regarding all parameter concentrations and assess them with regard to the clinical status, imaging methods and other factors.

The basic test is the plasma or serum Ca, P and Mg assay, namely total plasma Ca, P and Mg concentrations and not the metabolically active ionised fraction, which is technically more difficult. The plasma Ca, P and Mg test is a relatively insensitive tool for calcium and phosphate metabolism examination for the following reasons:

- A change in calcium concentration with a stable pH value occurs when regulatory mechanisms become depleted, i.e. PTH, 1,25 $(OH)_2$ D and calcitonin. So calcium concentration measurement is a relatively rough tool to detect calcium and phosphate metabolism disorders. If additional PTH or 1,25 $(OH)_2$ D tests are chosen, they will help detect imbalances which have not led to a fluctuation of Ca concentration so far;
- Total serum calcium concentration depends on albumin concentration. Changes in albumin concentration lead to a change in overall Ca concentration without substantially changing the ionised Ca level;
- Ionised Ca concentration depends on the current pH; the measurement is made on an ion-selective electrode with a pH measurement on POCT (Point-of-Care Testing) analysers;
- Mg is primarily an intracellular cation; its amount in the ECF is negligible compared to the ICF. Serum concentration tests only reflect the current intake rather than its tissue concentration;
- More broadly, assays for total or ionised serum Ca, P and Mg levels provide a picture of the current status in relation to regulating hormones;
- Tests of Ca, P and Mg loss in urine inform about the supply in the body assuming normal regulation in the kidneys. If kidney function is impaired, plasma and urinary concentrations together with regulating hormones should be examined;
- Urine tests are often subject to great error in the pre-analytical part (wrong urine volume measurement, instable solution – a stabilizing agent (HCl) should be added, sample should be stirred after collection and before taking a new sample, unwashed containers, etc.); there are great inter-individual differences due to diet, physical activity, etc;
- Assays for regulating hormones: PTH, 25 OH vitamin D, 1,25 OH_2 vitamin D, calcitonin and FGF23 are only performed using specialist sites; great differences among test manufacturers;
- The PTH assay and interpretation of results depend on the method selected, i.e. cross reactivity with fragments;
- The immunoassay selected for 25 OH vitamin D decides whether D2 or D3 is measured, or in what proportion. A method determining 100 % D2 and D3 should be used. The reference method is HPLC allowing for all vitamin D fractions and metabolites to be determined;
- So far the FGF23 assay is only a research method;
- The calcitonin assay has practically only one indication in the differential diagnosis of hypocalcaemia, where one of the possible causes is medullary cancer with the overproduction of calcitonin.

When a disorder of calcium and phosphate metabolism regulation is suspected, it is often necessary to prescribe a complete range of tests from plasma Ca, P and Mg concentrations through regulating hormones to urinary Ca, P and Mg losses. In primary hyperparathyroidism, for example, the laboratory findings are often a more sensitive indicator than imaging methods.

9.3.3. Sampling Variability and Sample Stability, Collection of Urine

Minimum diurnal variability except for the PTH; concentrations depend on food; collection should be performed in the fasted state.

Ca, Mg and P tests:

- Samples must not be taken into tubes with EDTA or citrate as they counterbalance all Ca and Mg in the sample;
- Haemolysis increases Mg and P values;
- PTH and PTHrP tests: If taken into EDTA tubes, samples are stable for 24 hours after centrifuging at 4 – 8 °C.;
- 25 OH vitamin D is a stable serum and plasma parameter and does not require other than usual measures.

9.3.3.1. Stability

Serum

Total Ca, P and Mg concentrations are stable for 7 days at 8 °C.

Whole blood

Not longer than 2 hours; haemolysis increases P and Mg, and does not affect the overall Ca level.

The total calcium level is affected by plasma albumin concentration; a low albumin concentration leads to hypocalcaemia, which is only relative because the ionised calcium level is usually normal. Ca concentration should be corrected with regard to albumin in all patients with hypoalbuminaemia.

Ionised forms of Ca^{2+} and Mg^{2+}: Collection the same as Astrup (heparinised venous, arterial, mixed blood). Process within 2 hours with pH measurement. Ionised Ca^{2+} and Mg^{2+} concentrations should be corrected with respect to current pH (usually done automatically by the analyser).

Urine

Ca, P and Mg values are stable at -20 °C for at least 3 weeks, at +4 °C for 3 days, though they tend to precipitate or sediment; the sample should be stabilised and thoroughly stirred before the analysis.

Urine collection: itself introduces many inaccuracies to the measurement. Collection is easy from patients with a urinary catheter. There is a high risk of collection errors in patients without a catheter.

24-hour urine collection: Thoroughly wash the container for urine collection, then provide 50 – 100- ml graduations, measure total volume in ml after collection. Collection containers should be washed with hot water only without detergents or disinfectants. Urine should be collected from 6:00 a.m. to 6:00 a.m. Start collecting urine in the morning after waking up by urinating outside the container, then save all urine in the collection container, also at night, and take the last urine in the morning after waking up 24 hours from starting the collection. Before starting the collection, add 10 ml 6 M HCl to the urine. Keep the container in a dark, cool place; stir thoroughly after collection, take a sample into a test tube and bring for the test. Adhere to usual drinking and dietary regimen during urine collection.

A sufficient volume of urine should be reached by drinking steady and suitable amounts of liquids during urine collection. A suitable intake of liquids should produce 1500 – 2000 ml of urine from an adult over 24 hours. This means that the patient should drink around ¾ of a litre of liquids (water or mineral water) for every 6 hours of urine collection (not at night).

Nordin index – 2-hour urine collection

Urine is collected from fasting patients from 8:00 a.m. to 10:00 a.m. The patient urinates before the test, then drinks at least 250 ml of liquid, and all urine is collected during the 2 hours. The test is used to measure Ca, creatinine and the Ca/Cr ratio.

Most common mistakes in urinary Ca, P and Mg assays:

- Wrong volume measurement – thorough instruction, ideally a special collection container with 50-ml graduations;
- Insufficient stirring of the collected volume before taking the sample leads to falsely lower concentrations;
- Washing the collection container and leaving phosphate detergent residues – falsely high urinary phosphorus loss;
- Ca, P and Mg precipitations – acidification is required to protect the urine from insoluble complex formation.

Most pre-analytical errors in urine collection lead to falsely lower values. Falsely higher values are received if only night and first morning urine is collected, when urinary Ca, Mg and P losses are elevated.

In common outpatient practice, urine is mostly collected at home; handling HCl (acidifying agent) is associated with too many risks of patient and material injury, so collection is usually made without the agent.

Other options for urinary Ca, P and Mg loss determination:

- 24-hour calcium loss conversion to 1 kg of body weight – very useful mainly in children and overweight patients;
- Ca/Mg ratio determination – nephrolithiasis risk assessment;
- Calculation of Ca, P and Mg fractional excretions and comparison with reference intervals.

9.3.4. Test Methods

- Photometric tests are used to determine total Ca, Mg and P concentrations. A false increase in Mg level often occurs due to haemolysis;
- The ionised Ca (Ca^{2+} and Mg^{2+}) level is measured using ion-selective methods with concurrent pH measurement. Collection must be performed immediately prior to the measurement, which is done on the acid-base analyser, and values are adjusted to the current pH. Mg^{2+} measurements are rare;
- Immunoassays based on a specific antigen-antibody reaction are used to measure PTH, PTHrP and FGF23. These are very sensitive methods where the selected antibody is essential for the test specificity, i.e. cross reactivity with related substances or degradation products;
- PTHrP and FGF23 assays are not routine tests, so the immunoassays are usually manual (ELISA or EIA);
- The PTH assay is a routine test available on automatic immunoanalysers. Currently used tests measure "intact" PTH, i.e. a mix of intact 1-84 PTH (biologically active PTH) and its 7-84 PTH fragment (35% : 65% ratio). At present there are only a few methods to determine 1-84 PTH only (Diasorin, Roche).
- Vitamin D assay:
    - Although an overwhelming majority of circulating vitamin D is the D3 form, most vitamin supplements contain D2. They have equivalent biological effects. In addition to this, 25 OH D and 1,25 $(OH)_2$ D assays should be distinguished. The 25 OH D level provides information about the status of the vitamin D supply, and is therefore a key parameter measured. Active 1,25 $(OH)_2$ D hormone measurement provides information about renal 1α-hydroxylase activity, and warns of potential ectopic production. 1,25 $(OH)_2$ D concentrations produced locally by tissues cannot be detected;
    - The basic test method is either HPLC or liquid chromatography – mass spectrophotometry. These methods provide accurate information about the concentrations of D3 and D2 isoform fractions, main metabolites 25 OH D and 1,25 $(OH)_2$ D and catabolites 24,25 OH D. The test is very accurate and specific but the instruments are expensive, it is difficult to handle while the instruments have a small capacity;
    - Immunoassays are practically only used to measure 25 OH D on automatic analys-ers; the measurement is less accurate and the antibody quality determines whether both D3 and D2 isoforms will be detected. Ideally the method detects D3 and D2. The immunoassay for determining 1,25 $(OH)_2$D requires extraction (sample preparation), and is performed only manually using enzyme or radioisotope methods.

9.3.5. Interpretation of Results

9.3.5.1. Reference Limits:

Serum, plasma

- Ca: 2.18 – 2.6 mmol/l (Klatovy donors data, n=96);

- Ca^{2+}: 1.15 – 1.35 mmol/l;
- Mg: 0.7 – 1.1 (1.3) mmol/l;
- P: 0.84 – 1.45 mmol/l;
    - 25 OH D: 75 – 150 nmol/l;
    - Intact PTH: 22 – 77 pg/ml (N-tact PTH, ILMA, Dia Sorin, Klatovy hospital donors n= 96) limits are method-dependent;
    - 1-84 PTH: 8 – 28 pg/ml (ILMA, Dia Sorin, Klatovy hospital donors n= 96).

Urine

- 24-hour Ca loss, reference limit under 6.20 mmol Ca/24 hrs for women, and under 7.5 mmol Ca/24 hrs for men of average body weight;
- 24-hour Ca loss per kg – reference limit under 0.1 mmol Ca/1 kg b.w.;
- Ca Nordin index – reference limit under 0.6 mmol Ca/mmol creatinine;
- Ca/Mg index reference interval 1.0 – 2.0;
- Fractional excretion: FE Ca under 5 %, FE P 5 – 20 %, FE Mg under 3 %;
- 24-hour Mg loss: reference interval with average body weight 3 – 5 mmol Mg/24 hrs;
- Mg/creatinine 0.2 – 0.05 mmol/mmol crea.

9.3.5.2. Calculations

- Formula for total Ca adjustment to current serum albumin level
  Ca adjusted = Ca measured (mmol/l) + 0.020*(41.3 – c. albumin in g/l);
- Formula for ionised calcium Ca^{2+} adjustment to pH 7.40
  Ca^{2+} adjusted pH 7.40 = Ca^{2+} measured (mmol/l) * (1-0.53*(7,4 – pH measured));
- Fractional excretion
  FE Ca = UCa* SCrea/ Ca^{2+} estimate */UCrea (mmol/l) … similarly for P and Mg;
- Estimated ionised calcium concentration
  Ca2+estimate = 878*SCa/(SAlbumin *15.384+1053) SCa in mmol/l, Albumin in g/l;
- Nordin index
  UCa_{NI} = UCa/UCrea (mmol/l);
- Lithogenic index = UCa/UMg (mmol/l).

9.3.5.3. Hypercalcaemia

In addition to primary hyperparathyroidism, some of the other most common causes are tumour diseases with bone metastases and PTHrP overproduction, multiple myeloma, Paget's disease, granulomatous diseases with vitamin D overproduction (sarcoidosis).

Diff. dg. of hypercalcaemia:

- Increased calcium intake: Milk-alkali syndrome;
- Increased calcium absorption in the intestine:
    - Vitamin D intoxication;
    - Sarcoidosis, tuberculosis;
    - Acromegaly;
    - Addison's disease.
- Increased bone resorption:
    - Primary hyperparathyroidism (sporadic, MEN 1 or MEN 2A, familial);
    - Ectopic secretion of PTHrP;
    - Thyrotoxicosis;
    - Haematological malignancies (osteoclast-activating factor, lymphotoxin, TNF);
    - Thiazide diuretics;

    - Lithium;
    - Immobilisation.
- Decrease in renal excretion – familial hypocalciuric hypercalcaemia;
- Abnormal serum proteins (increase in bound fraction):
    - Hyperalbuminemia, multiple myeloma, hyperglobulinemia;
    - Waldenström's macroglobulinemia.

9.3.5.4. Hypocalcaemia

Relative hypocalcaemia is most common in hospitalised patients – the decreased overall Ca concentration is determined by the low albumin concentration.

Acute conditions include hypocalcaemia in acute pancreatitis from repeated administration of transfusion preparations containing citrate, and complex formation in crush syndrome.

Acute hypocalcaemia is also responsible for clinical symptoms of hyperventilation tetany. Quickly induced respiratory alkalosis leads to a drop in ionised calcium (through an increased albumin bond) and clinical symptoms of hypocalcaemia (tetanic spasms), which even worsen the clinical presentation.

Most common chronic causes of hypocalcaemia:

Conditions associated with insufficient calcium and vitamin D intake (malabsorption, osteomalacia), chronic kidney disease, hypomagnesemia. More rarely it can be due to hypoparathyroidism (insufficient PTH production), most commonly iatrogenic, or medullary cancer (calcitonin overproduction).

Diff. dg. of hypocalcaemia:

- Hypoparathyroidism;
- Hypomagnesemia;
- Resistance to PTH;
- Osteoclast-inhibiting drugs (calcitonin, bisphosphonates);
- Vitamin D deficiency (hereditary and acquired d.);
- Resistance to vitamin D (vitamin D receptor deficiency);
- Acute administration of complex-forming agents or calcium depositions;
    - Acute (also iatrogenic) hyperphosphatemia;
    - Crush syndrome with myonecrosis;
    - Acute pancreatitis;
- Transfusion containing citrate.

9.3.5.5. Hypophosphatemia

Acute phosphatemia occurs when P is transferred to the ICF after administering glucose (anabolic phase), also in alcoholics.

Diff. dg. of chronic hypophosphatemia:

- Malabsorption;
- Severe vitamin D deficiency;
- Primary hyperparathyroidism;
- Renal disorders – losses (Fanconi syndrome) or tumours producing FGF23;
- STH deficiency;
- Hyperinsulinism;
- Phenylketonuria;
- Alcoholism;

- Tumours producing PTHrP;
- De Toni-Debre-Fanconi syndrome = distal tubular acidosis and glycosuria, renal rickets, aminoaciduria;
- Phosphate diabetes = Albright-Buttler-Bloomberg syndrome – retarded vitamin D-resistant rickets.

9.3.5.6. Hyperphosphatemia

Major causes: Renal failure, hypoparathyroidism, acute metabolic acidosis (transfer of P from the ICF), crush syndrome and acute tumour lysis syndrome

Diff. dg. of hyperphosphatemia:

- Renal failure;
- Hypoparathyroidism;
- Pseudohypoparathyroidism;
- Acute metabolic acidosis;
- Crush syndrome;
- Acute tumour lysis syndrome;
- Addison's disease;
- STH overproduction.

9.3.5.7. Hypermagnesemia: is quite rare

Diff. dg. of hypermagnesemia:

- Acute kidney injury and chronic kidney disease;
- Addison's disease;
- Acute kidney injury and chronic kidney disease;
- Hyperparathyroidism;
- Hypothyroidism;
- Addison's disease;
- Lithium intoxication;
- Surplus of antacids and laxatives containing magnesium;
- Familial hypocalciuric hypercalcaemia.

9.3.5.8. Hypomagnesemia

Causes are similar to those of hypocalcaemia; also primary hyperaldosteronism, Bartter syndrome, Gitelman syndrome; most common causes include insufficient intake from food, or malabsorption or increased renal losses concurrently with calcium disorders.

Diff. dg. of hypomagnesemia:

- Acute respiratory alkalosis;
- Acute pancreatitis;
- Intestinal malabsorption;
- Renal tubular injury;
- Drug-induced (furosemide, aminoglycosides, digoxin, cisplatin);
- Hypercalcaemia and hypophosphatemia;
- Hyperaldosteronism;
- Hyperthyroidism;
- Hypoparathyroidism;
- Alcoholism.

9.3.5.9. *Vitamin D*

Classification of vitamin D insufficiency based on serum 25 OH D levels

Severe deficiency	< 25 nmol/l
Insufficiency	25 – 49 nmol/l
Slight lack	49 – 74 nmol/l
Adequate level	**75 – 150 nmol/l**
Risk of intoxication	> 500 nmol/l

Reduced levels are usually due to a lack of exposure to sunlight (also due to the use of sunscreens), impaired intake (malabsorption) or impaired conversion to active vitamin D in kidney damage. Impaired 1,25 $(OH)_2D$ production due to FGF23 overproduction is rare.

Increased levels are found in granulomatous 1,25 $(OH)_2D$ diseases (sarcoidosis) and due to substitution preparation overdosing – increased 25 OH D level. The immunoassay is negative in the case of overdosing on active vitamin D metabolites (HPLC or mass spectrophotometry detection is required).

9.3.5.10. *PTH*

The laboratory diagnosis of hyperparathyroidism is based primarily on tests for PTH, Ca, Ca^{2+} and their urinary losses; 25 OH vitamin D is added in differential diagnosis.

Primary hyperparathyroidism most often afflicts post-menopausal women, and is also part of familial MEN1 and MEN2B syndromes, mostly involving parathyroid adenoma, or carcinoma in rare cases.

Most common causes of secondary hyperparathyroidism are: Vitamin D insufficiency, malabsorption syndromes, and bone disease in kidney damage.

Tertiary hyperparathyroidism is caused as autonomous parathyroid hyperplasia with previous secondary hyperparathyroidism in kidney damage.

9.3.6. Summary – Key Information:

9.3.6.1. *Indications*

- The basic test is the plasma or serum Ca, P and Mg assay, namely overall plasma Ca, P and Mg concentrations and not the metabolically active ionised fraction, which is technically more difficult;
- Assays for overall or ionised serum levels provide a picture of the current status in relation to regulating hormones;
- Plasma Ca, P and Mg assays are not very sensitive tools for examining calcium and phosphate metabolism since changes in the levels occur when regulating mechanisms have been depleted;
- The overall plasma calcium concentration depends on albumin concentration. Changes in albumin concentration lead to a change in total Ca concentration without substantially changing the ionised Ca level (patients have low Ca levels but do not display symptoms of hypocalcaemia);
- Ionised Ca concentration depends on the current pH; the measurement is made on an ion-selective electrode with a pH measurement on POCT (Point-of-Care Testing) analysers;
- Mg is primarily an intracellular cation; serum concentration tests only reflect the current intake rather than its tissue concentration;
- Tests of Ca, P and Mg loss in urine point to information about the supply in the body assuming normal regulation in the kidneys, and are also subject to a high error rate during sample collection;
- Assays for regulating hormones: PTH, 25 OH vitamin D, 1,25 OH_2 vitamin D, calcitonin and FGF23 are only performed by specialist sites; great differences between test manufacturers lead to possible different interpretations.

9.3.6.2. *Stability*

- Samples for Ca, Mg and P assays must not be taken into tubes with EDTA or citrate as they counterbalance

all Ca and Mg in the sample; overall Ca, P and Mg concentrations are stable at 8 °C for 7 days; haemolysis increases Mg and P levels;

- PTH and PTHrP tests: If taken into EDTA tubes, samples are stable for 24 hours after centrifuging at 4 – 8 °C;
- 25 OH vitamin D is a stable serum and plasma parameter and does not require other than usual measures;
- Ionised forms of Ca^{2+} and Mg^{2+}: Collection identical to Astrup (heparinised venous, arterial, mixed blood). Process within 2 hours with pH measurement;
- Urinary Ca, P and Mg: Ca, P and Mg values are stable at -20 °C for at least 3 weeks, at +4 °C for 3 days, though tend to precipitate or sediment; the sample should be stabilised (10 ml 6Mol HCl for 24-hour collection) and thoroughly stirred before the analysis.

Test Methods:

- Immunoassays are used for determining PTH, PTHrP and FGF23; different tests use different reference limits (differences in test specificity and sensitivity);
- Currently widely available PTH tests detect 1-84 PTH as well as 7-84 PTH fragment. Methods measuring only biologically active 1-84 PTH are also possible, however these have substantially lower reference limits;
- Liquid chromatography-mass spectrophotometry is ideal for measuring vitamin D as it measures all metabolites, but this is not a routine test. Immunoassays are used for routine testing with varying specificity to D2 and D3 isoforms (equal measurement of 25-OH-vitamin D2 and D3 isoforms is ideal).

Interpretation:

- Formula for overall Ca adjustment to current serum albumin level
  Ca adjusted = Ca measured (mmol/l) + 0.020*(41.3 – c. albumin in g/l);
- Major causes of hypercalcaemia: Primary hyperparathyroidism, tumour diseases with bone metastases and PTHrP overproduction, multiple myeloma, Paget's disease, granulomatous diseases with vitamin D overproduction (sarcoidosis);
- Acute hypocalcaemia: quickly induced respiratory alkalosis leads to a decrease in ionised Ca – hyperventilation tetany;
- Chronic hypocalcaemia: conditions associated with insufficient calcium and vitamin D intake (malabsorption, osteomalacia), chronic kidney disease, hypoparathyroidism, medullary carcinoma;
- Hypophosphatemia is most common in the following conditions: Malabsorption, severe vitamin D deficiency, primary hyperparathyroidism, renal disorders, tumour diseases;
- Hyperphosphatemia is most common in the following conditions: Renal failure, hypoparathyroidism, acute metabolic acidosis (transfer of P from the ICF), crush syndrome and acute tumour lysis syndrome;
- 25 OH vitamin D is an indicator of vitamin D supply status: adequate level 75 – 150 nmol/L (severe deficiency < 25 nmol/L, often associated with secondary elevation of PTH);

For differential diagnosis of primary and secondary hyperparathyroidism see Fig. 9.1

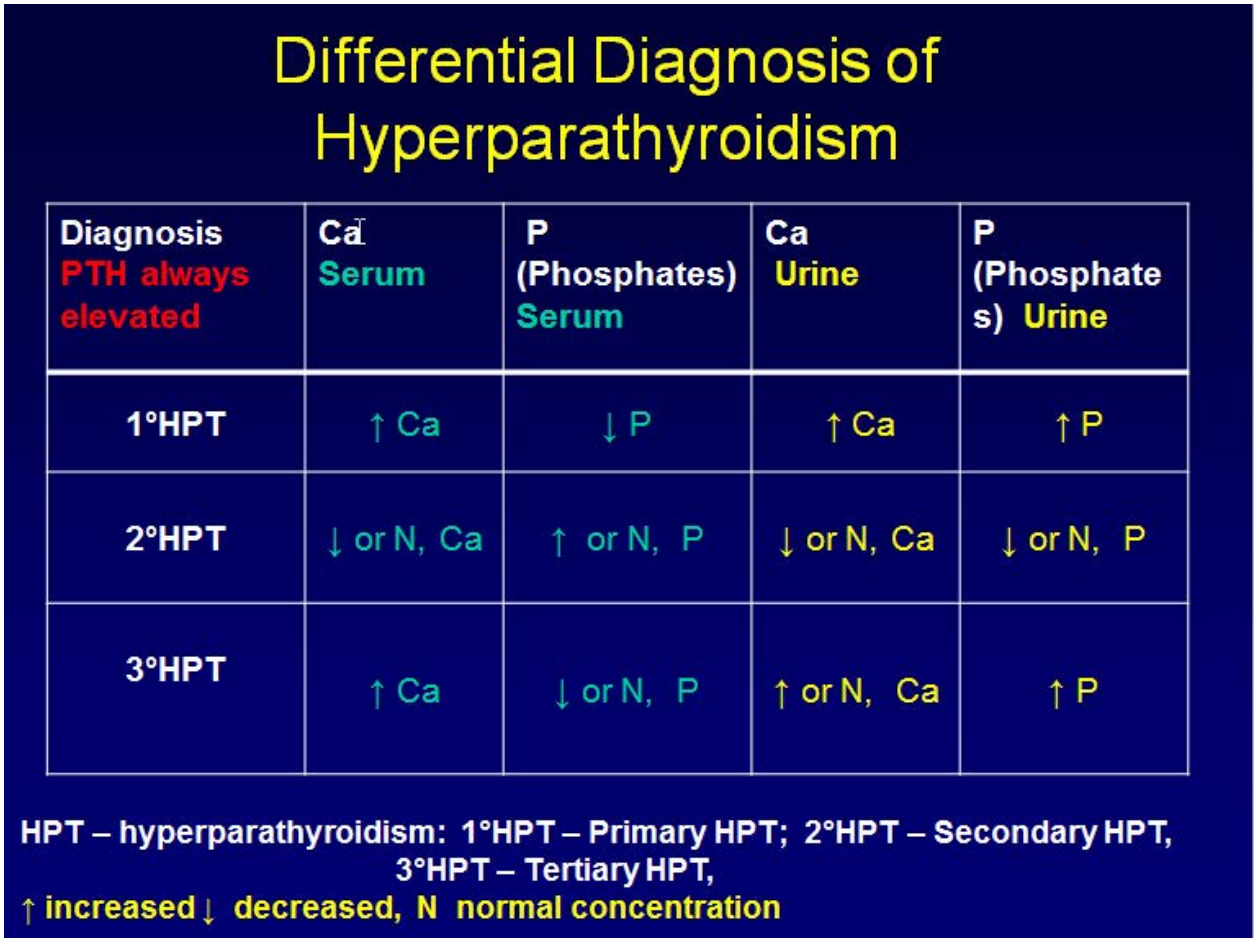


Figure 9.9. Differential diagnosis of hyperparathyroidism (HPT)

9.4. References:

1. P. Broulík, *Poruchy kalciofosfátového metabolismu* , Grada Publishing,. Praha 2003
2. A. Jabor a kol, *Vnitřní Prostředí*, GradaPublishing, Praha, 2008
3. I.Sotorník, Š. Kutílek a kol.; *Konstní minerály a skelet při chronickém onemocnění ledvin*, Galén, Praha 2011
4. C.J. Rosen; *Primer on Bone Metabolic Diseases and Disorders of Mineral Metabolism 7th Ed, ASBMR*, Washington, 2008
5. L. de Groot, J.L.Jameson; *Endocrinology 4th Edition, Vol2W.B Sounders*, 2001
6. L. Thomas: *Clinical Laboratory Diagnostics*, TH-Books, 1998

CHAPTER 10

10. Trace Elements

Authors: doc. MUDr. Milan Dastych, CSc., MBA; MUDr. Milan Dastych

Reviewers: prof. MUDr. Antonín Kazda, Ph.D.; prof. MUDr. Tomáš Zima, DrSc., MBA

10.1. Esential Trace Elements

The human body as part of the terrestrial ecosystem draws its elementary composition from the elements of the Earth's crust. If a sufficiently sensitive analytical method is used, practically all the elements of Mendeleev's periodic system can be demonstrated in the human body. However, it is very difficult, and perhaps impossible, to interpret such finding at present.

Table 10.1 shows the percentage (elemental) composition of the Earth's crust and the human body. Given the prevalence of organic matter, the building blocks of organic compounds (O, C, H, N) dominate in the elemental composition of the human body.

Element Percentage (%)							
Earth's Crust				Human Body			
O	47	Mg	2.1	O	62.8	Na	0.26
Si	28	Ti	0.4	C	19.4	K	0.23
Al	8	H	0.1	N	9.3	Cl	0.18
Fe	5			Al	5.1	Mg	0.04
Ca	4			Ca	1.4	**Fe**	**0.009**
Na	2.8			P	0.63	**Zn**	**0.00003**
K	2.6			S	0.60	**Cu**	**0.000001**

Table 10.1. Percentage (elemental) composition of the Earth's crust and the human body

Trace elements are found in the body in very low, "trace" amounts not exceeding 0.01 % of body weight, with typical concentrations of µmol/l (µg/l) and lower.

They can be divided into essential and non-essential trace elements. Trace elements essential for humans are: Fe, Zn, Cu, Se, Cr, Co, Mn, Mo, I; possibly essential trace elements: Al, Si, F, V, Ni; trace elements with hitherto unknown physiological function, considered toxic heavy metals: Pb, Cd, Hg, As, Bi, Sn, Ag, Au, Pt, Ti, Rb, Sr, Ba, Y (Table 2).

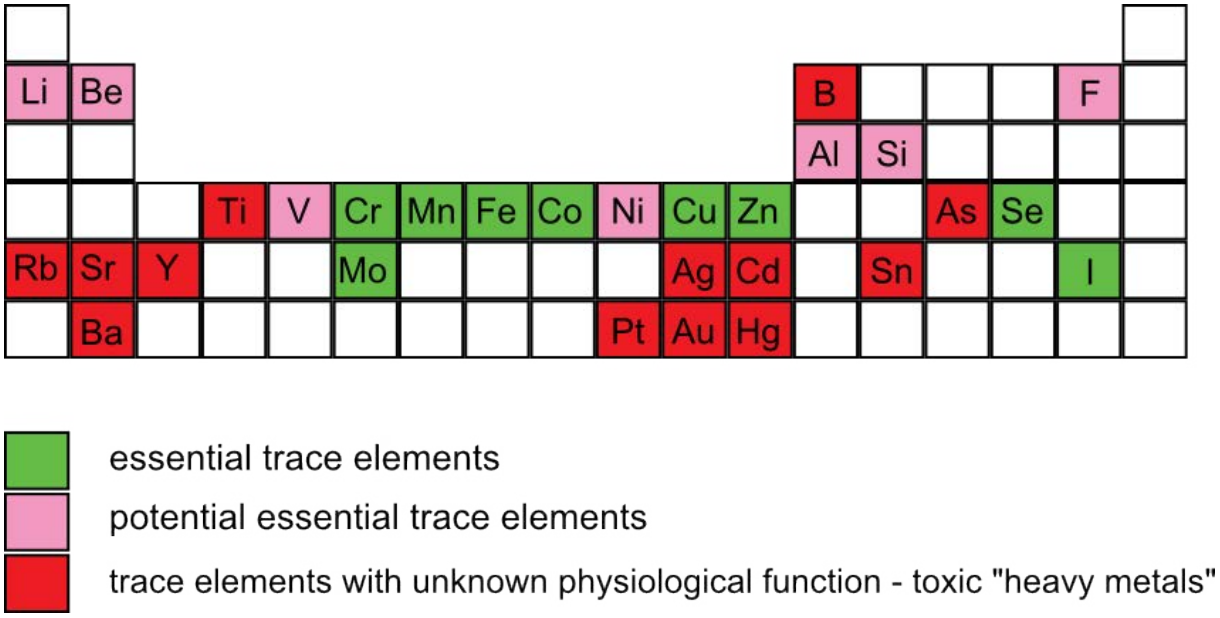


Table 10.2. Arrangement of the elements listed above in Mendeleev's periodic system

10.1.1. Definition of essentiality

Essentiality is generally defined by the occurrence of a specific defect in the structure or function of a tissue or organ resulting from a deficiency of the relevant element, and the remedy or prevention of such defect when such element is replenished. The definition of essentiality is currently supplemented by the requirement for knowledge and description of the element's action at the molecular level, including the specific physiological effect in the body.

Most essential trace elements have an important function as co-factors of enzymatic catalysis processes such as metalloenzymes and non-enzymatic metalloproteins.

Trace element deficiency may be caused by insufficient intake in food or by their reduced utilization from food, decreased absorption (maldigestion, malabsorption), increased losses (diarrhoea, bleeding in the GIT, exudative enteropathies) or by increased requirement (stress, fever, sepsis, tumours, traumas, burns). The early years of total parenteral nutrition without the supplementation of essential trace elements can be used as a model case of induced essential trace element deficiency. Numerous interactions between trace elements and other nutrients are known which may also cause their deficiency in the organism. There are also genetic defects in the population that have a considerable effect on trace element metabolism (Zn-Acrodermatitis enteropatica, Cu-Menkes disease, Wilson's disease, congenital thyroid disorders, haemochromatosis).

In addition to their essentiality, some trace elements show high toxicity at higher concentrations (Se); others are toxic at high doses in connection with poisoning or chronic, often occupational exposure.

10.1.2. The diagnosis of essential trace element deficiency

The diagnosis of essential trace element deficiency is currently based on the following three principles:

- Laboratory assays of the trace element content in different body fluids (plasma, serum, urine, saliva) and cells or tissues (erythrocytes, leukocytes, thrombocytes, hair, nails, bioptic material);
- Determination of the activity of trace element-dependent enzymes;
- Clinical signs of the deficiency.

Clinical signs are usually accompanied by severe whole-body deficiency conditions. The absence of a sufficiently sensitive and specific test to reveal a slight or medium deficiency of trace elements presents a continuing problem.

It is advisable to combine all three procedures alongside their comprehensive assessment and evaluation, respecting all available knowledge about specific and non-specific signs of deficiency, including other influences on the monitored parameters. In such situation, a therapeutic test continues to have a high predicative ability. Where causal connection exists, administration of the suspected deficient micronutrient will lead to an improvement in the clinical and laboratory signs.

10.1.3. Methods for trace element determination in biological material

Flame or electrothermal atomic absorption spectrometry (AAS) is the commonest used method for determining trace element concentration (except for iodine). It is absolutely essential to respect all precautions for trace assays in the pre-analytical and analytical phase, to use special collection sets, acid-washed laboratory glassware, "suprapure" reagents, the highest quality of water and gases (for AAS), and to maintain a clean, dust-free laboratory environment, especially for ultra-trace assays.

10.1.4. General rules for adequate and safe intake of essential trace elements from food

As the essentiality of each trace element for the human body was discovered, the essential trace elements were included in recommendations for their dietary intake from food. Issues concerning the definition and determination of recommended values for the optimum daily dietary intake of essential micronutrients (trace elements and vitamins) continue to be discussed today. Such issues have a long history. The first recommendations, **Recommended Dietary Allowances (RDA)**, were published in the US in 1941 in connection with World War Two and the introduction of the food rationing system. They included recommended nutrient doses, including micronutrients (particularly vitamins), intended to ensure their sufficient intake for the general population. Emphasis was put on preventing the occurrence of nutritional rickets, scurvy and pellagra. RDAs for essential trace elements were only specified in later revisions of this document. Today, RDAs are being replaced with a newly defined term, Daily Recommended Intake (DRI), intended to incorporate four other terms: Estimated Average Requirement (EAR), Recommended Daily Allowance (RDA), Adequate Intake (AI) and tolerable Upper Limit. The Population Reference Intake (PRI) is defined in Europe. Definitions of the above terms differ in details and in the range of their application to differently stratified populations of healthy and diseased people.

In general, they meet the following criteria:

- Minimum micronutrient amount preventing a specific disorder from occurring;
- Amount required to maintain an optimum nutritional status;
- Daily intake with the lowest risk of potential toxic effect.

In respect of the health of the local population, it can be stated that the normal mixed diet is a sufficient source of nutrients, including trace elements, except perhaps for iodine and selenium.

10.1.5. Food chain and food production

The content of essential trace elements in food largely relates to the food chain (soil, plants, livestock and humans). The content of trace elements in the soil, which is subject to considerable, in global terms often vast, regional differences, may directly influence grazing cattle in particular. Other farm animals, for example pigs and poultry, fully rely on the supply of trace elements in prepared feed and feed mixes. Serious deficiencies afflicting entire breeds may occur and do actually exist even here. The influence of local food production is decreasing to some extent due to ongoing globalization within the worldwide trade, export and import of staple foods.

10.1.6. Utilization of trace elements from food – interactions

Trace elements are an integral part of the human diet. In particular, they come from animal products, whose organically bound elements are better utilizable for the human body than their inorganic forms. The mutual interactions of micro and macronutrients and other food components should be understood and respected when considering the adequate intake of essential trace elements. The commonest interfering substances are phosphates, fibre, phytanic acid (inositol-hexaphosphate) and other chelating agents. In general, one element supplied in excess may negatively affect the absorption and metabolism of others.

The text below deals with Iodine, Iron, Zinc, Copper and Selenium in detail. The other essential trace elements, Chromium, Cobalt, Manganese and Molybdenum will be mentioned only marginally, since they have yet to find any clear practical application in clinical medicine.

10.2. Iodine

10.2.1. Biochemistry and metabolism

Iodine is absorbed in the intestine in the form of iodide. It is very efficiently taken up from the blood plasma by the thyroid parenchyma using the iodide pump with the participation of Na/K dependent ATPase under TSH control. Through the action of thyreoperoxidase, the tyrosine molecule is gradually iodized to form mono, di, tri and tetraiodotyrosine (thyroxine), which is saved (concentrated) in the thyroidal follicle colloid. The main active form of the thyroid hormone triiodothyronine is produced by cleaving one iodine atom away from the tetraiodotyrosine (thyroxine) molecule through the action of selenium-dependent 5´-deiodase.

Iodine is excreted from the body in the urine. Iodine output correlates with iodine intake from food, and so iodineuria can be used to estimate the intake.

10.2.2. Biological functions

Iodine is numbered among the essential trace elements for its single but essential and unique function in the body. It is indispensable for the synthesis and function of thyroid hormones, triiodothyronine T3 (3 I atoms) and tetraiodothyronine T4 (4 I atoms).

Inadequate iodine supply from food leads to a reduction of T4 and T3 synthesis, as well as the T4 and T3 blood level.

On the negative feedback principle, the hypophysis reacts to this situation by increased secretion of TSH, which leads to thyroid hypertrophy and, at a later stage, goitre.

Due to the suboptimal function of thyroid hormones, iodine deficiency causes disorders of CNS development and function, growth and sexual maturation disorders, disorders of energy metabolism regulation, temperature regulation and other functions.

In adulthood, iodine deficiency is manifested by complex symptoms of myxoedema, usually with goitre.

A guaranteed source of iodine in food is sea fish; iodine in other vegetable and animal products depends in the iodine content in the ambient environment. An important source can be the targeted or even accidental enrichment of foodstuffs with iodine (e.g. contamination of milk by traces of iodine agents used for disinfecting dairy vessel and tubing).

The recommended daily dose of iodine for men is 150 µg, 120 µg for women and 170 µg in pregnancy.

10.2.3. Deficiency – causes and signs

Iodine Deficiency Disorders (IDD) are globally regarded as the most widespread cause of CNS damage with signs of mental retardation which could be removed through targeted and consistent prevention.

Based on the WHO records, about 1/3 of the world's population (about 2.2 billion people) is at risk from iodine deficiency signs. Iodine is naturally found in soil and seawater. There are regions on Earth with very low iodine content in the ambient environment, with a high incidence of iodine deficiency signs and marked clinical signs such as goitre and mental disorders. The severest intellectual disorders (e.g. imbecility) specific to these regions have given rise to the name ***endemic cretinism***.

The name ***sporadic cretinism*** refers to the same clinical picture caused by a congenital thyroid function defect that was not identified in time and treated immediately after the birth of the afflicted child.

Observations of the widespread global endemic occurrence of iodine deficiency signs have led to efforts to eliminate them through widespread iodine supplementation. This was started in the US in 1920 and has spread over time to most regions at risk of iodine-deficiency in more than 140 countries of the world. From the many possible ways of iodine supplementation, table salt iodization has proved the optimal method. Salt is accessible to all social and economic groups, shows an even intake without seasonal swings, there is no risk of overdosing, and iodization by adding KI is easy and inexpensive.

10.2.4. Diagnosis of the deficiency

Iodine deficiency is accompanied by a drop in the thyroid hormone (fT4, fT3) level and a feedback increase in the TSH level.

Insufficient iodine intake from food leads to decreased iodine excretion in urine (iodineuria); < 20 μg per litre is indicative of a serious deficiency compared to normal levels of 100 – 200 μg per litre.

10.2.5. Toxicity

Excessive iodine intake may go together with an excessive intake of sea products, mostly, however, without marked somatic signs. An iatrogenic excessive intake (e.g. KI) may cause signs of thyroid hyperfunction, including hyperfunctioning goitre.

10.3. Iron

Iron is the most abundant essential trace element in the human body. The global incidence of iron deficiency resembles that of iodine.

10.3.1. Biochemistry and metabolism

Most iron in the body is bound in haemoglobin (60 %), the rest in myoglobin, ferroproteins and iron-dependent enzymes. The metabolic pathway of iron from food to target molecules in cells, tissues and organs requires the synergism of different systems. The prevalent form of iron in food (Fe^{3+}) must be reduced to Fe^{2+} to be absorbed in the small intestine. This reduction is enhanced by HCl in the gastric juice, vitamin C, citric acid and lactic acid. Following absorption, iron is re-oxidized by action of caeruloplasmin (also called ferroxidase) and embedded in the transferrin molecule. Transferrin brings iron to cells, into which it is incorporated through surface receptors. Iron is incorporated in haemoglobin in bone marrow, myoglobin in muscles, and iron-dependent enzymes and reserve metalloproteins (ferritin, haemosiderin) in other cells.

Daily iron loss (0.35 mg) is due to occult blood losses in the digestive system, bile, skin and urine (80 μg). These losses may rise significantly during GI tract diseases, particularly bleeding and menstrual cycle disorders in women.

The daily requirement depends on age and gender, the quality and composition of food and the amount of losses; 10 mg for men and 15 mg for women (25 mg during pregnancy).

10.3.2. Biological functions

The biological function of iron is closely connected with oxygen energy metabolism. Hexavalent iron in haemoglobin (four valences go to nitrogen ions in the four pyrrole nuclei, the fifth link binds globin-coupled histidine, and the sixth valence is free for oxygen) provides a unique transport of oxygen to tissues and cells, where cytochrome redox chains (Fe^{2+} … Fe^{3+}) release energy for producing the high-energy phosphate bonds of ATPs.

Primarily, iron is essential for haemoglobin synthesis in erythrocytes. Iron deficiency is responsible for microcytic hypochromic anaemia.

10.3.3. Deficiency – causes and signs

Iron deficiency is the commonest nutritional deficiency and the commonest cause of anaemia. Sideropenic anaemia afflicts 500 – 600 million people. Sideropenic anaemia is the commonest blood disease occurring during childhood.

Iron deficiency in the body is caused by an unequal relationship between iron intake and increased losses (sometimes even just obligatory losses).

Iron deficiency is more common among women due to blood losses during menses.

Another cause may be inadequate dietary composition with a low iron content, the presence of substances reducing iron utilization (oxalic acid and polyphenols in coffee, tea and chocolate which form insoluble complexes with iron), or an excessive supply of trace elements (Zn, Cu, Mn) causing a competitive decrease in iron absorption in the

intestine.

10.3.4. Diagnosis of the deficiency

Marked iron deficiency in the body leads to hypochromic microcytic anaemia. Iron assays and serum transferrin, ferritin and soluble transferrin receptor saturation assays are used to confirm the aetiological role of iron deficiency, see Table 9.3.

S-iron	9.0 – 30.0 µmol/l
Serum transferrin saturation	< 16 %
S-ferritin	< 15 µg/l
S-soluble transferrin receptors	> 8.5 mg/l

Table 10.3. Laboratory parameters for the diagnosis of iron-deficiency anaemia.

10.3.5. Toxicity

Typical symptoms of the toxic action of iron surplus in the body accompany haemochromatosis with congenital abnormally high iron absorption in the intestine, which leads to liver damage, diabetes and skin pigmentations (bronze diabetes). Secondary disorders may be due to excessive iron intake or abundant transfusion. Iron toxicity relates to the pro-oxidation action of iron. Excessive Fe^{3+} not bound to binding proteins may induce an increased production of free radicals and tissue damage by oxidation.

Molecular Level	**(Patho)physiological Functions**	**Clinical Signs**
Fe-metalloenzymes		
Cytochromes C	Oxygen energy metabolism	Hypoxia
Cytochromes P450	Transformation of xenobiotics	Inefficiency/toxicity of drugs
Fe-proteins		
Haemoglobins	Oxygen transport in blood	Hypoxia
Myoglobin	Oxygen metabolism in muscles, intracellular O_2 transport	
Transferrin	Fe-carrying protein, acute-phase protein	Anaemic conditions
Ferritin	Fe-storage proteins	Anaemic conditions
Haemosiderin	Proteins saving iron surplus	

Table 10.4. Molecular basis of clinical signs of iron deficiency (examples)

10.4. Zinc

Professor Prasad was the first to describe completely documented clinical and laboratory observations of severe human zinc deficiency in 1961 in Iran.

Nevertheless, it was only the discovery by Barnes and Moynahan (1970) of the role of zinc in the aetiology and clinical signs of acrodermatitis enteropathica, and its causal treatment with a complete remission of signs following zinc supplementation, which led to the inclusion of zinc among the essential trace elements and the assessment of the RDA for zinc (1974).

10.4.1. Biochemistry and metabolism of zinc

Total plasma zinc (13 – 18 µmol/l) includes zinc freely bound to albumin, more strongly bound to other plasma proteins, particularly alpha-2-macroglobulin, diffusible zinc bound to plasma amino acids, particularly histidine, cysteine and lysine, and a low amount of ionized zinc. Plasma zinc represents only about 0.1 % of the total zinc content in the body.

The recommended daily dose of zinc in a normal mixed diet containing animal proteins ranges between 8 and 15 mg/day.

The output side in the zinc balance consists mainly of zinc losses in the digestive tract, stool 15 – 30 µmol/day (1 - 2 mg/day), urine 7.5 µmol/day (0.5 mg/day) and sweat 7.5 µmol/day (0.5 mg/day).

10.4.2. Biological functions of zinc

More than 200 enzymes are known to require the presence of zinc in their molecule to be functional. They include almost all groups of enzymes through which zinc widely participates in the metabolism of carbohydrates, lipids, proteins and nucleic acids, including many other metabolic processes.

Zinc is required for the synthesis and metabolism of nucleic acids (DNA, RNA), through which it is closely related to the cell division rate and therefore growth, differentiation, morphogenesis, regeneration, the healing of wounds, etc.

As part of Zn,Cu-superoxide dismutase, zinc is an essential part of the anti-oxidative protection of cells against superoxide radical ($O_2^{\cdot-}$) damage.

Zinc is important in maintaining the integrity and proper barrier function of the skin. Severe deficiencies are accompanied by various skin alterations, parakeratosis, vesiculo-pustular exanthema with crusts, and frequent infections with a major occurrence of circumoral and genitoanal alterations and alterations on peripheral extremities. The most pronounced signs accompany acrodermatitis enteropathica with a zinc absorption defect in the intestine.

The participation of zinc in the effective function of the immune system is extremely complex. The cellular part of immunity is afflicted most; the T lymphocyte count and differentiation decreases, as well as the T_4^+/T_8^+ lymphocyte sub-population ratio.

This condition is associated with thymus atrophy accompanied by a marked decrease in thymulin concentration in the plasma. The clinical sign of the effect of zinc deficiency on immune functions may be a reduced immunity to infections, relapsing infections and chronic inflammatory affections of the skin and mucosa.

10.4.3. Deficiency of zinc – causes and signs

Depending on its severity, zinc deficiency may induce serious, life-threatening conditions, including developmental defects, retardation of growth and sexual maturation, or just discrete skin alterations, slightly lower stature, taste disorder, night blindness or a greater liability to different infections.

In general, zinc deficiency in the body may be provoked by lower zinc intake, increased losses, increased requirement or a combination of these effects. Animal products represent a more valuable source of zinc due to the higher zinc utilization during absorption in the digestive tract. Legumes and nuts are the best source of zinc for vegetarians.

Zinc absorption disorder goes together with chronic intestinal diseases such as Crohn's disease, coeliac sprue, ulcerous collitis, chronic diarrhoeas, jejunal bypass, external pancreatic secretion insufficiency, etc.

Increased zinc losses, especially in the urine, are associated with increased catabolism such as burns, polytraumas, sepsis, neuroinfections, etc.

An increased requirement of zinc is typical of fast growth periods, pregnancy, lactation and the anabolic phase following heavy strain with catabolism.

10.4.4. Diagnosis of the deficiency

Serum zinc is used most often for assays due to its sampling accessibility. Albumin (hypalbuminaemia reduces zincaemia) and CRP levels must be known at the same time. Infections and inflammations cause zinc redistribution from plasma to cells through cytokines, so a reduced serum zinc concentration is not indicative of any whole-body zinc deficiency.

Preventing the contamination of collection and transport containers poses a substantial pre-analytical problem. Even brief contact of blood or urine with a rubber stopper may cause gross errors. Rubber filler contains ZnO and zinc elution from rubber is very fast.

Attempts at assays for zinc in leukocytes, thrombocytes, hair and muscles have not led to any wider practical use.

The determination of zinc-dependent cellular enzyme activities seemed to be promising. These enzymes included erythrocyte carboanhydrase, leucocyte alkaline phosphatase, erythrocyte Zn,Cu-superoxide dismutase and others. However, findings have not been so far clearly and consistently interpreted.

Given the difficulty of obtaining clear laboratory proof of boundary zinc deficiency in particular, a therapeutic test is indicated in cases of clinical suspicion of whole-body deficiency. Signs induced by zinc deficiency will promptly disappear following zinc supplementation.

The problem of zinc substitution in the acute conditions of critically ill patients with hypozincaemia (as in the case of iron) remains unresolved.

10.4.5. Toxicity of zinc

Zinc is a relatively low-toxic element. Acute and chronic zinc intoxications are described as occupational diseases in people inhaling metal zinc fumes or fine dust (metal fume fever or zinc shakes) with manifestations such as fatigue, headache, cough, chills and fever.

Molecular Level	**(Patho)physiological Functions**	**Clinical Signs**
Zn-metalloenzymes		
Thymidine kinase DNA polymerase	Nucleic acid metabolism Cell division, growth, regeneration	Delayed growth, dwarfism, poor healing of wounds
Zn,Cu-superoxide dismutase	Anti-oxidative protection	Oxidative stress signs
Retinol dehydrogenase	Visual purple regeneration	Night blindness
Collagenase	Collagen fibre metabolism	Connective tissue disorders
Zn-proteins		
Metallothionein	Essential and other trace element metabolism Cell membrane stabilization	Signs of deficiency, toxicity of elements
Gene regulation, Zn-finger-proteins and expression of different protein genes	Reduced expression of different protein genes, cell growth, differentiation, morphogenesis	Developmental defects of neural tube, spina bifida, meningocele, anencephaly
	Spermatogenesis	Lower sperm count, lower testosterone level, hypogonadism
	Barrier functions of skin	Parakeratosis, genito-oral and acral vesiculo-papilous exanthema
Thymulin	Differentiation of T lymphocytes, chemotaxis, antigenic T-cell response Bb antigens	Reduced immunity to infections, relapsing infections, chronic inflammatory affections of skin and mucosa
Glucocorticoid and oestrogen receptors	Specific cortisol and oestrogen bond	Hormonal disorders

Table 10.5. Molecular basis of clinical signs of zinc deficiency (examples)

10.5. Copper

As an essential trace element, copper is third in abundance in the human body after iron and zinc.

In routine clinical practice, copper metabolism disorder is usually encountered in the differential diagnosis and monitoring of Wilson's disease (hepatolenticular degeneration).

10.5.1. Biochemistry and metabolism of copper

Copper is widespread in biological tissues, where it is found in the form of organic complexes, mostly metalloproteins and metalloenzymes.

Copper in the blood is generally divided between erythrocytes and plasma. Most copper in red blood cells is contained in Cu,Zn-superoxide dismutase. Physiological plasma copper concentrations range from 11 to 22 µmol/l. Women have about 10 % higher plasma copper concentrations with up to a threefold increase in the late stage of pregnancy, and levels are also increased in women using oestrogen-based contraceptives. Copper along with caeruloplasmin levels increase during non-specific reaction to stress exposure (traumas, operations, malignant tumours, biliary obstructions, etc.).

Copper metabolism is very closely connected with the synthesis and functions of caeruloplasmin.

Caeruloplasmin is a plasma protein synthesized in the liver, containing 6 – 7 copper atoms and showing marked enzyme oxidase activity. Copper bound in caeruloplasmin represents 80 – 93 % of plasma copper concentration. The remaining 7 % is freely united to albumin and some plasma amino acids.

The major copper excretory pathway is copper excretion in the bile in the form of poorly absorbable complexes leaving the body with the stool. Only very small amounts are excreted in the urine in normal circumstances (32 – 64 µg/24 hrs, 0.5 – 1.0 µmol/24 hrs).

Given their high reactivity, copper atoms bind to different carriers to be transported within and between cells of the body. The copper transport system in the plasma consists of caeruloplasmin, albumin and some amino acids. Copper absorbed in the intestine is bound to albumin through amino nitrogen and histidine imidazole nitrogen, and is carried to the liver by portal blood. Here it is embedded in caeruloplasmin and released with it into the circulatory system.

10.5.2. Biological functions of copper

The numerous physiological functions of copper are derived from the well-known activities of copper-dependent metalloenzymes. These activities include, first and foremost, the action of caeruloplasmin as ferroxidase I in iron and haemoglobin metabolism, and the lysyl oxidase enzyme, which is a key enzyme in tissue cross-linking.

An important oxidase activity of caeruloplasmin in the plasma acts as an essential means of bivalent (Fe^{2+}) to trivalent (Fe^{3+}) iron oxidation. This allows the mobilisation of reserve iron from ferritin, its incorporation into transferrin and its use for the synthesis of haemoglobin.

Lysyl oxidase catalyzes the oxidation of lysyl and hydroxy lysyl residues in polypeptide chains of elastin and collagen. The condensation of the formed aldol groups produces configurations bridging polypeptide chains which are crucially important for the strength, elasticity and flexibility of major fibrous connective tissue proteins.

Wilson's disease, a genetic, familial disease, represents a separate entity in copper pathophysiology. A pathognomonic sign is the considerably reduced incorporation of copper in caeruloplasmin and reduced copper excretion in the bile. This leads to a positive copper balance and gradual copper accumulation in the liver. Non caeruloplasmin-bound plasma copper increases, which leads to increased excretion of copper in the kidneys, and copper deposits in other extrahepatic tissues such as the cornea, where it forms the well-known Kayser-Fleischer ring, the brain (primarily in basal ganglia regions), kidneys, muscles and bones.

Reduced or even zero caeruloplasmin concentrations are typical in laboratory reports. Serum copper is usually reduced together with caeruloplasmin, but normal or elevated levels may be found, too. Nevertheless, the "free", non caeruloplasmin-bound copper, which is toxic, is indicative of decompensation of the disease with frequently occurring accompanying intravascular haemolysis. Copper excretion in the urine is elevated.

An increased copper content in the liver (more than 300 µg/g of dry tissue) can be proved.

An afflicted individual's siblings must be examined very carefully and rigorously, as there is a higher chance of a copper metabolism disorder in the family. An active approach in diagnosing Wilson's disease can prevent delayed dia-

gnosis and permanent consequences.

10.5.3. Copper deficiency – causes and signs

A severe, purely nutritional copper deficiency with clinical signs is quite rare. Such cases may occur in infants on artificial nutrition, re-alimentation lacking trace elements following severe malnutrition and catabolic conditions, and during long-term, insufficiently balanced parenteral nutrition.

Developed copper deficiencies occur more often due to enteropathies and malabsorptions of different origin.

The most abundant source is liver, lobster meat, oysters, as well as nuts, peas and chocolate.

Menkes disease is a rare inherited copper metabolism disorder with fatal consequences and clinical signs of copper deficiency. The disease is caused by a copper absorption, transport and utilization disorder.

10.5.4. Laboratory diagnosis of copper deficiency

Decreased serum caeruloplasmin and copper levels may warn of copper deficiency. A constant sign of copper deficiency is hypochromic microcytic anaemia not responding to iron administration, and **leucopenia.** Significant alterations can be seen in bones, especially in children. These are signs of osteoporosis, fiberized calyciform metaphyses of long bones and periostal appositions accompanied by an increased serum alkaline phosphatase activity.

10.5.5. Toxicity of copper

If inhaled in the form of aerosol or fine dust, copper induces signs of acute intoxication (“metal fever”) with the same signs as those typical of zinc poisoning: cough, chills, fever, fatigue and headache.

Molecular Level	**(Patho)physiological Functions**	**Clinical Signs**
Cu-metalloenzymes		
Cytochrome C oxidase	Electron transport in respiratory chains	Non-specific, erythropoiesis disorders
Superoxide dismutase	Scavenging superoxide radical	Oxidative stress signs
Monoamine oxidase	Catecholamine metabolism	Non-specific CNS disorders
Tyrosinase	Melanin synthesis	Skin pigmentation disorders
Lysyl oxidase	Connective tissue metabolism, cross-links	Collagen disorders, cutis laxa, laxity of joints
Cu-proteins		
Caeruloplasmin	Acute phase proteins, Fe^{2+} oxidation	Anaemia
Metallothionein	Cu binding and transport in cells	Cu deficiency/toxicity

Table 10.6. Molecular basis of clinical signs of copper deficiency (examples)

10.6. Selenium

Unlike zinc and copper, selenium is a highly controversial element in medical terms. In the 1930s, selenium was identified as a major toxic element in veterinary and human medicine. It was surprising, therefore, when its essentiality was recognized and serious signs of selenium deficiency found. Selenium substitution is anticipated to have a positive effect on tumour incidence in regions with a low intake of selenium, while selenium sulphide is considered carcinogenic, with anti-oxidizing as well as pro-oxidizing qualities. It is evident that its atomic structure provides selenium compounds with diverse, often contrasting physical, chemical and biological properties. Selenium has therefore been subject to intense basic as well as applied research in technical and biological fields.

10.6.1. Biochemistry and metabolism

Selenium is part of numerous metalloenzymes and metalloproteins, where it is mostly found in the form of seleni-

um methionine and selenium cysteine. Selenium methionine is solely of vegetable origin and represents about 50 % of the daily selenium intake. Among other things, it is used for selenium cysteine synthesis in the body.

Between 50 – 60 % of plasma selenium is bound to selenoprotein P, 30 % is found in glutathione peroxidase (GSH-Px-3) and the remaining part is found in selenium methionine.

Reference limits in this population, rather lower than in Europe in general, are 0.8 – 1.4 μmol/l.

The recommended daily intake for adults is 55 μg. Cereals, legumes and in particular nuts are a good source of selenium.

10.6.2. Biological functions

Selenium is involved in human tissue protection against oxidative stress, anti-infection protection, and modulating growth and development.

There are two groups of selenium-dependent enzymes. The first contains glutathione peroxidases in four isoforms and thioredoxin reductase involved in antioxidation protection against reactive oxygen radicals.

The second group consists of three types of selenium-dependent 5'-deiodinases essential for thyroxine (tetraiodothyronine T4) conversion into a physiologically more efficient (active) triiodothyronine (T3).

The three isoforms differ in their quantity and function in different tissues and organs.

Concurrence of selenium and iodine deficiencies is known, as are their synergies in hypothyroidism signs. Selenium is important for spermatogenesis and the proper function of sperm, muscle metabolism as well as the function and immunity of the body.

10.6.3. Deficiency – causes and signs

Nutritional deficiency of selenium fundamentally (significantly) depends on the regional selenium content in the soil at the beginning of the food chain.

The clinical signs of severe selenium deficiency described below do not occur in this population. However, very similar signs are not rare in some livestock breeds in this country.

Kashin-Beck disease is an endemic, chronic osteoarthropathy in children in China, Russia and Korea, with primary involvement of articular cartilages (atrophy, degeneration, necrosis) and marked signs of deformed, arthrotic joints. Another apparent sign is shorter stature due to multifocal necroses in the region of growth plates in long bones. This nosological entity was first described by Kashin in the Baikal region (1848) and Beck in China (1906). The aetiology of the disease is not quite clear yet. Apart from selenium deficiency, the effect of mould-contaminated grain and other factors are considered.

Keshan disease is an endemic disease afflicting mostly children and young women in certain Chinese regions with low selenium content in the soil. The predominant pathological and anatomical finding is multifocal myocardium necroses leading to dilatation of the heart, congestive heart failure and death. Although the incidence of acute and sub-acute cases of the disease has decreased dramatically following selenium supplementation, the aetiological role of selenium has recently undergone change; today it is seen rather as a predisposing factor.

Clinical signs of selenium deficiency, caused mainly by malnutrition in local conditions, are non-specific and complex in the same manner as our definition of the effects of increased oxidative stress on the body remains non-specific.

Populations living in areas with low selenium content in the soil (Nordic countries and also the CR) have a lower selenium concentration in the serum. As a result of worldwide globalization of food sources and large-scale livestock production (feed mixes, etc.), the soil foodweb has partly lost its immediate importance. The discussion on the need for widespread selenium substitution in this population still remains open.

10.6.4. Laboratory diagnosis of deficiency

The commonest method is to determine the selenium concentration in the blood, plasma or serum using atomic

absorption spectrometry. Another method used is the determination of GSHPx activity in erythrocytes or selenoprotein P concentration.

When evaluating the findings of the above methods, one should take into account that all of the parameters are affected by redistribution and plasma level decrease due to stress, trauma, inflammation, etc.

10.6.5. Toxicity of selenium

Of more than 1500 species of *Astragalus* plants, only about 25 are primary accumulators of selenium. They are able to accumulate selenium from soil with selenium levels ranging from normal up to highly toxic concentrations (up to 1000 ppm). Secondary accumulators of selenium are plants accumulating selenium in regions with high selenium content in the soil. If consumed by farm animals, both types of plants cause lethal diseases referred to as alkali disease in beef cattle, and blind staggers in sheep.

The toxic selenium effect in humans manifests itself in the same way as in animals: brittleness, alterations and loss of hair, nail alterations (brittleness, cracks, colour changes), and including the typical garlicky breath odour. Selenium toxicity consists in its interaction with thiols; selenite reacts with glutathione and H_2Se to form a superoxide radical ($O_2^{\cdot-}$). The subsequent peroxidation of biogenic molecules and membranes represents the molecular basis for the toxic effects of certain forms of selenium.

Molecular Level	**(Patho)physiological Functions**	**Clinical Signs**
Se-metalloenzymes		
Glutathione peroxidases (1,2,3,4)	Decomposition of inorganic and organic peroxides	Oxidative stress signs
Thioredoxin reductase	Intracellular redox status	Ditto
5'-deiodinases (1,2,3)	Thyroxine conversion into triiodothyronine	Signs of hypothyroidism
?	Connective tissue metabolism	Connective tissue disorders
Se-proteins		
Selenoprotein P	Se-storage form in plasma	Signs of deficiency
Selenoprotein W	Muscular form of Se	Signs of deficiency
Selenoprotein – sperm	Proper sperm function	Infertility?

Table 10.7. Molecular basis of clinical signs of selenium deficiency (examples)

10.7. Chromium

Chromium plays a role in the normal metabolism of glucose and cholesterol and affects spermatogenesis and fertility. Sufficient valid information is not available for determining the daily intake. Cases of chromium deficiency are rare, perhaps as part of cases of severe malnutrition.

Chromium deficiency may cause glucose tolerance disorders, hypercholesterolaemia, decreased sperm count and fertility disorders.

Cr^{3+} has a high allergenic potential that leads to a high incidence of occupational contact eczema in people working in the building industry (cement), leather industry, printing and metal-working industries.

10.8. Manganese

Manganese forms part of numerous metalloenzymes (arginase, pyruvate carboxylase, Mn-superoxide dismutase in mitochondria, etc.).

Manganese has a non-specific activation effect on some enzymes (hydrolases, decarboxylases and transferases), in which the function of manganese may be replaced by other trace elements such as Fe and Cu.

The signs of the deficiency are non-specific.

10.9. Molybdenum

Molybdenum forms part of three metalloenzymes (xanthine oxidase, aldehyde oxidase and sulfite oxidase).

Congenital molybdenum deficiency with mortality in the first year of life has been described.

10.10. Cobalt

The only function known to determine the essentiality of cobalt is its presence in the vitamin B_{12} molecule (Cobalamin).

Reference limits; output in urine; recommended doses (table)

	Reference Limits (Serum)	24-hr Output in Urine	Rec. Daily Intake (Oral)	Rec. Daily Intake (Parenteral)
Zinc	10.7 – 18.4 µmol/l	3.0 – 12.0 µmol 196 – 780 µg	12 – 15 mg	3.2 – 6.5 mg
Copper	M 11.0 – 22.0 µmol/l F 12.6 – 24.3 µmol/l	0.5 – 0.9 µmol 30 – 60 µg	0.9 – 1.2 mg	0.3 – 1.3 mg
Iron	M 11.6 – 30.4 mol/l F 9.0 – 27.0 µmol/l	1.8 – 3.6 µmol 100 – 200 µg	M 10 mg F 15 (25*) mg	1.2 mg
Selenium	0.7 – 1.24 µmol/l	0.1 – 1.3 µmol 10 – 100 µg	55 µg	30 – 60 µg
Iodine	0.28 – 0.25 µmol/l	0.8 – 1.6 µmol 100 – 200 µg	120 – 150 µg	130 µg
Chromium	1.0 – 10.0 nmol/l	9.6 – 38.5 nmol 0.5 – 2.0 µg	50 – 200 µg	10 – 20 µg
Manganese	7.0 – 13.0 nmol/l	9.0 – 36.0 nmol 0.5 – 0.2 µg	2 – 5 mg	0.2 – 0.3 mg
Molybdenum	1.0 – 30.0 nmol/l	208 – 313 nmol 20 – 30 µg	50 µg	19 µg
Cobalt	1.9 – 7.6 nmol/l	8.8 – 35.1 nmol 0.5 – 2.0 µg		

*Table 10.8. Reference limits, output in urine, recommended oral and parenteral doses of essential trace elements (*pregnancy)*

Content of essential trace elements in pharmaceutical products

Parenteral products (table)

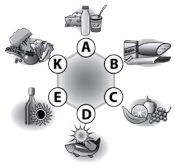
	TRACUTIL [10 ml]	ADDAMEL [10 ml]	ELOTRACE [100 ml]
Zinc	3.3 mg	6.6 mg	6.0 mg
Copper	0.76 mg	1.3 mg	1.2 mg
Iron	2.0 mg	1.1 mg	1.1 mg
Selenium	24 µg	32 µg	12 µg
Iodine	127 µg	127 µg	127 µg
Chromium	10 µg	10 µg	X
Manganese	550 µg	110 µg	300 µg
Molybdenum	10 µg	20 µg	20 µg
Fluorine	570 µg	950 µg	950 µg

Table 10.9. Content of trace elements in products for parenteral administration

Enteral products (table)

	NUTRIDRINK [100 ml]	FRESUBIN [100 ml]	NUTRISON [100 ml]
Zinc	2.9 mg	1.5 mg	1.8 mg
Copper	0.43 mg	0.3 mg	0.27 mg
Iron	3.8 mg	2.0 mg	2.4 mg
Selenium	14 µg	10 µg	8.5 µg
Iodine	132 µg	30 µg	20 µg
Chromium	16 µg	10 µg	10 µg
Manganese	800 µg	400 µg	500 µg
Molybdenum	24 µg	15 µg	15 µg
Fluorine	200 µg		

Table 10.10. Content of trace elements in products for enteral administration



CHAPTER 11

11. Vitamins

11.1. Vitamins – Preparation

Authors: MUDr. Michaela Králíková, Ph.D.; MUDr. Milan Dastych

Reviewer: MUDr. Daniel Rajdl, Ph.D.

11.1.1. Vitamins

Human nutrition consists of primary nutrients (carbohydrates, proteins and fats); in addition to this, food should contain other essential substances which the body is unable to produce by transforming from other components. Such components include essential amino acids, essential fatty acids, mineral compounds and vitamins. Vitamins are classified as either fat-soluble or water-soluble. Among other things, this classification helps estimate what foods the vitamin may be present in, and under circumstances vitamin deficiency can be expected. A balanced diet prepared from fresh foodstuffs usually contains the necessary amounts of all vitamins. A prolonged low supply of some of the vitamins may cause either hypovitaminosis with non-specific symptoms, or more serious disorders with typical symptoms (avitaminosis). Vitamin deficiency may also occur as a result of a disorder of their absorption in the digestive tract, or an increased vitamin requirement during diseases or other conditions (convalescence, pregnancy, breastfeeding). Excess hydrophilic vitamins taken from food are usually quickly excreted from the body in urine. On the other hand, fat-soluble vitamins may deposit in the body for a longer time: vitamins A and D, and vitamin K to a limited extent in the liver, and tocopherols in the adipose tissue. Toxic symptoms can only be caused by repeated high doses of retinol or calciols (sometimes referred to as hypervitaminosis A or D). These almost always occur as a consequence of inappropriate dosage for therapeutic use.

11.1.2. Water-Soluble Vitamins

They do not deposit in the body (except for vitamin B_{12}); excess amounts are excreted in the urine, meaning their consistent intake from food is important. They are enzyme cofactor components in the body.

11.1.2.1. Thiamine (Vitamin B_1)

Thiamine is a pyrimidine and thiazole derivative. Thiamine is fairly unstable in solutions and unstable to heat. Thiamine diphosphate coenzyme is a biologically active form, which binds the reaction intermediate in the course of pyruvate and 2-oxoglutarate oxidative decarboxylation. This makes it essential for degrading all nutrients, carbohydrates in particular. The approximate daily requirement for adults is 1 – 2 mg. The requirement is much greater in the case of excess sugar in the food, profuse sweating when working in the heat, and in alcoholics. Thiamine is present in many vegetable and animal foods; the main source is meat, pork in particular, giblets, yeast, and wholemeal cereals and bread. White bread and glazed rice lose most of their original thiamine content during processing. The initial signs of thiamine insufficiency include increased tiredness, muscle weakness, and proneness to neuritis. Extreme avitaminosis known as beriberi in Southeast Asia is rare here; typical features include various cardiovascular and neurological disorders and gastrointestinal symptoms. This disease occurs if diet consists almost solely of peeled rice.

11.1.2.2. Riboflavin (Vitamin B_2)

Riboflavin consists of a heterocyclic isoalloxazine ring bound to ribitol. It is the component of flavin mononucleotide (FMN) or flavin adenine dinucleotide (FAD), cofactors of flavin dehydrogenases and other oxidoreductases. The daily riboflavin requirement is about 2 mg for an adult. The source of riboflavin in food is milk and dairy products, eggs, meat and a variety of vegetable foods. Riboflavin is usually present as bound in FMN or FAD; only milk contains free

riboflavin. Avitaminosis B_2 is not known. A limited intake usually manifests itself by mucous membrane involvement (inflammation of the lips, tongue, corners of the mouth, corneal edge vascularisation) and facial skin.

11.1.2.3. Niacin (Vitamin B_3)

Niacin is the name for two pyridine derivatives: nicotinic acid and its amide, nicotinamide. Although they may be produced in the human and other mammalian body from tryptophan, niacin belongs to B-group vitamins. The biochemical importance of nicotinamide is in that it is part of pyridine nucleotides, NAD^+ and $NADP^+$, working as coenzymes for hydrogen transfer in reactions catalysed by many dehydrogenases. The daily niacin requirement for a healthy adult is 13 – 20 mg. The intake of sufficient amounts of biologically valuable proteins from milk or eggs covers the niacin requirement despite its very low content, because niacin is produced in cells from tryptophan. Liver, meat and yeast are rich in niacin. Severe niacin deficiency causes pellagra, a disease characterised mainly by diarrhoea, skin inflammation with pigmentation (dermatitis) and psychiatric disorders (dementia) – disease of the three Ds, and cardiac and neurological disorders at a later stage. Pellagra is endemic only where corn (low tryptophan content) is eaten as a staple food.

11.1.2.4. Pantothenic Acid (Vitamin B_5)

Pantothenic acid is an amide formed from beta-alanine by binding the acyl radical of pantoic acid to its amino group. It is plentiful in natural resources (*pantothen* from the Greek "from everywhere"), with sufficient amounts in food. It is present in the blood, in cells mostly as a pantethine component, a part of coenzyme-A and fatty acid synthase multienzyme complex. The daily requirement of about 6 – 8 mg is covered by a varied diet. Rich sources are liver or egg yolks. Pantothenate supplementation is required in complete parenteral nutrition. Deficiency does not display typical symptoms. Calcium salt or alcohol precursor panthenol (dexpanthenolum) may be used as supplements improving the healing of uninfected burns, superficial injuries or catarrh of the upper respiratory tract.

11.1.2.5. Vitamin B_6

Vitamin B_6 is a common name for three related derivatives of substituted pyridine. All of them have the same biological activity, and differ from each other in the group bound at position 4; pyridoxine (pyridoxol) has a primary alcoholic group, pyridoxal an aldehyde group and pyridoxamine a primary amine group. Pyridoxal 5-phosphate cofactor is the active form, a prosthetic group of enzymes primarily contributing in the conversion of amino acids (aminotransferase, decarboxylase). The daily requirement of about 2 mg in adults is covered by a normal, varied diet, but increases greatly during lactation, in all catabolic conditions, or large consumption of ethanol. Absolute deficiency of vitamin B_6 is rare; insufficiency may manifest itself in alcoholics, pregnant or breastfeeding women by skin and mucosa disorders or anaemia.

11.1.2.6. Folic Acid (Folate, Vitamin B_{11})

Folic acid is a common name for pteroyl-glutamic acid. Animals are unable to synthesise these compounds, and depend on their supply from food. Tetrahydrofolate, a cofactor transferring one-carbon residues in different oxidation states, is the active form in cells. The daily folate requirement for an adult is about 0.15 – 0.2 mg. The richest source is liver, vegetable foods such as leafy vegetables (cabbage, kale, spinach) and yeast. Folate insufficiency is usually caused by insufficient resorption, with manifestations in the blood count (megaloblastic anaemia, thrombocytopenia). Lack of folates during pregnancy is associated with congenital developmental defects of the foetus, in particular neural tube defects.

11.1.2.7. Vitamin B_{12} (Cobalamin)

Cobalamin has a fairly complex structure. It is based on the tetrapyrrole-corrin ring with a cobalt ion bound in the centre. It is only produced by microorganisms, and is only found in animal foodstuffs, not in vegetable foods. A specific glycoprotein excreted by the gastric mucosa, called the intrinsic factor, is required to resorb cobalamin supplied from food. Resorbed cobalamin is stored in the liver. When in cells, cobalamin is converted to cofactors methylcobalamin and deoxyadenosylcobalamin. They are required to eliminate some amino acids and, together with folic acid, for methylation reactions (remethylation of homocysteine to methionine, one of the requirements for continuous synthesis of nucleic acid base pairs). The daily requirement of B_{12} is about 2 µg in adults. Rich sources are giblets, followed by meat, dairy products and eggs. Cobalamin is also formed by the gut flora, but does not absorb and is excreted in the

stools. It is relatively rare for cobalamin deficiency to be caused by an insufficient diet such as vegans who avoid foodstuffs of animal origin. Intrinsic factor deficiency and insufficient cobalamin resorption from food are more common causes. Manifestations include megaloblastic or macrocytic anaemia or neurological disorders.

11.1.2.8. Vitamin C (Ascorbic Acid)

L-ascorbic acid is a monosaccharide derivative. This is not carboxylic acid; the distinct acidic character is derived from the presence of two enolic hydroxyles. It is synthesized by plants and most animals except for humans, apes and guinea pigs. It is an important hydrophilic reducing agent (antioxidant). L-ascorbic acid readily oxidized to biologically inactive L-dehydroascorbic acid; the reaction is reversible, however a sufficiently effective reducing agent for reverse reduction is usually missing. It is involved in hydroxylation reactions in the synthesis of collagen, bile acids and adrenalin, and in tyrosine degradation by keeping metalloenzyme metal ions in the reduced state. Vitamin C also increases iron resorption in the digestive tract. Unlike other vitamins, the daily requirement is one or two orders higher, 50 – 70 mg for adults. L-ascorbate is mainly present in plants, namely fresh fruits and vegetables. The richest source is citrus fruit, kiwi fruit, blackcurrant, peppers, potatoes. Ascorbate is readily deteriorated by oxidation from atmospheric oxygen at higher temperatures, especially in the presence of traces of heavy metals, Cu and Fe. People with a relatively low vitamin C intake get easily tired, more prone to minor infections, sometimes have swollen and bleeding gums or hypodermic haemorrhage from fragile capillaries. Extreme avitaminosis (scurvy, *scorbutus*) is currently rare. Prolonged ingestion of high ascorbate doses (over 0.5 g daily) is not considered beneficial. Excess vitamin C is quickly excreted in urine; larger ascorbate amounts are degraded to oxalate (risk of urolithiasis), and there may be a higher risk of acid-base disorders (acidosis). Ascorbate as an antioxidant protects against the action of free oxygen radicals however too high concentrations support the presence of reduced forms of metal ions, which catalyse the formation of aggressive oxygen radicals (Fenton's reaction).

11.1.2.9. Biotin (Formerly Vitamin H)

The biotin molecule contains a condensed heterocyclic system consisting of thiolane and imidazolidine. It acts as a cofactor of carboxylases catalysing carboxylation reactions. The nitrogen heteroatom binds a CO_2 molecule to form carboxybiotin, which is a donor of the carboxy group for carboxylation of compounds such as pyruvate to oxaloacetate, or acetyl-coenzyme-A to malonyl-coenzyme-A. It thus plays an important role in gluconeogenesis (synthesis of glucose from lactate and some amino acids), for the citric acid cycle (synthesis of oxaloacetate), and for synthesis of fatty acids. Biotin is found in most foodstuffs. In addition to this, a considerable part of the required amount is synthesised by the large bowel flora. Biotin deficiency is therefore quite rare in humans. It can be induced by excess consumption of raw egg whites. They contain a thermally unstable glycoprotein avidin, a specifically biotin-binding protein (the complex is not broken down by digestive enzymes) that prevents biotin absorption.

11.1.3. Fat-Soluble Vitamins

They have a non-polar (hydrophobic) character, since they are isoprenoids themselves (A and D), or their molecule contains an isoprenoid chain (E and K). In vegetable and animal foodstuffs they are usually dissolved in fats that are absorbed through the intestines. Long-term fat digestion disorders lead to reduced fat resorption, and hypovitaminosis may develop despite sufficient intake of fats from food. Fats enter the circulation as part of chylomicrons.

11.1.3.1. Vitamin A (Retinol)

Retinol ranks among isoprenoids. Retinol is produced by oxidative cleavage of beta-carotene provitamin present in vegetables and fruit. The liver has some retinol reserves in the form of its esters. Retinal aldehyde is part of rhodopsin (visual purple in the rods and cones of the retina). Retinoic acid in cells plays a role in controlling the expression of some genes. One of the first signs of retinol deficiency is night blindness followed by skin and mucosa disorders (dryness and keratinisation), and conjunctiva and cornea damage (xerophthalmia). Repeated multiple doses of retinol are toxic and may, among other things, pose a risk to foetal development during pregnancy. Major sources in food are liver, oily fish, yolk and butter. Carotenes are contained in yellow, red and green vegetables and fruit.

11.1.3.2. Vitamin D

Calciol (vitamin D_3, cholecalciferol) and ercalciol (vitamin D_2, ergocalciferol) are 9,10-secosterols. Calciol may form in the skin by 7-dehydrocholesterol exposure to UV radiation, and similarly, ercalciol forms in plants from ergosterol.

Their hydroxylation in the liver and kidneys produces the hormone calcitriol (1,25-dihydroxycalciol) which, together with parathormone and calcitonin, plays a major role in the Ca^{2+} and phosphate management. Calcitriol induces the synthesis of a protein, which enables resorption of Ca^{2+} in the intestine, and regulates bone tissue mineralisation. Calciol deficiency results in insufficient mineralisation of the organic bone component. Serious deficiency is manifested as rickets in children, and osteomalacia in adults. Repeated administration of excessive vitamin D doses leads to hypercalcaemia, sometimes even undesirable calcifications in various tissues or calcium urolithiasis. The main source is fish oil, cod liver, oily fish, and also enriched tub margarines.

11.1.3.3. Vitamin E

Vitamin E refers to a group of natural antioxidants featuring qualitatively the same biological activity as alpha--tocopherol, which is the most effective and most naturally widespread of them. Vitamin E was first isolated from wheat-germ oil. All of these antioxidants are derivatives of a chromate with a 16-carbon isoprenoid residue attached at position 2. The phenolic hydroxyl at position 6 is essential for its function. Some of the free radicals are neutralised by the hydrogen atom, and the phenoxy-radical thus formed is either regenerated by ascorbate to the original tocopherol, or is irreversibly degraded by another radical. Tocopherols reduce peroxidation of membrane lipids and lipoproteins unsaturated fatty acids, and also prevent other important compounds from oxidation. The daily requirement is about 15 mg of alpha-tocopherol, and increases by boosting intake of poly-unsaturated fats. Rich sources include vegetable oils, nuts, seeds, and cereal germs. The tocopherol content is greatly reduced in dehydrated or frozen foods due to auto-oxidation. Tocopherol deficiency does not present typical manifestations in humans; perhaps may cause anaemia and some neurological symptoms.

11.1.3.4. Vitamin K

Vitamin K is a common name for a group of naphthoquinone derivatives referred to as antihaemorrhagic vitamin. Naturally occurring are phyloquinone (K_1, 3-phytylmenadione) and menaquinones (K_2) with an attached polyisoprene side chain, which makes them fat-soluble. Synthetic, water-soluble menadiol or menadione has the same effect. They are essential for the final biosynthesis of blood coagulation factors II (prothrombin), VII, IX and X, and proteins important for bone mineralisation due to their contribution to forming binding sites for Ca^{2+} (by introducing another carboxy group to side chains of glutamic acid residues). This explains why they are used to prevent or treat increased bleeding. Their deficiency in adults is not common since vitamin K is relatively plentiful in foods (most in leaf vegetables), and is also produced by large bowel flora. A bleeding tendency due to vitamin K deficiency may occur in newborns (their gut is sterile), in all fat resorption disorders, most commonly in insufficient bile acid secretion, or in the absence of gut flora caused by broad-spectrum antibiotics. Anticoagulants (warfarin) are important vitamin K antagonists.

11.2. Vitamins

In addition to primary nutrients (carbohydrates, proteins and fats), human nutrition has to contain other essential components that the human body is unable to synthesize. Such components include essential amino acids, essential fatty acids, mineral compounds and vitamins. Vitamins and cofactors derived from these are indispensable for the normal course of many metabolic processes. Depending on their solubility, vitamins are classified as either fat-soluble (A, D, E, K) and water-soluble (the other vitamins). This classical classification is also practical – among other things it helps estimate in what foods the vitamin may be present, and under what circumstances vitamin deficiency can be expected. A balanced diet prepared from fresh foodstuffs usually contains the necessary amounts of all vitamins.

A prolonged low supply of some of the vitamins may cause either hypovitaminosis, at first subclinical, when a reduced amount of that particular vitamin in the body can be proved (common in vitamins D or B_{12}), and later clinical hypovitaminosis with non-specific symptoms, or more serious disorders with typical symptoms (avitaminosis). Vitamin deficiency may also occur as a result of a disorder in absorbing these in the digestive tract, or an increased vitamin requirement when suffering from diseases or other conditions (convalescence, pregnancy, breastfeeding), or their increased excretion.

Excess hydrophilic vitamins taken from food are usually excreted quickly in urine from the body (with the exception of vitamin B_{12} – see below). On the other hand, fat-soluble vitamins may deposit in the body for a longer time: vitamins A and D, and vitamin K to a limited extent in the liver, and tocopherols in the adipose tissue.

Toxicity manifestations (sometimes referred to as hypervitaminosis) can only be found in vitamins A and D. They almost always occur as a consequence of wrong dosage for therapeutic use.

11.2.1. Recommended Daily Allowances of Vitamins

The Recommended Daily Allowance (RDA) is the daily dietary intake level of a nutrient considered sufficient to meet the requirements of 97% – 98% of healthy individuals in each life-stage and gender group. RDA is derived from research studies on the physiological needs of healthy people. For example, the RDA of a vitamin is determined from the amount required to maintain its adequate blood level in a tested healthy person.

These values may have different interpretations, and may be more or less varied between countries. Neither are RDA values recommended for optimum nutrition. Although a lot of research has been carried out, the definitive approach to optimum vitamin intake is still being debated (vitamin D is currently a good example).

Vitamin	Unit	Amount
B_1	mg	1.1
B_2	mg	1.4
Niacin	mg	16
Pantothenic acid	mg	6
B_6	mg	1.4
Folic acid	µg	200
B_{12}	µg	2.5
Vitamin C	mg	80
Biotin	µg	50
A	µg	800
D	µg	10 (400 IU)
E	mg	12
K	µg	75

Table 11.1. RDAs of vitamins pursuant to Decree No. 225/2008 Coll.

http://eagri.cz/public/web/mze/legislativa/ostatni/100065067.html

11.2.2. Vitamin Test Methods

There are different approaches to determining the vitamin level in the body, when vitamin deficiency is suspected. The vitamin may either be tested directly (most often in the serum, however plasma, whole blood, RBC, WBC or biopsy tissue sample level can also be tested), by urinary vitamin excretion tests (without or after load), or tested using indirect methods. Such methods include tests for typical metabolite of the vitamin, or a typical product of the vitamin's activity (after load), and tests for catalytic activity of a suitable enzyme, whose cofactor is that vitamin. The specific options for each vitamin are listed.

11.2.3. Water-Soluble Vitamins

These vitamins are not stored in the body (except for vitamin B_{12}); excess amounts are excreted in the urine, meaning so their consistent intake from food is important. They are enzyme cofactor components in the body. Hydrophilic vitamins (again except for vitamin B^{12}) are absorbed through the saturable active Na^+-dependent multivitamin transporter; simple diffusion also occurs in the case of higher concentrations in food.

11.2.3.1. Thiamine (Vitamin B_1)

Thiamine is the oldest known vitamin – since 1905. It is a derivative of pyrimidine and thiazole linked by a methylene bridge. Thiamine diphosphate (TDP) is the biologically active form.

Major sources of thiamine in food are cereal germs, yeast, legumes, nuts, meat, milk and dairy products. By contrast, its content in white bread and glazed rice is low, so cereals and bread are enriched with thiamine in many countries (including the CR).

Thiamine is readily absorbed from food. Absorption takes place in the upper intestine using two mechanisms. Active saturable transport by the thiamine Na^+-dependent transporter (symport) takes place when low concentrations are involved. Passive diffusion transport takes place when the daily intake is greater than 5 mg. This is important for alcoholics since alcohol only inhibits the active transport.

Animation: Absorption of vitamin B_1 (low vitamin – symport with Na, high vitamin - transporter is saturated, diffusion begins)

Thiamine is transported through the portal vein to the liver where it is phosphorylated by the thiamine disphosphotransferase enzyme (also present in the brain) into active TDP. ATP is the source of diphosphate. 90% of thiamine transported in the plasma is free thiamine. The body contains roughly 30 mg of thiamine, about one half in the muscles, and the rest in the myocardium, liver, kidneys and nerve tissue. 90 % occurs as TDP, and 10 % in the nerve tissue as thiamine triphosphate. Excess thiamine is excreted in urine.

TDP is a cofactor for oxidative decarboxylation of 2-oxo acids (examples: conversion of pyruvate → acetyl-CoA, 2-oxoglutarate → succinylCoA, citric acid cycle; 2-oxo acids from valine, leucine and isoleucine; 2-oxobutyrate in methionin catabolism, etc.) and the transketolase reaction of the pentose cycle. It is therefore essential for carbohydrate and other nutrient metabolism. If TDP is deficient, pyruvate cannot metabolize to AcCoA, the level of lactate rises and lactate acidosis refractory to hydrogen carbonate develops. Accumulation of 2-oxo acids produced from branched amino acids leads to maple syrup disease symptoms. Thiamine also stimulates the function of neutrophils and lymphocytes; when deficient, proneness to infections develops.

Thiamine deficiency often occurs due to malnutrition, chronic alcoholism, administration of diuretics or malaria. Typical causes of insufficient intake include consumption of unfortified glazed rice or wheat products and also raw fish containing microbial thiaminase. Deficiency in alcoholics develops due to reduced intake, and also due to limited absorption as a result of active transport inhibition. In old age, insufficient intake typically combines with the administration of diuretics mentioned above (mainly for cardiac failure treatment).

Signs of deficiency include mood changes (irritability, depression), memory disorders, peripheral neuropathy; severe deficiency results in beriberi. Manifestations of beriberi include confusion, gastrointestinal symptoms (loss of appetite, poor digestion, weight loss), muscle weakness, ataxia, peripheral paraesthesia, ophthalmoplegia, oedemas (wet beriberi), tachycardia, cardiac failure (dry beriberi) or even coma. Typical manifestations in alcoholics include Wernicke-Korsakoff encephalopathy (loss of memory, confabulation, delirium, nystagmus, ataxia + beriberi symptoms).

The test for transketolase in erythrocytes reveals information about the thiamine reserve in tissues. Enzyme acti-

vity is measured in RBC haemolysate before and after TDP addition.

11.2.3.2. Riboflavin (Vitamin B_2)

A riboflavin molecule consists of the heterocyclic isoalloxazine ring attached to alcohol ribitol. It is a basic component of FAD and FMN flavin nucleotides used as the prosthetic groups of many dehydrogenases and other enzymes needed for electron transport.

Bound riboflavin is present in food in the form of FAD and FMN; free riboflavin is contained in milk. Major sources are dairy products, eggs, liver, meat, broccoli, parsley, yeast.

Through the action of gastric HCl, flavinoids taken from food released from the noncovalent bond to proteins and the action of intestinal enzymes releases free riboflavin. Riboflavin is mainly resorbed in the jejunum through the saturable Na^+-dependent active transport. Resorption is boosted by bile acids, reduced by metals (Cu, Zn, Fe), vitamin C, nicotinamide, tryptophane, caffeine, theophylline and some drugs (antibiotics, antacid drugs etc.). Blood transports B_2 bound to albumin, immunoglobulins and other proteins. B_2 is converted to cofactors in the cells of most tissues, in particular enterocytes, the liver, myocardium and kidneys. FMN is first produced by phosphorylation through the action of riboflavin kinase in the presence of ATP, and a greater part further converts to FAD through the action of ATP-dependent FAD synthase. The body has a low B_2 reserve; excess vitamin is excreted in the urine, meaning renal excretion reflects the daily intake of B_2.

FAD and FMN take part in many redox processes – respiratory chain (cytochrome C reductase), citric acid cycle (succinyl CoA-dehydrogenase), synthesis and elimination of fatty acids, metabolism of purins (xanthine oxidase), some amino acids, xenobiotics (part of CYP 450), part of glutathione reductase (anti-oxidative function), peroxidase, 2-oxo acid oxidases and other enzymes.

Riboflavin deficiency is usually accompanied by a deficiency in the other B-group vitamins. Typical symptoms include angular stomatitis (inflammation of the corners of the mouth), cheilosis, glossitis, swelling of mucous membranes, seborrhoeic dermatitis and other dermal defects, conjunctivitis, corneitis and nervous disorders.

When riboflavin deficiency is suspected, glutathione reductase activity in haemolysed erythrocytes is determined before and after supplying FAD.

11.2.3.3. Niacin (Vitamin B_3)

Niacin refers to pyridine derivatives, nicotinic acid and its amide - nicotinamide, a structural component of many NAD and NADP dehydrogenase cofactors.

Good sources are foods such as meat (turkey, tuna), liver, sunflower seeds, peanuts, brown bread, legumes, however a major part of niacin is produced endogenously from tryptophan.

Nicotinic acid absorbs in the stomach, and nicotinamide also absorbs in the small intestine. The process is referred to as saturable Na^+-dependent active transport. When the level in food is high and the transporter saturated, passive diffusion transport is also possible. Nicotinamide is the most common form circulating in the blood. Hepatocytes contain a small amount of free nicotinamide, and transformation into nucleotides takes place in other cells. Nicotinamide deamination to nicotinic acid and production of desamidoNAD^+ take place in cytosol. DesamidoNAD^+ is adenylated using ATP to active NAD^+. $NADP^+$ is phosphorylated in mitochondria. The body has a low reserve; excess niacin is excreted primarily as N-methyl nicotinamide produced by methylation in the liver.

Target organs include all tissues, liver and muscles in particular. Nicotinamide cofactors are essential for redox reactions in the citric acid cycle, respiratory chain, glycolysis, metabolism of fatty acids, amino acids and many other processes; they affect the function of more than 200 enzymes. They also play a role in vasodilation, non-redox reactions in DNA replication and reparation, apoptosis, increase the HDL level, and decrease the LDL, cholesterol and triglyceride level.

Deficiency occurs due to insufficient intake from food, alcoholism, cirrhosis, low intake (corn-based diet) or tryptophan absorption disorders such as carcinoid, when tryptophan excessively (up to 60 % vs. normal 1 %) converts to serotonin, 5-hydroxytryptophan and 5-hydroxyindoleacetic acid. Acute deficiency may occur as a result of isoniazid treatment (competitive inhibition of pyridoxal phosphate needed converting tryptophan to niacin).

Pellagra is a typical vitamin deficiency disease – disease of the three Ds – dermatitis, diarrhoea and dementia. Pigment rash primarily affects areas exposed to sun; also the oral cavity and tongue mucosa are affected. Diarrhoea develops as a result of mucosal inflammation; other neurological symptoms include depression, irritability, disorientation, hallucinations or catatonia.

Niacin is used as a drug for dyslipoproteinemia (antiatherogenic activity), carcinoid and congenital tryptophane absorption disorder (Hartnup disease). It can be used as a supplement for diabetics (improves tolerance of saccharides), or patients with vascular diseases. Clinical monitoring is advisable in respect to overdose symptoms such as vasodilatation (reddening, tingling, headache, nausea, vomiting), hepatitis, myopathy and thrombocytopathy.

Determination of urinary excretion of N'-methyl nicotinamide and N'-methyl-2-pyridone-5-carboxylamide metabolites using liquid chromatography is used to assess deficiency.

11.2.3.4. Pantothenic Acid (Vitamin B_5)

Pantothenic acid is a six-carbon branched-chain pantoic hydroxy acid bound to alanine. Coenzyme-A and acyl carrier protein (ACP) are its active forms.

It is plentiful in foodstuffs such as animal foodstuffs, wholemeal products, vegetables, legumes. It is mainly found as a component of coenzyme-A (CoA) and its esters. They are hydrolysed in the intestine and pantothenate is subsequently absorbed through the saturable active Na^+-dependent multivitamin transporter; simple diffusion also occurs in the case of higher concentrations in food. Pantothenic acid is transported in blood to hepatocytes and other tissues (muscles, kidneys, adrenal glands, testes) where it enters through a mechanism similar to that used for enterocytes, and synthetises CoA or ACP here.

CoA and ACP are needed for transporting acyl residues, ACP in fatty acid synthesis, CoA in many reactions: the citric acid cycle, oxidation and synthesis of fatty acids, synthesis of cholesterol, glycolysis, gluconeogenesis, catabolism of amino acids, xenobiotic acetylating reactions, haem synthesis and other reactions. Pantothenate increases resistance to infections and allergies and also improves wound healing.

Isolated deficiency does not occur in humans; reduced serum levels can be found in anorexia nervosa, malabsorption syndrome, sepsis, immunodeficiency.

Therapeutic administration is used for malnutrition, as a supplement in catabolic conditions or ethanol intoxication.

Laboratory tests include serum, whole blood and urine tests while CoA can be measured in erythrocytes; these are scarcely used in practice, however.

11.2.3.5. Vitamin B_6

The term vitamin B_6 includes pyridine derivatives: pyridoxine (pyridoxol), pyridoxal, pyridoxamine and their 5-phosphates. The active form of the vitamin is pyridoxal-5'-phosphate (PLP) and pyridoxamine-5'-phosphate produced by ATP-dependent phosphorylation by pyridoxal kinase in the cytosol.

Vitamin B_6 is widely distributed in animal and plant tissues with predominating PLP and pyridoxamine-5'-phosphate; plants also contain pyridoxin-5'-glucoside with a different mode of absorption. Vitamin B_6 occurs in the form of pyridoxin hydrochloride in drugs. Good sources are foods such as giblets, poultry meat, fish, potatoes, nuts, legumes, bananas, yeast and wholemeal cereals.

Phosphorylated forms taken from food are hydrolyzed by intraluminal alkali phosphatase in the small intestine, and unphosphorylated vitamin is absorbed by passive diffusion primarily in the jejunum. Pyridoxin-5'-glucoside is hydrolysed intracellularly by non-specific glucosidases. Phosphorylation takes place in enterocytes. Unused PLP is dephosphorylated again in the enterocyte (non-specific alkaline phosphatase of the plasma membrane), and enters the liver through the portal vein and to the circulation. Here it is transported as pyridoxal attached to albumin or haemoglobin in red blood cells. PLP is the main form present in tissues and muscles are the main reserve of B_6 (about 80 %). All forms of B_6 are excreted in urine, 4-pyridoxine acid (PA) being the main metabolite in high concentrations.

PLP is a cofactor for more than 100 enzymes – amino transferases (ALT, AST), amino acid decarboxylases (formation of catecholamines and serotonin) and deaminases, transsulphurases, cystathionine beta synthase (CBS, homocysteine

metabolism), glycogen phosphorylase and other enzymes. It is also essential for gluconeogenesis, synthesis of haem, myelin, DNA and taurine, modulates hormone effect and is important for immune functions.

Deficiency is rare and occurs as part of a mixed deficiency in other B-group vitamins. Symptoms may manifest themselves as hypochromic sideroblastic anaemia, dermatitis with cheilosis and glossitis, follicular hyperkeratosis, inflammations of the corners of the eye or epilepticogenic spasms. Clinical deficiency of PLP-dependent enzymes such as CBS (homocystinuria – mental retardation, skeleton deformities, thromboembolism) is also known. Vitamin B_6 (under 300 mg/day) is administered in all of these conditions as well as in treating tuberculosis with isoniazid or together with penicillamine (treatment for Wilson's disease or cystinuria).

Excess B_6 (from vitamin preparations) may lead to the development of neurotoxic and photosensitive symptoms; taking more than 100 mg B_6 per day is therefore not recommended.

There are direct and indirect methods for testing the B_6 level in the body. Direct methods include the whole blood PLP and urinary 4-PA excretion assays, mainly using HPLC. The determination of enzyme activity of erythrocyte ALT and AST in the haemolyzate before and after adding PLP is an indirect method.

11.2.3.6. Folic Acid (Folate, Vitamin B_{11})

Folic acid is pteroyl-glutamic acid, a substituted pteridine with p-aminobenzoic acid (PABA) and glutamic acid. The active form in the body is tetrahydrofolate transferring one-carbon residues. Among other things, it is required for synthesising nucleic acids.

Animals cannot synthesize PABA and have to take folate in their food. The main sources are liver, kidneys, leaf vegetables and citrus fruit.

Folic acid is taken in the form of polyglutamate from food. Intestinal mucosa enzymes hydrolyze it into monoglutamate, which is absorbed through active transport (low concentrations) and diffusion (higher concentrations). It is then reduced and methylated in enterocytes, and 5-methyltetrahydrofolate is released into the circulation. It is transported as bound to albumin and folate-binding protein. The main target organ is the liver, however folates are utilised in most tissues. Cells are the site of demethylation and conversion into polyglutamate through the action of polyglutamate synthase, which helps keep folates within the cells, because reverse conversion to monoglutamate is required for release into the circulation. The folate reserve in the body covers 2-4 months. Folates are excreted in the urine and bile via the enterohepatic circulation.

The main function of folic acid in the body is the transfer of one-carbon residues. They are used for the synthesis of purine nucleotides, dTMP and in converting homocysteine to methionine in the form of S-adenosyl methionine, which is the main and universal donor of methyl in methylation reactions (synthesis of creatine, adrenaline, phosphatidycholine, carnitine, methylation in nucleic acids).

Folic acid deficiency commonly occurs due to malabsorption (resection, celiac disease, inflammatory bowel diseases), increased losses (dialysis), decreased intake from food (chronic alcoholism), increased requirement (pregnancy, lactation, hepatic diseases, malignancies), and drug interactions (methotrexate, phenytoin). The condition leads to cell division disorders, typically manifested as pancytopenia with megaloblastic anaemia. Gastrointestinal, neurological and psychiatric symptoms are less frequent. Folate deficiency in the 1st trimester of pregnancy is associated with neural tube defects (spina bifida, encephalocele, anencephaly), so supplementation should be started in the preconception period.

The serum and erythrocyte level are determined to examine the folate supply. Erythrocyte concentration is more adequate for monitoring the long-term supply. Just as is the case for vitamin B_{12}, chemiluminescence immunoassay is used, while testing both vitamins together is advisable.

11.2.3.7. Vitamin B_{12} (Cobalamin)

The term of vitamin B_{12} includes a group of cobalamins differing from each other in the side group type bound to cobalt in the corrin ring. Active forms of the vitamin are methylcobalamin and deoxyadenosylcobalamin; hydroxycobalamin and cyanocobalamin are taken from the diet or therapeutic drugs. All of these forms are of microbial origin; the main source is solely animal foodstuffs – meat and dairy products. Although B_{12} is also produced by the large bowel flora, it does not absorb and is excreted in the stools.

Vitamin B_{12} released from food binds to the protein haptocorrin in the stomach. The complex disintegrates in the small intestine, and haptocorrin is broken down by pancreatic enzymes. The released B_{12} binds to the intrinsic factor (IF) – a glycoprotein secreted by parietal cells of the gastric mucosa. The B_{12}-IF complex in the distal ileum enters the enterocyte by endocytosis via the CUBN receptor (cubulin; in the epithelium of the ileum, proximal renal tubules and yolk sac; the cubulin-megalin complex = multiligand receptor for the transport of most proteins such as transferrin in reabsorption in the kidneys). The complex disintegrates in the enterocyte, and free B_{12} binds to holotranscobalamin (HTC, formerly referred to as transcobalamin II). The B_{12}-HTC complex leaves the enterocyte by active transport through the transporter, and enters the liver through the portal vein. B_{12} is stored here and released into the circulation if necessary. Here it occurs in a complex with HTC (active B_{12} available to cells, half-life 5 – 6 minutes) or haptocorrin (80 – 85 %, a strong bond lasting a few months to one year, half-life > 2 weeks). Approximately 1 mg B_{12} is normally stored in the liver, which covers the body's normal requirements for about 2000 days (5.5 years). Symptoms of deficiency thus occur after up to five year's latency.

B_{12} is constantly excreted to bile and reversely resorbed from the intestine – enterohepatic circulation amounts to 70 – 80 %. If the amount of circulating vitamin B_{12} exceeds the HTC capacity, it is excreted in the urine. Roughly 0.2 % of the total daily reserve is excreted in bile, stools and urine.

Vitamin B_{12} is a cofactor of many enzymes, of which methionin synthase (MS) and methylmalonyl-CoA mutase are most important. Methylcobalamin is a cofactor for MS. This enzyme is essential for converting homocysteine to methionine, converting folate coenzymes, and all methylation reactions in the body. Its deficiency leads to the development of hyperhomocysteinaemia and folate trap (accumulation of N_5-methyltetrahydrofolate which cannot give its methyl group to B_{12}, while levels of other folate forms required for the synthesis of nucleic acid bases decrease). A cofactor for methylmalonyl-CoA mutase is deoxyadenosylcobalamin. This enzyme is essential for the catabolism of some amino acids and formation of succinylCoA (anaplerotic reaction of the citric acid cycle). Its deficiency leads to the development of methylmalonic aciduria and metabolic acidosis. Vitamin B_{12} is required for the continuous synthesis of nucleic acids, cell proliferation, formation and stability of membranes, synthesis of myelin, adrenaline, creatine and other compounds, RBC maturation and haematopoiesis, while it also has an antipernicious effect.

Deficiency occurs due to malnutrition (vegans), malabsorption (typically in atrophic gastritis when IF is lacking, or intestinal inflammations), chronic pancreatitis and interaction with some drugs and alcohol. Deficiency is classified in four stages: 1. ↓ plasma concentration; 2. ↓ cell concentration; 3. metabolic disorders (hyperhomocysteinemia, homocystinuria, acidosis, etc.); 4. clinical manifestation of the deficiency. This includes megaloblastic, pernicious anaemia (blockage of DNA synthesis in the bone marrow leads to the accumulation of megaloblasts) and neuropathy (paresthaesia, pain in the limbs, hypotonia, spasms, paralysis, position sense disorders, confusion, memory disorders, bradypsychism, dementia, depression).

A chemiluminescence immunoassay is most commonly used for the laboratory determination of serum concentration. As B_{12} function is linked to the folate function, the two vitamins should be tested at the same time.

11.2.3.8. Vitamin C (Ascorbic Acid)

Ascorbic acid is a derivative of monosaccharides, a gulonic acid lactone. It is an important antioxidant due to the presence of two enolic hydroxyls. It converts into biologically inactive dehydroascorbic acid during the reaction. The reaction is reversible, however the agent for reverse reduction to ascorbate is usually missing. Vitamin C is synthesized by plants and animals except for humans, apes and guinea pigs.

Major local food sources include potatoes, peppers, sauerkraut, citrus fruit, strawberries and blackcurrant. Vitamin C is sensitive to freezing, prolonged heating, drying, salting, exposure to oxygen, light and metal ions (Cu, Fe, Al and Pb in particular).

Vitamin C absorbs in the jejunum and ileum via the Na^+-dependent active transport. This also plays a main role in renal tubules; vitamin C enters other cells in transporters similar or identical to the glucose transporter GLUT-1 (facilitated diffusion); simple diffusion plays a minor role. This vitamin does not accumulate in the body except for the adrenal glands. The highest concentrations can be found in growing and young cells, particularly glands, leukocytes, liver and the eye lens. The main metabolite is oxalic acid. This is important when high doses of this vitamin are consumed, which leads to the risk of oxalate urolithiasis. Unused vitamin C is excreted in the urine.

Normal doses of vitamin C have antioxidant effects; vitamin C reduces ferric ions to ferrous ions thus helping their absorption, protecting against oxidation (adrenaline), is essential for the function of many hydroxylases such as lysine and proline (important for the formation of cross-linking bridges in collagen), phenylalanine and tyrosine (their cata-

bolism + formation of catecholamines), cholesterol (conversion to bile acids), formation of steroid hormones, serotonin, metabolism of folic acid, histamine, carnitin, etc. It is essential for functioning of the immune system, protecting against the toxic effects of metals.

Long-term use of high doses of vitamin C however leads to manifestations of toxicity. The risk of oxalate urolithiasis increases. The body becomes acidified (the risk of metabolic acidosis may increase) and theoretically osmotic diruresis may occur. Excessively formed reduced metal ions (Fe^{2+} in particular) increase the production of reactive oxygen forms, meaning a pro-oxidative effect of the vitamin prevails. Haemolytic anaemia may develop in patients with a glucose-6--phosphate dehydrogenase deficiency.

Vitamin C deficiency manifests itself as fatigue, increased proneness to infections, worse healing of wounds, weakness, muscle pain, loss of appetite, disorders of the mucous membranes, depression, and hypercholesterolemia as a result of reduced conversion to bile acids. Scurvy is a typical deficiency disease. Symptoms include bleeding from the gums and mucous membranes, loosening of the tooth attachment apparatus, haematomas and anaemia.

When vitamin C deficiency is suspected, its plasma level is tested (mostly using HPLC) even though this test does not offer information about vitamin C reserve in the tissues. This is measured in leukocytes. Leukocytes and thrombocytes take up ascorbate from the plasma even against the concentration gradient. Another option is the saturation test: at least 50 % of ascorbate is excreted following administration of 500 mg ascorbate; deficiency leads to lower amounts.

11.2.3.9. Biotin (Formerly Vitamin H)

Biotin is a derivative of 2 condensed heterocycles, imidazolidine and thiolane, with valeric acid.

Foods rich in biotin include yolks, liver, chocolate, yeast, cereals, legumes, sea fish or nuts. A great portion of the human requirement is covered by biotin synthesis from the gut flora. Biotin is mostly bound to proteins in food. These are cleaved to biotinylated peptides, from which biotin is released by biotinidase. Egg white contains a thermally unstable protein avidin which binds strongly to biotin, thus preventing its absorption. Absorption involves a saturable active Na^+-dependent multivitamin transporter (competition with pantothenate). Biotin uses the same mechanism to enter cells, mainly in the liver, muscles and kidneys, where it is present in cytosol and mitochondria as a cofactor for carboxylases. The active form is enzyme-bound carboxybiotin requiring HCO_3^-, ATP, Mg^{++} and AcCoA to be produced. Major carboxylation reactions include acetyl-CoA carboxylase, fatty acid synthesis, production of oxaloacetate and succinyl-CoA (anaplerotic reaction of the citric acid cycle). Biotin is also essential for cholesterol and leucin metabolism, gluconeogenesis and cell growth.

Deficiency may occur due to malnutrition, poor parenteral nutrition or a large intake of raw egg white. Biotinidase deficiency is rare. Clinical symptoms include dermatitis, alopecia, depression, nausea and vomiting.

Biotin mitigates muscle pain, prevents grey hair and hair loss, improves nail quality and helps treat skin diseases.

Laboratory tests include direct determination of serum/plasma concentration, or indirect determination of biotin--dependent enzyme activity. Urinary metabolite (3-hydroisovaleric acid) levels may be used to measure renal excretion.

11.2.4. Fat-Soluble Vitamins

These vitamins in vegetable and animal foodstuffs are dissolved in fats which are absorbed through the intestines. Hypovitaminosis may therefore also occur due to prolonged malabsorption of lipids despite sufficient intake of fats from food. They are absorbed in the proximal small intestine, absorption is facilitated by bile acids. Vitamins in enterocytes become part of chylomicrons which are released into the lymph and then the blood.

11.2.4.1. Vitamin A (Retinol)

Vitamin A ranks among isoprenoids. Its molecule contains 4 double bonds in the side chain which may form corresponding cis and trans-isomers, of which only two are biologically active.

Vitamin A is present in food in the form of retinol or carotenoids, of which beta-carotene is the best known and is 6 times less efficient than retinol. A part of beta-carotene is hydrolysed to retinol in the intestine by the beta-carotene 15,15'-dioxygenase enzyme. Therefore, beta-carotene is a provitamin of vitamin A. Liver, yolk, rich dairy products and fish are sources of retinol. Carotenoids are present in leaf, yellow and orange vegetables and fruits.

Vitamin A is resorbed in the duodenum. Retinol esters are first hydrolysed, free retinol then enters enterocytes where it is oxidised to retinal. Retinal is either reverse reduced to retinol (most of it) or oxidised to retinoic acid. Retinol is re-esterified by fatty acids (mainly palmitic acid) and built in chylomicrons which enter the lymph and eventually the blood. Retinol esters with chilomicron remnants enter the liver where they are stored (retinyl palmitate, 80 – 90 % of total vitamin A amount in the body).

If needed, retinol is released from the liver and transported by blood to extrahepatic tissues in a bond to retinol--binding protein (RBP, MW 22,000) which then binds to prealbumin. Retinoic acid binds to albumin and carotenoids circulate in a bond to LDLs and HDLs. The Retinol-RBP complex uptake by tissues is enabled by specific receptors on the cell surface.

11-cis-retinal and retinoic acid are active forms of vitamin A in cells. 11-cis-retinal is essential for reproduction and sight; deficiency causes night blindness. The biologically active form of vitamin A in the retina is all-trans-retinol; as soon as it enters the cell it is esterified and 11-cis-retinol is eventually released by hydrolysis. It is oxidised by alcohol dehydrogenase to 11-cis-retinal, which becomes part of rhodopsin pigment together with the opsin protein (the gene mutation for rhodopsin leads to retinitis pigmentosa). After photon absorption, isomerisation of 11-cis-retinal to all-trans-retinal takes place, causing it to be released from the bond to opsin. All-trans-retinal is reduced to all-trans--retinol and the whole cycle starts again. Retinoic acid is responsible for the other effects of vitamin A, which include the induction of growth and differentiation of epithelial cells, bone growth, is important for lipoprotein and lysosome integrity, and is essential for the synthesis of steroid and thyroid hormone as well as calcitriol. Retionic acid is produced from retinol by oxidation in target tissue cells. It binds to specific nuclear receptors, and the activated complex interacts with chromatin and activates the transcription of specific genes such as the gene for keratin synthesis in epithelial cells.

Figure: Resorption, transportation and conversions of vitamin A (adapted from Lippincott's illustrated reviews Biochemistry, 5th ed., p. 383, cannot insert to GD)

Symptoms of vitamin A deficiency include night blindness, xerophthalmia, keratinisation of epithelial cells, xeroderma, hyperkeratosis or ichthyosis, dental caries, loss of appetite. Deficiency leads to a predisposition for immunity disorders, infections, particularly those of the digestive and respiratory system, and serious deficiency leads to corneal keratinisation or ulceration and blindness.

The total vitamin A level in the serum may decrease until the liver reserve is depleted. This occurs in diseases associated with fat-soluble vitamin deficiency (malabsorption of lipids, cystic fibrosis, inflammatory bowel diseases, defective apoB48, etc.), and severe hypoproteinemia as a result of RBP and prealbumin deficiency.

Retinol and its precursors (esters, carotenes) are used to treat deficiency. Retinoic acid is used in dermatology to treat acne (13-cis-retinoic acid and isotretinoin orally; all-trans-retinoic acid and tretinoin locally) and psoriasis (tretinoin), and in oncology to treat promyelocytic leukaemia (tretinoin).

Excessive intake (100 times the RDA for adults) of vitamin A is toxic and teratogenic, whereas increased intake of carotenoids is not. Pregnancy should be avoided while being treated with retinoic acid; 3 mg of retinol is considered a risk dose, especially if the woman takes it in the first 10 weeks of pregnancy. In this regard every woman planning to become pregnant should be advised to avoid using vitamin preparations and foodstuffs with a high content of vitamin A.

Early manifestations of hypervitaminosis include dryness and itching of the skin followed by alopecia, hepatomegaly or even cirrhosis, neurological disorders (increased intracranial pressure, cephalalgia, diplopia). Prolonged deficiency also leads to an increase in the HDL/LDL ratio and bone disorders. Teratogenic effects include abortions and vitamin A embryopathy (congenital developmental defects of the heart, ears and nose, mandibular hypoplasia, cleft palate, hydrocephalus).

Laboratory tests allow serum vitamin A concentration measurement, however this test does not provide any information about the vitamin reserve in tissues. Liquid chromatography is the method used for the tests.

11.2.4.2. Vitamin D

The term of vitamin D refers to sterols ercalciol (ergocalciferol, vitamin D_2) and calciol (cholecalciferol, vitamin D_3), which, in themselves, are not biologically active. Their conversion in the body produces the active hormone 1,25-dihydroxycholecalciferol (calcitriol).

Ercalciol originates from vegetables and is taken solely in food. Calciol is either taken from animal foodstuffs, or is

produced from 7-dehydrocholesterol by skin exposure to UV radiation. 15 minutes of everyday exposure of face and arms to sunlight (without sunscreens) is enough to reach sufficient plasma levels. In old age, synthesis of 7-dehydrocholesterol and its conversion to calciol (cholecalciferol) decreases. Main food sources include oily fish, egg yolk, liver, cheese, butter, fortified margarines and other fats.

Lipophilic vitamin D absorbs from food in the proximal small intestine. In the enterocyte it becomes part of chylomicrons and is eventually transported in the blood to the liver. In the liver it is hydroxylated to calcidiol (main reserve form of vitamin D in the body), which is further hydroxylated in the kidneys (and placenta) to biologically active calcitriol. Calcidiol and other metabolites are subject to enterohepatic circulation. Vitamin D is transported in the plasma as bound to alpha1-globulin and the DBP (vitamin D-binding protein, transcalciferin), while also binds to albumin and lipoproteins with a lower affinity. The DBP has a high affinity to calcidiol and its inactive form, 24,25-dihydroxyD_3, but it poorly binds calcitriol, which makes this hormone readily biologically available. Adipose tissue presents a major reserve of calciol; more than 50 % of this vitamin is stored here under physiological conditions.

Together with parathormone (PTH), and calcitonin, calcitriol plays a major role in calcium and phosphate management. It binds to nuclear receptors in target cells (intestine, bones, kidneys, placenta, breast), and directly affects transcription. In the intestine it increases Ca and phosphate absorption by its easier access to the enterocyte through the cytosolic calmodulin conformational change, through its easier transport in the enterocyte through the calbindin synthesis induction, and by easier transport of Ca from the enterocyte through Ca^{++}-ATPase synthesis induction. In the bones it accelerates osteoclast maturation, while also affects mineralisation by osteoblast stimulation. In the kidneys it increases Ca and phosphate resorption while also increases their transport in the placenta and mammary glands. Recent research has also shown that calcitriol has immunomodulatory, antiproliferative and pro-differentiation effects.

Synthesis and the calcitriol level are regulated by many mechanisms. The main site of regulation is the last step of calcitriol synthesis – reaction catalysed by 1-alpha-hydroxylase. Here the main activator is parathormone, while Ca (directly and indirectly via ↓ PTH), phosphates and the reaction product (calcitriol) have an inhibitory effect. Enzyme activity is also modulated by the acid-base balance and other hormone concentrations (calcitonin, insulin, STH--IGF-I, prolactin, sex hormones). Another mechanism of regulating calcitriol homeostasis is activation of synthesis by 24,25-dihydroxyD_3 calciol and inhibition of PTH. 24,25-dihydroxyD_3 is considered to be an inactive degradation product; its function is not quite clear.

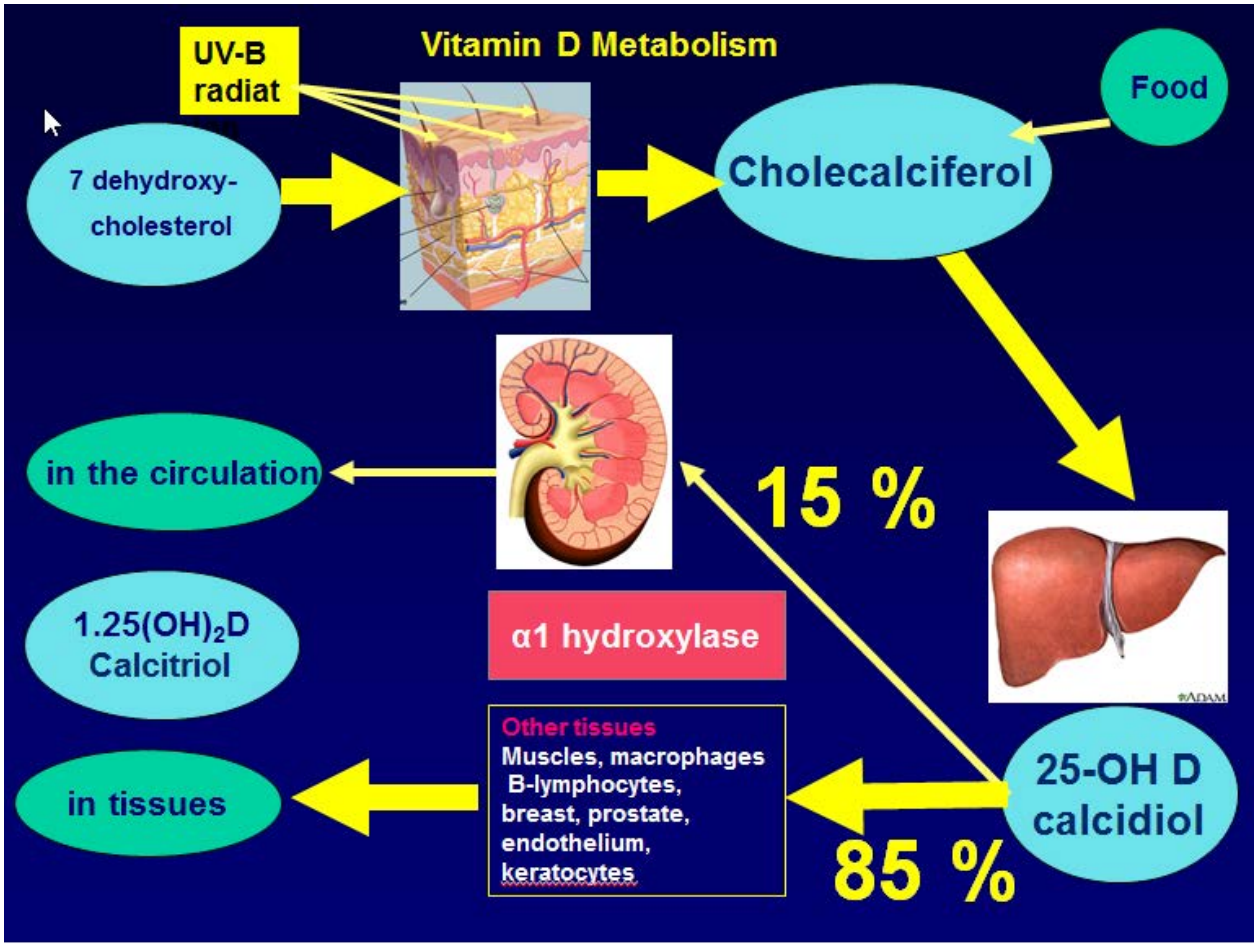


Figure 11.1. Regulation of calcitriol level and synthesis (is it next to Ca, bones, hormones?)

Deficiency occurs due to insufficient intake from food, disorders of absorption, cholestasis (lack of bile acids), insufficient exposure to sunlight (season of the year, indoor lifestyle, sunscreens), reduced hydroxylation (hepatic and renal diseases, hypoparathyroidism), nephrotic syndrome. Demineralisation of the bones leads to the development of rickets in children, and osteomalacia in adults. Vitamin D is used for treatment, and calcitriol for resistant cases. In chronic renal failure osteodystrophy develops as a result of reduced 1-alpha-hydroxylation. Patients are given calcitriol and the phosphate level should be decreased to prevent formation of calcium-phosphate stones. A long-term low vitamin D saturation is currently estimated to exist in at least 1/3 of the healthy European population. It is not associated with typical symptoms of deficiency, but increases the risk of autoimmune (thyroiditis, rheumatoid arthritis) and lifestyle diseases (tumours, depressions, etc.).

Symptoms of toxicity such as loss of appetite, nausea, diarrhoea and vomiting, thirst, itching or stupor may occur as a result of prolonged very high doses of vitamin D. Hypercalcaemia develops and leads to calcifications of the vessel walls and kidneys.

Plasma concentration of calcidiol and calcitriol is currently determined by immunochemical techniques or HPLC. Calcidiol (25-hydroxycholecalciferol) is proving to be the most adequate method for measuring total reserve of vitamin D in the body.

Note on the intake and RDA: 1 µg = 40 IU; RDA = 10 µg = 400 IU

Vitamin D intake	µg	IU	% RDA
10 – 15-minute whole body exposure to sun	250 – 500	10 000 – 20 000	2 500 – 5 000
Salmon /100 g	12	480	120
Sardines /100 g	5	200	50
Butter /100 g	1	40	10

Table 11.2. Vitamin D – sources

11.2.4.3. Vitamin E

Vitamin E refers to a group of eight tocopherols and tocotrienols, of which alpha-tocopherol has the highest biological activity. It is also the most common in food, with cereal germs, nuts, poppy seed and egg yolk being the richest sources.

Lipopholic vitamin E resorbs from food with fats (the resorption rate is roughly 35%); in enterocytes it is built in chylomicrons and goes to the lymph and blood. It enters the cells of tissues with active lipoprotein lipase; the rest with remnants goes to the liver. Here it is built in VLDL and re-enters the circulation. The target organs are all tissues; higher content can be found in cells with high pO^2 (erythrocytes, lungs); it is stored in the adipose tissue. 70 – 80% is excreted through the liver, the rest in urine as tocopheric acid and glucurono-gamma-lactones.

Vitamin E has antioxidant properties. The body uses it principally to protect poly-unsaturated fatty acids in plasma membranes and lipoproteins against the action of free radicals, also to prevent oxidation of other important compounds. It is important for membrane integrity, protecting erythrocytes against haemolysis, inhibiting mutagens in the digestive tract, and playing a role in cell signalling pathways.

Vitamin E deficiency may occur as a rare complication of severe and long-term steatorrhoea and poor parenteral nutrition or malnutrition. Children in particular may develop haemolytical anaemia as a result of corrupted erythrocyte membrane stability. The life of erythrocytes is shorter and thrombocyte aggregability is greater. Changes in peripheral nerve function, elevated excretion of creatinine and increased CK activity occur as a result of damaged membrane and skeletal muscle cell disintegration. Long-term deficiency is accompanied by myopathy, necrosis of muscles, hypo- or even areflexia, spinocerebellar ataxia and retinopathy.

Toxic overdosing symptoms include gastrointestinal problems, tiredness, headaches, muscle weakness. Overdosing in pregnant women (20 – 70 times RDA) may cause damage to the foetus (congenital heart defects, omphalocele, non-standard birth weight).

Liquid chromatography with UV or fluorescent detection is commonly used to determine plasma or serum vitamin E levels; alternatively gas chromatography with mass spectrometry can be used.

11.2.4.4. Vitamin K

Vitamin K refers to derivatives of naphthoquinone: phylloquinone (phytomenadione, K_1) and menaquinones (K_2). The source of phylloquinone is food such as leaf vegetables, broccoli and oatflakes; menaquinone is synthesised by intestinal bacteria (synthesis normally covers demand for vitamin K) and is also contained in fermented foodstuffs (yoghurt, cheese) and the liver of ruminants. Synthetic derivatives menadiol and menadione are soluble in water and have the same effects as the natural vitamin. Resorption takes place in the small and large intestine, 10 – 80% of ingested vitamin K, which is transported by the lymph followed by the blood to target tissues. K_2 is stored in the liver; the reserve is limited, which requires permanent substitution.

Vitamin K is needed for post-translational conversion of glutamate residues in proteins into gamma-carboxyglutamate, which can subsequently bind Ca^{2+}. This conversion takes place in blood coagulation factors II, VII, IX and X, protein C and S and osteocalcin, a protein produced by osteoblasts, which is required for new bone formation and remodelling.

Vitamin K deficiency most commonly occurs as a result of anticoagulant treatment (coumarins = warfarin – vitamin K cycle inhibition in the endoplasmic reticulum of the liver), and the typical manifestation is proneness to bleeding. Less common causes of deficiency include malabsorption of lipids (obstructive icterus) and lack of gut flora resulting from antibiotic therapy. Deficiency and required substitution are physiological in newborns as vitamin K does not pass the placental barrier and the newborn gut is sterile.

Vitamin K_1 is used as an antidote for oral anticoagulant overdosing, in addition to blood plasma transfusion. As its effect persists several days to weeks, short-term anticoagulant treatment with heparin is required when bleeding stops.

Liquid chromatography is commonly used to determine vitamin K_1 concentration in plasma.

11.2.5. Case Reports

A patient with icterus and pain in the right epigastrium was hospitalised. He is pale, feverish and has chills.

Laboratory findings: ALP 8.5 micro kat/l, AST 0.64 micro kat/l, total bilirubin 418 micro mol/l, conjugated bilirubin 311 micro mol/l, positive urinary bilirubin, negative urobilinogen, prolonged prothrombin time, total cholesterol 7.5 mmol/l.

What is the most likely cause of icterus in this patient? Why is prothrombin time prolonged?

Answers: Icterus is caused by an extrahepatic obstruction. Prothrombin time prolongation is caused by lacking vitamin K required to activate prothrombin. Vitamin K is fat-soluble and malabsorption of lipids occurs in cholestasis.

Prothrombin time improved after administering vitamin K as an injection.

CHAPTER 12

12. Thyroid Gland

12.1. Thyroid Gland – Preparation

Author: MUDr. Richard Pikner, Ph.D.

Reviewers: MUDr. Daniel Rajdl, Ph.D.

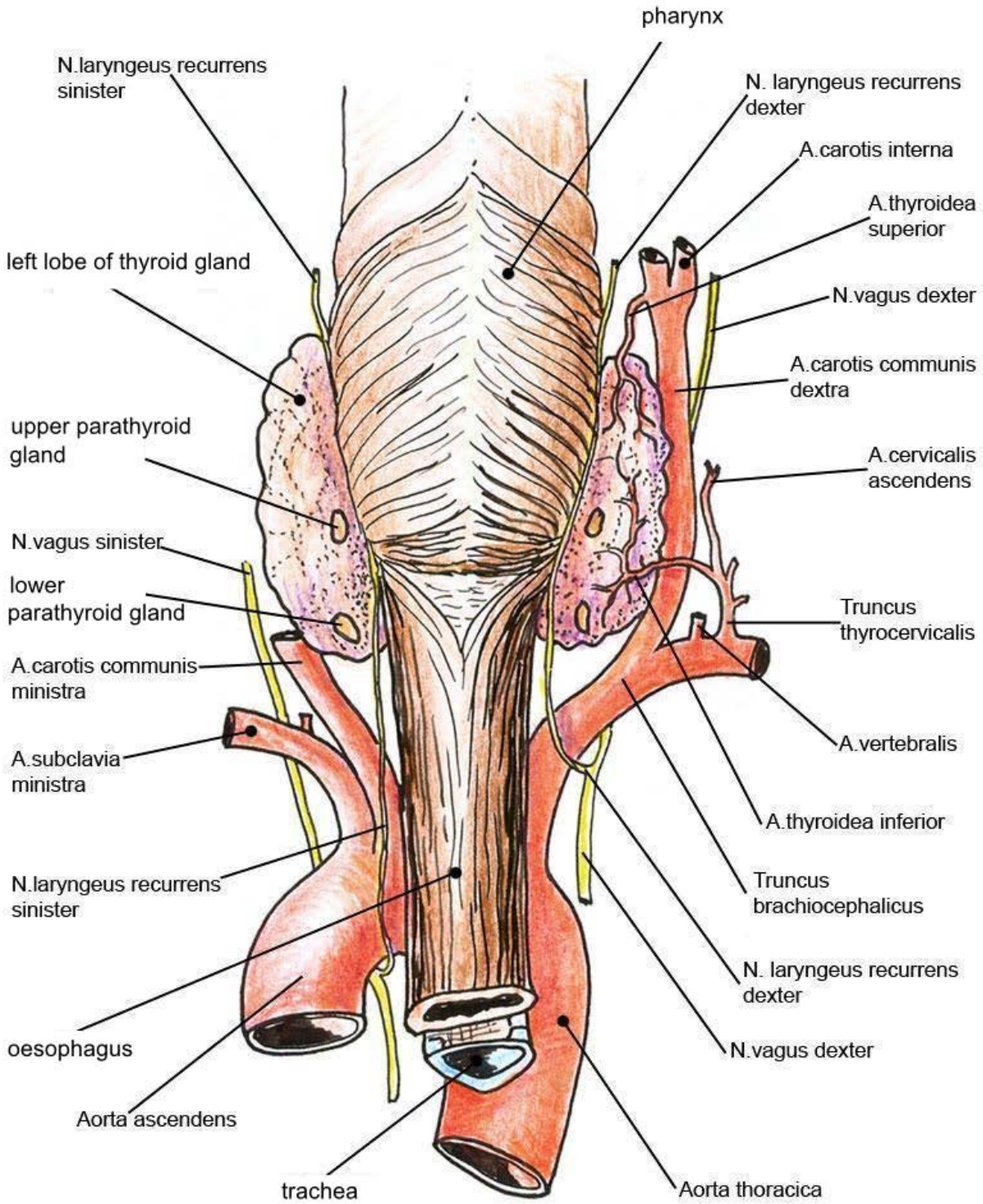
12.1.1. Introduction

The thyroid gland is an organ weighing roughly20 g and with a volume of 18 – 22 ml in adults. Thyroid disorder affects the intermediary metabolism of all cells within the human body. Thyroid dysfunction with obesity and diabetes present a triad of the most common endocrinopathies afflicting the population. Laboratory testing is the basic pillar in diagnosing, treating and monitoring thyroid disorders, because a change in parameters often comes before developing clinical symptoms.

12.1.2. Notes to Anatomy and Histology

The thyroid gland is formed as a diverticulum in the floor of the pharynx, descending to the base of the tongue and down in front of the trachea (this path is called the thyroglossal duct, *ductus thyreoglossus*). Ultimately the thyroid gland is located just below the cricoid cartilage and above the jugular notch at the level of C5-Th1 vertebra. It consists of the right and left lobes connected via the thyroid isthmus. The size of the lobes depends on the iodine supply, age and gender. When the iodine supply is sufficient, the normal thyroid size is under 18 ml in adult women and under 22 ml in men (measured by sonography).

The thyroid is richly supplied with blood from both sides of *a. thyroidea superior* and *a. thyroidea inferior*, supplying the two thyroid lobes, and *a. thyroidea ima* supplying the isthmus. The blood is drained via the *v. thyroidea superior, media* and *inferior*. Close to the thyroid anatomy and along both lobe edges pass *n. laryngeus recurrens dexter* and *sinister*, which, when damaged, often present a complication of thyroid surgery. On the rear surface of the lobes, lying on the posterior side of the lobes at their upper and central part, on each side there are two parathyroid glands. These are responsible for producing parathormone. (Figure 11p.1)



The relation of glandula thyreoidea to the surrounding structures (from behind)

Figure 12.1. Thyroid gland and surrounding structures – from behind

The follicle is the basic functional unit of the thyroid gland. These almost spherical formations are 300 µm in diameter. The wall consists of follicle cells, whose base lies on the interfollicular space with a rich vascular and lymphatic supply. The inner, apical part is turned to the follicle lumen and contains a colloid with thyroglobulin as the main component. Apart from the rich vascular and lymphatic drainage system, the interfollicular space contains parafollicular cells (C-cells) producing calcitonin which travelled to the thyroid during embryogenesis. (Figs. 2 and 3)

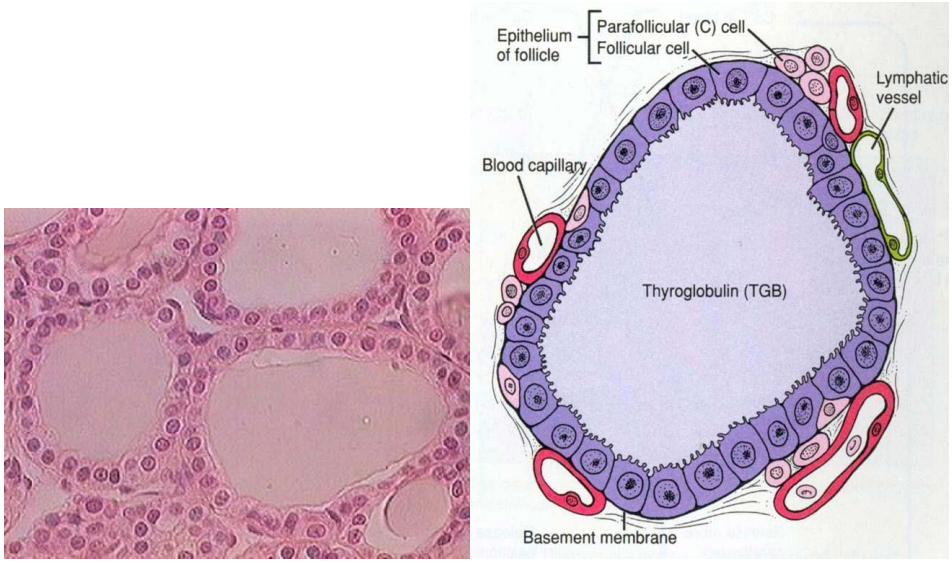


Figure 12.2. Thyroidal follicles

12.1.3. Metabolism of Thyroid Hormones

The thyroid gland is the site where thyroid hormones T_4 (tetraiodothyronine, thyroxine) and T_3 (triiodothyronine) are produced. T_4 is a major thyroid product, while T_3 is a cellularly active hormone. The human body thus has to transport the produced T_4 and subsequently convert it to active T_3, followed by deactivation and degradation of T_3. Tissue deiodinase enzymes are responsible for converting T_4 to T_3 and deactivating T_4 and T_3.

12.1.3.1. Metabolism of Iodine

Iodine and selenium are essential elements for thyroid hormone metabolism. A sufficient supply of iodine is required to produce T_4 (T_3) in the thyroid. Iodine deficiency leads to insufficient production of thyroid hormones with clinical symptoms from goitre to severe CNS damage (cretinism) in newborns. Iodine intake is geographically dependent on its content in natural resources. In terms of natural content, the Czech Republic ranks to iodine-deficient regions, meaning iodisation of food (salt) and water is necessary. The normal daily iodine requirement of an adult is about 80 µg/day, and up to 200 µg/day for pregnant women. Iodine is received by re-using the T_4 and T_3 hormones, and uptake of this from the blood. The main loss of iodine following T_4 and T_3 degradation takes place in the kidneys and in stools, with the loss in stools being relatively constant, about 16 µg/day. Iodineuria measurement provides the basic information about the iodine supply. Therefore, the optimum daily iodine intake is about 100 – 150 µg/day, and 150-200 µg/day for pregnant women.

12.1.3.2. Thyroid Hormone Synthesis

Iodine uptake on the basal membrane of follicle cells is an active process via the ATP dependent sodium-iodine symporter (NIS). Its activity is increased by the TSH and blocked by competitive inhibitors such as thiocyanate (SCN^-) and perchlorate (ClO^{4-}) anions. NIS is also physiologically present in other organs such as the mammary gland, salivary gland, placenta and the CNS. NIS transports iodine from the mammary gland to breast milk.

Iodine is transported towards the apical (luminal) side following on from uptake on the basal membrane. Here it is oxidized with thyroid peroxidase (TPO) to reactive free iodine, which immediately binds to tyrosines in the thyroglobulin molecule. This process takes place in the colloid close to the apical membrane. Thyroglobulin rich in iodinated tyrosines (monoiodotyrosine, MIT, and diiodotyrosine, DIT) is thus formed. Thanks to its spatial arrangement (proximity of MIT and DIT on thyroglobulin), thyroglobulin is the place where 2 tyrosine molecules combine to form triiodothyronine (T_3) and tetraiodothyronine – thyroxine (T_4), again in the presence of TPO (Figs. 4 and 5).

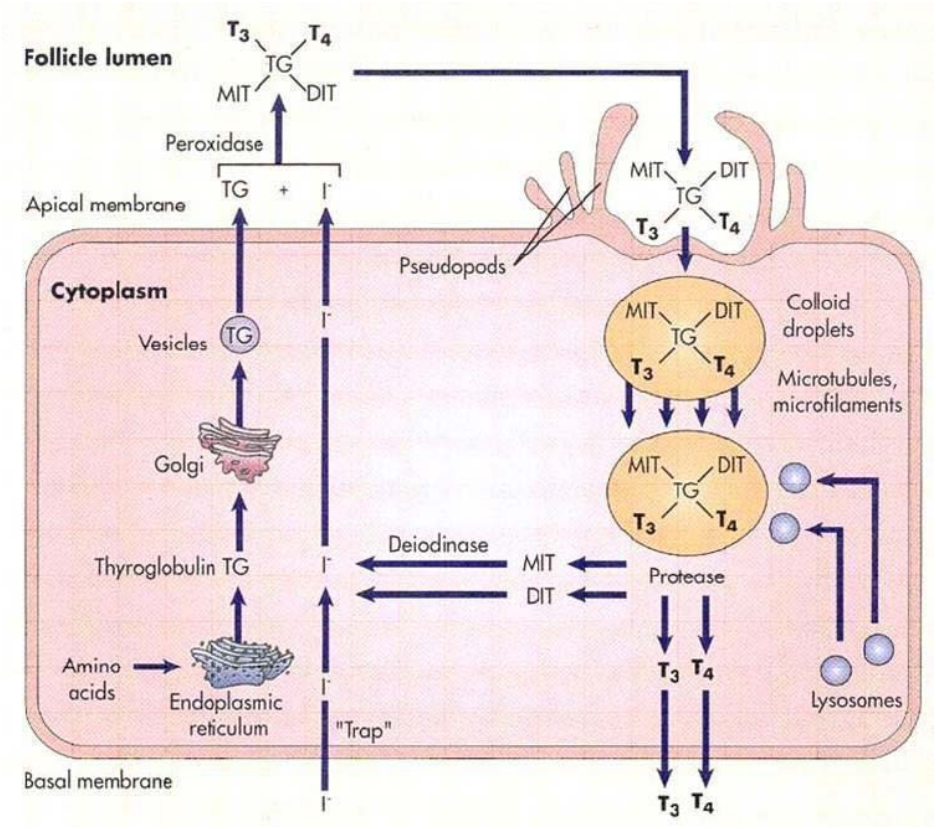


Figure 12.3. T-hormones synthesis

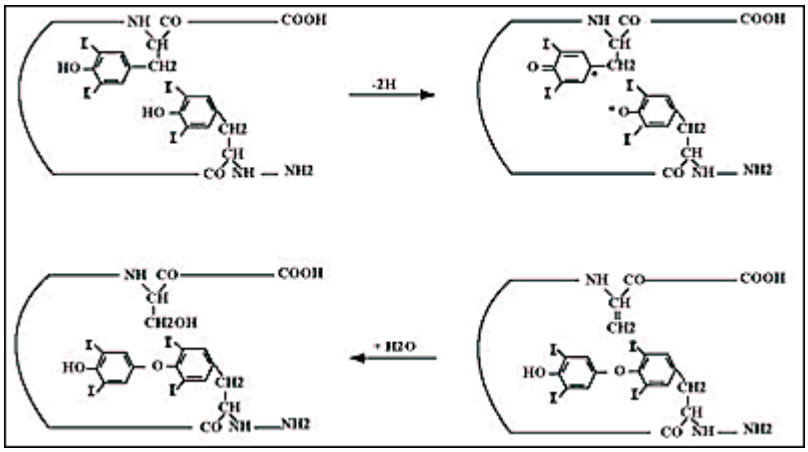


Figure 12.4. T-hormones synthesis - chemical formulae

Thyroglobulin containing T_4, T_3 and DIT and MIT is endocyted back to thyrocyte. Tg is hydrolysed through the action of phagolysosomes; Tg releases T_3 and T_4 to secretory granules, from where T_4 and T_3 are subsequently released into the circulation. MIT and DIT cleave away an iodide, which just like Tg is recycled inside the cell for further synthesis. T_4 and T_3 secretion into the circulation is regulated by the TSH. The thyroid gland produces more T_4 (70 %) than T_3. T_3 is more metabolically active (3-8 times), so T_4 is rather a prohormone. After being transported in the blood to peripheral tissues, T_4 is converted to active T_3.

12.1.3.3. Transport of T_4 and T_3 to Peripheral Tissue:

T_4 and T_3 are transported into the plasma as bound to transport proteins. Only a little part in the plasma is free (metabolically active), and is measured as free T_4 (0.02 – 0.05 % of total T_4 – TT_4) and free T_3 (0.01 – 0.3 %). The main transport proteins are thyroxine-binding globulin (TBG), prealbumin (transthyretin) and albumin. T_4 and T_3 bound to transport proteins present a thyroid hormone pool for tissues, ensure stable levels of free T-hormones, and protect T_4 and T_3 from getting lost in urine. On the other hand, free FT_4 and FT_3 levels practically do not change with changing binding protein concentrations, except in extreme cases. For example, a 200% increase in TBG occurs during pregnancy, while free FT_4 and FT_3 remain unchanged. In cases of malnutrition the TBG level drops while FT_4 and FT_3 levels do

not change.

12.1.3.4. Conversion and Elimination of T_4 and T_3:

Around1/3 of T_4 produced by the thyroid converts to T_3, another 1/3 to rT_3 (metabolically inactive), and the rest is degraded by glucuronidation and sulfation in the liver. The conversion of T_4 to active T_3 or inactive rT_3 is catalysed by tissue-specific deiodinase enzymes (selenium-dependent enzymes). These also ensure rT_3 and T_3 degradation to T_2. There are 3 types of deiodinases in the body, which differ in their localisation in organs, specificity to the substrate and sensitivity to propylthiouracyl (PTU). They require selenium to function; selenium subsequently removes iodide in the reaction.

Type I deiodinase. D1 – present in the liver, kidneys, thyroid; primarily degrades rT_3 to T_2, and T_4 to T_3, is the source of major T_3 pool in the circulation, and contributes to maintaining iodine for further thyroid hormone synthesis. This means that D1 activity increases in hyperthyroidism, causing the condition to worsen further; PTU is inhibited.

Type II deiodinase. D_2 – present in the brain, pituitary gland, heart, brown adipose tissue; locally produces T_3 from T_4, is only slightly sensitive or not at all to PTU.

Type III deiodinase. D_3 – brain, placenta, skin; responsible for T_4 inactivation to rT_3 and T_2, regulates the passage of hormones from the placenta to foetal circulation, only slightly sensitive or not at all to PTU.

12.1.3.5. The Physiological Effects of Thyroid Hormones:

Thyroid hormones have an intracellular receptor (TR). T_3 shows a greater affinity and biological activity. The receptors are located in the cell nucleus. A hormone bound to the receptor forms the thyroid response element (TRE), which binds to DNA and induces the expression of specific genes, or suppression of other genes. The TR is a part of the group of nuclear receptors including receptors for: oestrogens, progesterone, androgens, gluco- and mineralocorticoids, and vitamin D. TRs exist in at least 4 isoforms: $TR\alpha_1$, $TR\alpha_2$, $TR\beta_1$, and $TR\beta_2$. The DNA-binding domain contains zinc. The main metabolic effect of T_3 is regulating the basal metabolic rate. In addition to lipolysis, glycogenolysis and gluconeogenesis, T_3 increases cell sensitivity to catecholamines, while increasing O_2 consumption and heat production.

12.1.4. Regulation of Thyroid Hormones

12.1.4.1. Hypothalamus-Pituitary-Thyroid-Peripheral Tissue Axis

The top and first degree of regulation of thyroid hormone synthesis is the CNS. Therefore, the production of neurotransmitters has both a stimulating and inhibiting effect on the production and release of the TRH.

Regulatory mechanisms at the hypothalamic level:

- **Stimulating factors** of TRH production: Stress, cold, leptin, adrenergic receptor pathways;
- **Inhibiting factors:** Heat, T_3, dopaminergic and opioid receptor pathways.

Regulation at the pituitary level primarily involves the TRH, which stimulates TSH release in the pituitary, while somatostatin inhibits TSH secretion. Major feedback inhibition takes place via T_3 (T_4). In addition to these mechanisms, many other mediators, signalling molecules and hormones are involved in regulation.

Regulatory mechanisms at the pituitary level:

- **TSH release inhibition:** Androgens, T_3 (T_4), cortisol, inflammatory cytokinins (IL6, TNF α, IL1β), malnutrition, cholecystokinin (CCK), gastrin-releasing peptide (GRP), neuropeptide-Y (NPY);
- **TSH release stimulation:** Leptin, growth hormone (GH), oestrogens, arginine vasopressin (AVP), glucagon-like peptide-1 (GLP-1), galanin.

It should be kept in mind that the biological effect of TSH itself depends on post-translational modification, i.e. glycation. This fact may be manifested in some forms of central hypothyroidism, when the measured TSH does not match the clinical condition – TSH with a lower biological activity is produced.

TSH and feedback via T_3 (T_4) are the major regulatory mechanisms at the thyroid level. Thyroid hormone synthesis is also inhibited by excess iodine, namely by blocking hydrogen peroxide production with subsequent iodine oxidation

and bonding to thyroglobulin (referred to as the Wolff-Chaikoff effect). Thyroid hormone synthesis is also blocked by agents used to treat hypothyroidism, thioamides (propylthiouracil and methimazole).

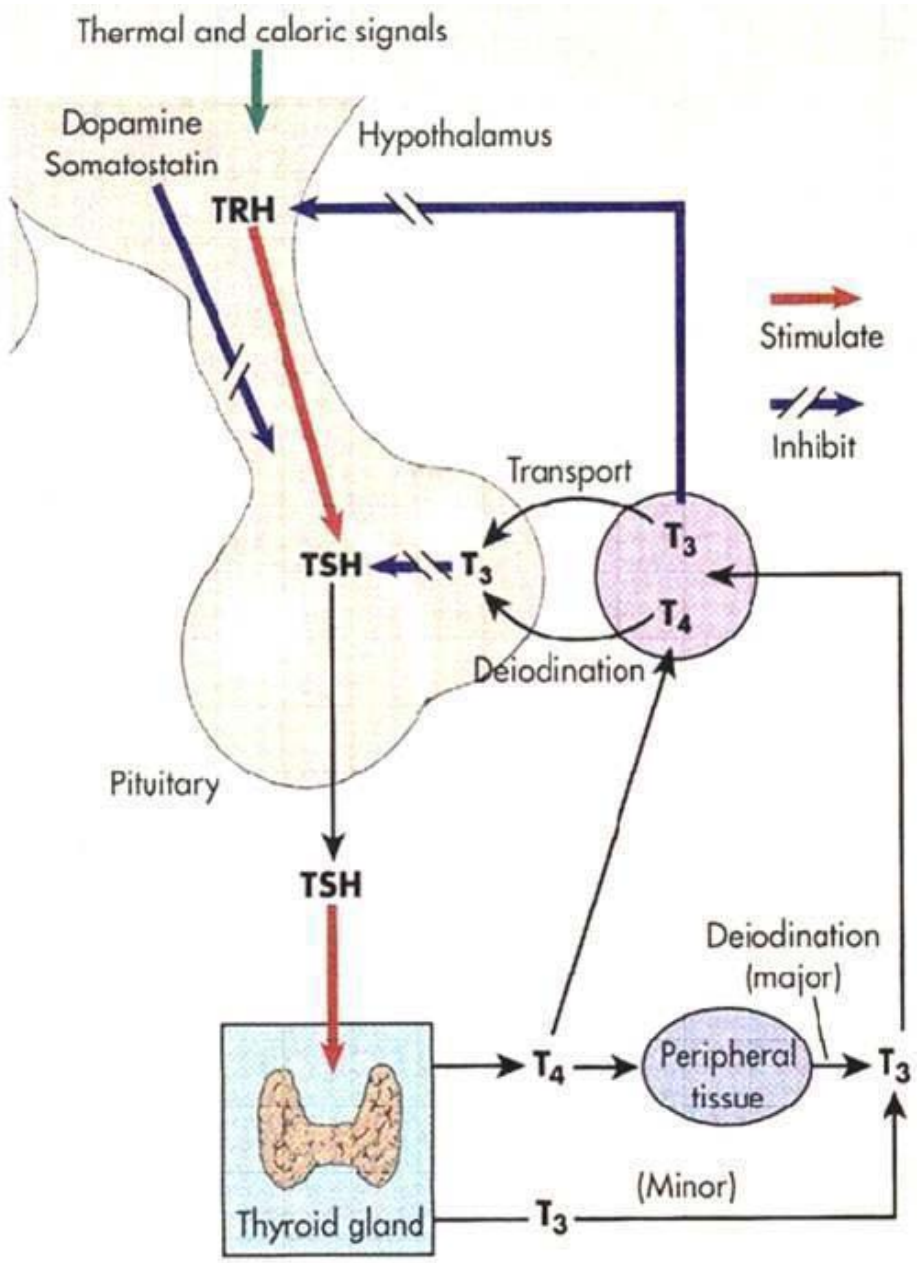


Figure 12.5. T-hormones regulation

There is a logarithmic correlation between FT_4 and TSH (see Fig. 7); TSH subsequently reacts 100 times more sensitively to changes in FT_4 levels, so TSH is also used as the most sensitive parameter to detect thyroid dysfunctions.

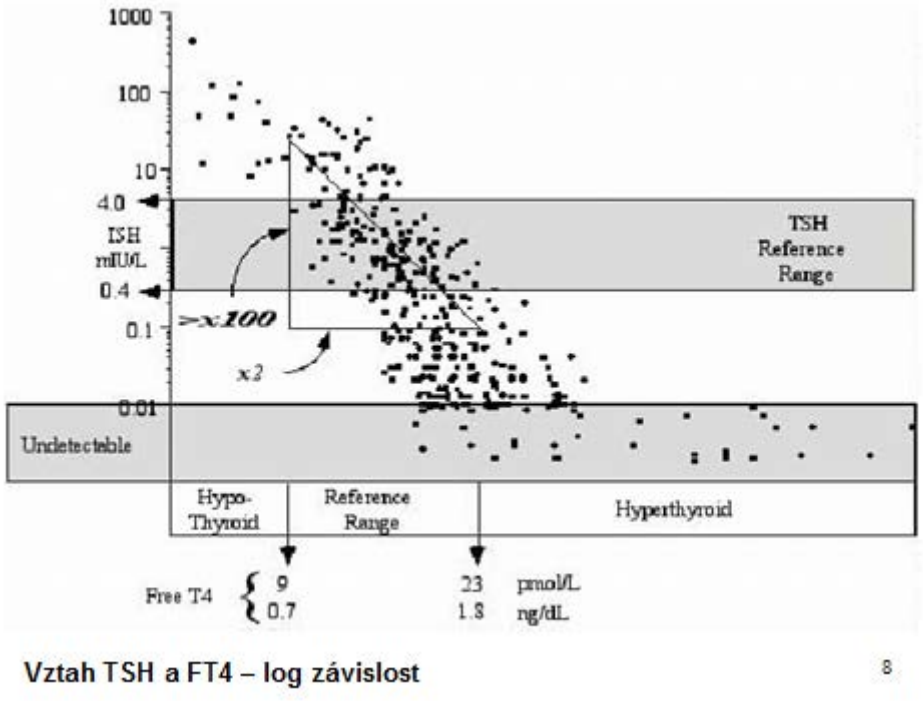


Figure 12.6. Connection TSH and FT_4 – log dependence

12.1.5. Description of Laboratory Parameters

12.1.5.1. Thyroid-Stimulating Hormone (TSH, Thyrotropin)

TSH is a glycoprotein hormone with a molecular weight of 28-30 kDA, produced by thyrotrope cells (basophilic cells) in the anterior pituitary gland. TSH ranks among the group of glycoprotein hormones (FSH, LH, hCG) with the same α-subunit and differing in the β-subunit. The α-subunit is coded by the common gene on chromosome 6, and consists of 4 exons and 3 introns. The β-subunit is coded by the gene on chromosome 1, and consists of 3 exons and 2 introns. Thus the protein itself consists of an α-chain of 92 amino acids, and a β-chain of 112 amino acids. This β-chain is in 38% homology with the hCG molecule β-chain. This grows in importance during pregnancy or in the case of hydatidiform mole (*mola hydatidosa*), when excess hCG and partial homology lead to the stimulation of TSH receptors in the thyroid and thereby to hyperthyroidism. Carbohydrates take 15-20% of the relative molecular weight of the TSH, and contain 3 oligosaccharide chains (2 of them bound to the α-subunit and 1 to the β-subunit). The process of glycation is very important for biological activity of the TSH. Some forms of central hypothyroidism are caused by reduced biological activity due to impaired TSH glycation. TSH release into the blood circulation is controlled by the hypothalamus through the TRH and somatostatin.

The TSH binds to its specific TSH receptor, which stems from the group of G-protein activating receptors.

The TSH receptor consists of:

- Extracellular domain: binds TSH and antibodies against TSH, 40% match with receptors for the FSH and LH, 398 amino acids;
- Transmembrane domain – 7 transmembrane domains, 70% match with the FSH and LH receptors, 346 amino acids, G-protein activation on the third loop.

Assuming an intact hypothalamic-pituitary axis, the TSH has a logarithmic relation with free T_4, which means that even a little change in FT_4 results in a 10 times greater change in the TSH.

TSH function:

- Stimulates various functions of thyrocytes: Uptake and organification of iodine, production of Tg and production and secretion of thyroid hormones;
- Supports thyrocyte growth – hypertrophy and hyperplasia;
- Inhibits thyrocyte apoptosis;

- Extrathyroidal functions: Affects lymphocytes, adipocytes, adrenal and testicular cells through its receptors.

12.1.5.2. Thyroid Hormones T_4 and T_3

Thyroid hormones are thyroxin (T_4), triiodothyronine (T_3) and reverse triiodothyronine (rT_3), diiodothyronines (T_2) and monoiodothyronines (T1). Most hormones circulating in the blood are bound to transport proteins, and thus biologically inactive. Only a small part of hormones present in free form in the blood is biologically active.

Overall thyroid hormone concentrations, total T_3 (TT_3) and total T_4 (TT_4), have a diagnostic value only if the binding capacity of transport proteins is not changed (use of drugs, serious systemic disease, etc.).

Conversely, free fractions of thyroxin (FT_4) and triiodothyronine (FT_3) provide information about current functioning of the thyroid gland.

12.1.5.3. Thyroxin-Binding Globulin (TBG)

TBG is the main transport protein of T-hormones into plasma, has a relative molecular weight of 5400 Da, is produced in the liver and consists of 395 amino acids. The TBG gene is localised on the X chromosome and consists of 6 exons. Indications for testing are very rare; these days it is primarily due to a congenital TBG deficiency and discrepancy of TT_4 and FT_4 results.

12.1.5.4. Thyroglobulin (Tg)

Human Tg is a glycoprotein containing iodine and has a relative molecular weight of 660 kDa. Tg consists of 2 identical polypeptide chains. Carbohydrates form less than 10% of its weight. Around 30% of iodine bound to Tg has the form of T_3 and T_4; the rest is inactive precursors monoiodotyrosine (MIT) and diiodotyrosine (DIT). Concentration in the follicular lumen reaches 200 – 300 g/l. The Tg gene is located on chromosome 8 and contains 48 exons. Tg is synthesized in follicular cells of the thyroid gland in normal conditions. Its synthesis is controlled by the TSH. The plasma half-life is 3 – 4 days.

1 g of normal thyroid tissue produces 1 – 2 ug/l of Tg at a normal TSH concentration, while its concentration correlates to the thyroid volume, thyroid activity (increased in the case of hyperfunction), it also increases with active inflammatory or tumour processes.

12.1.5.5. Autoantibodies

Autoimmune thyroid diseases are the most common cause of thyroid disorders. It is 4 times more frequent in women than the prevalence in men in regions without iodine deficiency. The test for autoantibodies is important for differential diagnosis and prognosis.

Methods of anti-TPO, anti-Tg and anti-TSH testing have highly heterogeneous results and different sensitivity and specificity, which is due to the fact that:

- Antibodies recognize different interindividual epitopes, have different affinity, and mostly comprise a mixture of heterogeneous antibodies;
- Manufacturers use different standards, different detection methods (competitive and non-competitive), and antigens are prepared using different methods.

These two basic postulates lead to different results between methods, which differ in the order of tens to hundreds of percent.

TSH Receptor Antibodies – Anti-TSH (aTSH, TRAb)

TSH receptor antibodies may be divided into 2 groups depending on their function:

- **TRAbS** (Thyroid receptor stimulating antibodies) – The bond to the N-terminal region leads to imitation of the TSH function, stimulation of production and release of thyroid hormones and cell growth. They are an aetiological factor of Graves' thyrotoxicosis;
- **TRAbB** (Thyroid receptor blocking antibodies) – The bond to the C-terminal region leads to the inhibition

of TSH function and reduced release of thyroid hormones.

These two antibodies are undistinguishable with regular immunoassays; they can be distinguished using biological analyses. One individual may possess both the stimulating and inhibiting antibodies (exceptionally), either simultaneously or consecutively, which greatly affects the clinical presentation. They are also responsible for clinical manifestation of endocrine ophthalmopathy in euthyroid individuals. The antibodies cross the placenta and are risk factors of foetal hyperthyroidism.

Thyroid Peroxidase Antibodies – Anti-TPO (aTPO, TPOAb)

Thyroid peroxidase is an enzyme located on the apical membrane of thyrocytes, and catalyses hydrogen peroxide formation followed by iodine organification and bond to Tg. It has a relative molecular weight of 100 kDa. The gene for TPO is situated on chromosome 2. In normal cases the gene acts as a code for two proteins (TPO_1 and TPO_2) formed by post-transcriptional modifications of mRNA.

Anti-TPO were originally described as autoantibodies against microsomes; only later was thyroid perodixase detected as a major antigen.

Anti-TPO are IgG-class antibodies with high interindividual variability in subclasses IgG1, IgG_2 and IgG_3 and reactivity. Unlike anti-Tg, they are able to fix the complement. They are the main marker of chronic lymphocytic thyroiditis.

Thyroglobulin Antibodies – Anti-Tg (aTg, TG Ab)

Unlike anti-TPO, they are unable to fix the complement, and have a higher heterogeneity given the large number of epitopes on Tg.

They are a similar marker of autoimmune thyroiditis, but have a higher incidence of positivity than anti-TPO in iodine-deficient regions and patients with modular goitre.

They can interfere with Tg tests (cause false positivity or negativity), and have a higher interindividual variability and variability of reactivity than anti-TPO.

12.1.6. Summary – Key Information

12.1.6.1. Metabolism of Thyroid Hormones

- Iodine is required for thyroid hormone synthesis; the daily requirement is around 100 – 200 µg/day; iodised salt is the source of iodine in the CR;
- Thyrocyte actively take up iodine from the circulation; iodine binds chemically on the apical side to tyrosine amino acid in thyroglobulin;
- Thyroid hormones are synthesized in the colloid, and thyroglobulin is the donator of tyrosine and also a matrix where iodinated tyrosyl amino acids combine to form tetraiodothyronine (T_4) and triiodothyronine (T_3);
- T_4 is the main product of the thyroid;
- T_4 and T_3 are bound in the plasma to transport proteins (TBG, transthyretin, albumin), which maintain a stable level of circulating T_4 and T_3;
- Thyroid gland is the source of only 20 % of T_3 concentration in the circulation; the majority of T_3 occurs as a result of deiodination from T_4 in peripheral tissues;
- Only 0.03 % of T_4 and 0.03 % of T_3 circulate in the free, biologically available form, referred to as free T_4 (FT_4) and free T_3 (FT_3), whose levels are responsible for the biological effect;
- T_4 is converted in peripheral tissues to biologically active T_3, or deactivated to reverse T_3 (rT_3) through the action of specific deiodinase enzymes (dependent on sufficient amounts of selenium – selenoenzymes);
- T_3 and partially T_4 are actively transported to cells, where the main effect of T_3 takes place through nuclear receptors for T_3;
- T_3 increases the basal metabolic rate in the first place (lipolysis, glycogenolysis and gluconeogenesis and also O_2 consumption and heat production);
- T_4 and T_3 degradation takes place as deiodination by deiodinases, and by sulphonidation and glucuronidation.

12.1.6.2. Regulation of Thyroid Hormones

- There are 4 key components of thyroid hormone production regulation (Fig. 4):
  - Hypothalamus (production of the thyrotropin-releasing hormone, TRH);
  - Pituitary gland (production of the thyroid-stimulating hormone, TSH);
  - Thyroid gland (production of T_4, minimal amounts of T_3 - 15 %);
  - Peripheral tissues (production of T_3).
- TSH plays a key role in regulation. TSH is stimulated by TRH and retroactively inhibited by free fractions of FT_4 and FT_3 independently of each other;
- Many local factors affect TRH and TSH secretion at the local level (e.g. somatostatin, dopamine, stress, cold); they are usually temporary and do not affect FT_4 and FT_3 levels;
- TSH has a stimulating effect on thyrocytes, which leads to their hypertrophy and proliferation, and finally to increased production of T_4;
- The TSH level is 100 times more sensitive to individual changes in FT_4 levels; the TSH is the most sensitive marker of changes in potential thyroid disorder;
- T_4 levels depend on TSH production, a sufficient amount of iodine, while overall fraction levels depend on binding protein levels;
- It is assumed that every individual has a genetically defined optimal level of thyroid hormones, so reference limits for FT_4 have a relatively broad range;
- The T_3 level depends on T_4 production, function of deiodinases (sufficient amount of selenium) and metabolic condition (see the low T_3 syndrome);
- Proper thyroid function thus requires normal functioning of the hypothalamic-pituitary axis, sufficient amounts of iodine and selenium, an undamaged thyroid (most commonly autoimmune diseases) and absence of serious co-morbidities.

12.1.6.3. Laboratory Parameters for Thyroid Tests

- TSH:
  - Common beta-subunit with the FSH, LH and hCG; excess levels cause thyroid stimulation (pregnancy, hydatidiform mole);
  - TSH stimulates hormone production in thyrocytes and also has a proliferative effect.
- T_4 and T_3:
  - Their free fractions are biologically active;
  - T_4 concentration reflects its production in the thyroid gland;
  - T_3 concentration reflects the activity of peripheral deiodinases (mainly from the liver), however thyroid hyperfunction also occurs due to isolated T_3 overproduction.
- TBG:
  - Production is stimulated by oestrogens (pregnancy); production disorders are very rare Tg (thyroglobulin);
  - Serum concentration correlates with the size and activity of the thyroid, and with the activity of thyroid diseases (inflammations, tumours).
- Anti-TSH:
  - These are mostly stimulating agents which cause Graves' thyrotoxicosis; they have rarely an inhibitory effect causing hypothyroidism;
  - They are responsible for endocrine orbitopathy in Graves' thyrotoxicosis.
- Anti-TPO, Anti-Tg:
  - Major markers of Hashimoto's thyroiditis – chronic lymphocytic inflammation; Anti-TPO are the major risk marker; Anti-Tg occur sporadically in iodine-deficient patients and patients with a nodular goitre;
  - These tests provide the most considerable differences in resulting concentrations determined using different methods;
  - Anti-Tg present in serum may interfere with the Tg test and lead to falsely positive or negative results.

12.2. References:

L. De Groot: *Thyroid Disease Manager*, , chapters 1 -7http://www.thyroidmanger.org

1. L.M Demers, C. Spenser: *LABORATORY MEDICINE PRACTICE GUIDELINES Laboratory Support for the Diagnosis and Monitoring of Thyroid Disease*;NACB 2002 , český překlad 2009, vydal AV Pron, Plzeň, ISBN 978-80-904492-0-6,
2. J.Fiala, J.Valenta : *Anatomie štítné žlázy*, CEVA [online] 15.11.2011, poslední aktualizace 15.11.2011 [17.11.2011]. Dostupný z WWW: http://www.ceva-du.cz/mod/resource/view.php?id=316 ISSN 1803-8999
3. Z. Límanová, Štítná žláza, kapitola 2 a 3, *Trendy soudobé endokrinologie*, Galén, Praha, 2006
4. L. J. DeGroot,J. L. Jameson *Endocrinology Volume 2*, W.B.Saunders Copany, 2001
5. L. Thomas: Clinical Laboratory Diagnostics, TH-Books, 1998

12.3. Laboratory Tests for Thyroid Gland Disorders

Author: MUDr. Richard Pikner, Ph.D.

Reviewer: doc. RNDr. Kristian Šafarčík, Ph.D., Ing. Vladimír Bartoš, Ph.D., MUDr. Daniel Rajdl, Ph.D.

12.3.1. Introduction

Specific laboratory tests for thyroid gland disorders are very important as laboratory signs of thyropathies are often more sensitive than clinical symptoms. On the other hand, some thyropathies do not manifest themselves in laboratory results (eufunctional goitre, thyroid nodules).

This text only deals with parameters covered by public health insurance and therefore used in normal practice.

Specific parameters of thyropathy tests include:

- Basic tests:
  - Thyroid function parameters: TSH Thyroid-stimulating hormone (TSH), free T_4 (FT_4), free T_3 (FT_3), total T_4 (TT_4), total T_3 (TT_3);
  - Autoimmune parameters: TSH receptor antibodies (Anti TSH, TRAK, TRAb), thyroid peroxidase antibodies (Anti TPO,TPOAb), thyroglobulin antibodies (Anti Tg, TgAb).
- Other tests: Thyroglobulin (Tg), iodineuria, Thyroxin-Binding Globulin (TBG), calcitonin

12.3.2. Indications for Testing

12.3.2.1. General

- TSH is a more sensitive parameter for thyropathies assuming proper functioning of the hypothalamus-pituitary axis and the absence of serious non-thyroidal disease;
- TSH tests with functional sensitivity of at least 0.02 mIU/l are recommended, as they allow thyroid hyperfunction to be distinguished from quite common TSH suppression;
- Tests for free T_4 and T_3 are preferred to their total concentrations;
- The T_3 test is only indicated for thyroid hyperfunction; it is not indicated for hypofunction since the major product of the thyroid gland is T_4;
- Autoantibodies are tested to clarify aetiology of the disease, assess the risk of thyropathy in the future and to monitor the effect of anti-TSH treatment;
- Anti-TPO and Anti-Tg tests are indicated every 2 years or when the clinical condition changes; always when potential thyropathy is suspected;
- Anti-Tg is tested with Tg if Tg is used as a tumour marker to monitor patients' condition after thyroid cancer treatment;
- Anti-TSH is indicated if Graves' thyrotoxicosis is suspected, to monitor the effect of its treatment, and to clarify the aetiology of exophthalmos;
- Thyroglobulin (Tg) test is indicated as a tumour marker in patients with differentiated thyroid cancer. Pre-operative tests are used to assess the scope of disease, the effect of treatment and for early diagnosis of a relapsed disease, always with Anti-Tg;
- In addition, Tg tests help monitor acute and subacute thyroid inflammations and when T_4 abuse is suspected;
- TBG tests are indicated if a genetic TBG production disorder is suspected;
- Iodineuria tests are primarily used in epidemiological studies to determine the status of iodine supply within the population;
- Calcitonin tests are indicated if medullary thyroid carcinoma is suspected for the same reasons as Tg tests; also in the differential diagnosis of hypocalcaemia.

12.3.2.2. Specific Diagnosis and Monitoring of Thyropathies

- The first-choice method for outpatients is the TSH test as this is also capable of diagnosing subclinical disorders as well;
- Laboratory tests (TSH, FT_4) of patients treated for thyroid hypofunction should be repeated after 6 weeks, when their condition is normalised every 6 months in the first 2 years and following on from this annually;
- Laboratory tests (TSH, FT_4, FT_3 and Anti-TSH for Graves' thyrotoxicosis) of patients treated for thyroid hyperfunction should be repeated after 6 weeks, and when their condition normalises every 6 months.

12.3.2.3. Specific Diagnosis and Monitoring of Thyropathy – Risk Groups

- In general, regular follow-up TSH tests should be indicated to risk groups;
- When a TSH level outside the reference interval is found, FT_4 should be tested (also FT_3 if hyperthyroidism is suspected), and repeated Anti-TPO, Anti-Tg using the same method are recommended either every year or every 2 years when the finding is negative;
- Groups at risk of thyropathy include:
  - Pregnant women – TSH, FT_4 and Anti-TPO at the beginning of pregnancy;
  - Patients with fertility disorders – TSH and Anti-TPO;
  - Women over 50 years – TSH every 4 years;
  - Patients with other autoimmune diseases (type I DM, celiac disease, etc.);
  - Patients with manifested IHD, tachyarrhythmias – TSH annually;
  - Patients with medication or treatment inducing thyropathies (amiodarone, lithium, interferons gamma or other, radiotherapy of the neck and chest) – TSH tests every 6 months during treatment.

12.3.3. Sampling – Sources of Variability during Sampling for Thyroid Tests

The presence of autoantibodies is not physiological, so there is no physiological biological variability, and all information relates to TSH, FT_4, FT_3, and Tg.

All thyroid analytes reveal higher inter-individual variability when compared with intra-individual variability. Inter-individual variability is particularly high for Tg and serum autoantibody concentrations 40 – 150 %. Serum TSH levels also show a high degree of inter- and intra-individual variability (25 %; 19 %). This fact primarily relates to the short TSH half-life (~60 minutes) and also circadian and diurnal variability.

Biological and analytical variability values could allow for changes (LSC) of minimal significance to be calculated for each parameter.

LSC refers to the change which has to occur in the measured parameter (considering the measurement uncertainty and biological variability) to have a significant change.

	LSC
FT_4	6.0 pmol/l
FT_3	1.5 pmol/l
TSH	0.75 mIU/l
Tg	1.5 µg/l (ng/ml)

LSC in % is more precise however absolute LSC values within the reference interval are presented just to give an illustration. For example, if TSH changes from the last test (6 weeks ago) from the baseline 3.6 mIU/l to 4.1 ml/l (0.5 mIU/l change), it is not a significant change, meaning the difference may be caused by the sum of the intra- and inter-individual variability error and analytical error.

- **Circadian rhythm:** Exists with lowest values in late morning and highest values before sleeping. TSH also shows pulsatile secretion with 2 – 6-hour intervals during the day and shorter intervals and frequency while asleep. TSH levels vary within the range of 0.5 – 2.0 multiples of the average daily level while the fluctuations are clinically insignificant.

- **Age** – the reference intervals of newborns and children are different. Foetus and newborn – TSH first appears around the 13th week of pregnancy, the values subsequently increase to many times the normal adult value. Childhood – TSH and reference interval values gradually decrease from birth; normal adult values are reached between 5 and 10 years of age; adult reference intervals are the same.

- **Gender** – no difference.

- **Pregnancy** – TSH drops in the first trimester in about 20 % of normal pregnancies, which is caused by thyroid gland stimulation by human chorionic gonadotropin (hCG). Minimum TSH and elevated FT_4 levels occur concurrently at week 10 – 12 of pregnancy. About 10 % of these cases (2 % of all pregnancies) may lead to gestational transient thyrotoxicosis (GTT) syndrome, which is typical of more or less clinically apparent symptoms of thyrotoxicosis. This clinical condition is often accompanied by more frequent vomiting in the first trimester of pregnancy. This is not serious in terms of clinical prognosis, and the problem usually subsides spontaneously.

Increased oestrogen production during pregnancy increases average TBG concentration 2 – 3 times. This results in a shift of the reference range for TT_4 and TT_3 in week 16 of gestation to about 1.5 times the non-pregnant value, but FT_4 and FT_3 practically do not change from non-pregnant values.

Serum Tg concentration usually rises throughout normal pregnancy. These values turn to normal 6 to 8 weeks after birth.

- **Hepatic or renal dysfunction** – mainly binding protein disorders; changes in free T_4 and T_3 are manifested only in severe deficiency cases.

- **Drugs** – thyroid parameters show frequent drug interactions; this is namely the case of binding protein concentration and thyroid hormone displacement from the bond to binding proteins. This can be avoided by choosing free T_4 and T_3 tests whose serum levels are relatively stable.
- Most drugs or pathological agents affect the TSH level by blocking or stimulating at the hypothalamus and pituitary level (typical examples are the inhibition of TSH secretion by dopamine, GH-RH analogs, stress, etc.).
- **Amiodarone** – a drug with a high iodine content; long-term administration induces thyropathy in about 15 % of patients, commonly with prior autoimmune thyroid disease (AITD).

Use of T_4 (levothyroxine) – use of drugs containing T_4 (Letrox, Euthyrox) markedly increases FT_4 levels from 30 – 60 % for at least 6 hours from the use of the drug. FT_3 levels respond similarly if combined formulations containing T_4 and T_3 are used. On principle, samples are taken from these patients before using drugs with T_4 or T_3.

- **Stability** – thyroid hormones are relatively stable; serum: 24 – 48 hrs at +22 °C, 5 – 7 days at +4 – 8 °C, and at least 6 months at -20 °C (this applies to all parameters and antibodies)

- **Freezing** – does not affect the levels provided the serum is properly mixed after thawing by inverting the test tubes (vortex mixing is not sufficient).

- **Haemolysis, lipaemia and hyperbilirubineamia** – do not significantly interfere with immunoassays.

- **Gel tubes** – samples taken into gel tubes do not affect the results of most TSH or thyroid hormone tests.

- **Heterophilic antibodies (HAMA)**

Patient serum may contain heterophilic HAMA antibodies classified into two classes. Either these are relatively weak multi-specific, polyreactive antibodies, often IgM rheumatoid factors, or alternatively can be widely reactive antibodies induced by infection or treatment containing monoclonal antibodies. They are sometimes referred to as human anti-mouse antibodies (HAMA). On occasions they may be specific human anti-animal immunoglobulins (HAAA) created against well defined specific antigens following the administration of a therapeutic agent containing animal antigens (mouse antibodies), or alternatively occur randomly by immunisation at the workplace (by handling animals). Both HAMA and HAAA rather affect the non-competitive test methods than competitive immunoassays by creating bridges between binding and detection antibodies, thus generating a false signal that causes inadequately high values.

- Non-Thyroidal Illness (NTI)

Patients with a serious illness often have pathological thyroid tests, although thyroid dysfunction is not found. This is caused by poorly regulated central inhibition of hypothalamic releasing hormones including the TRH.

FT_4 and TSH tests are useful only if clinical symptoms are present or the patient has a history of thyroid disorders. In addition to the cause mentioned above, the illness may be induced by various drugs such as dopaminergic drugs, glucocorticoids, furosemide or heparin, which either inhibit the pituitary TSH directly or indirectly by inhibiting the T_4 bond to serum proteins. The diagram of thyroid hormone changes in NTI is shown in Figure 11.1.

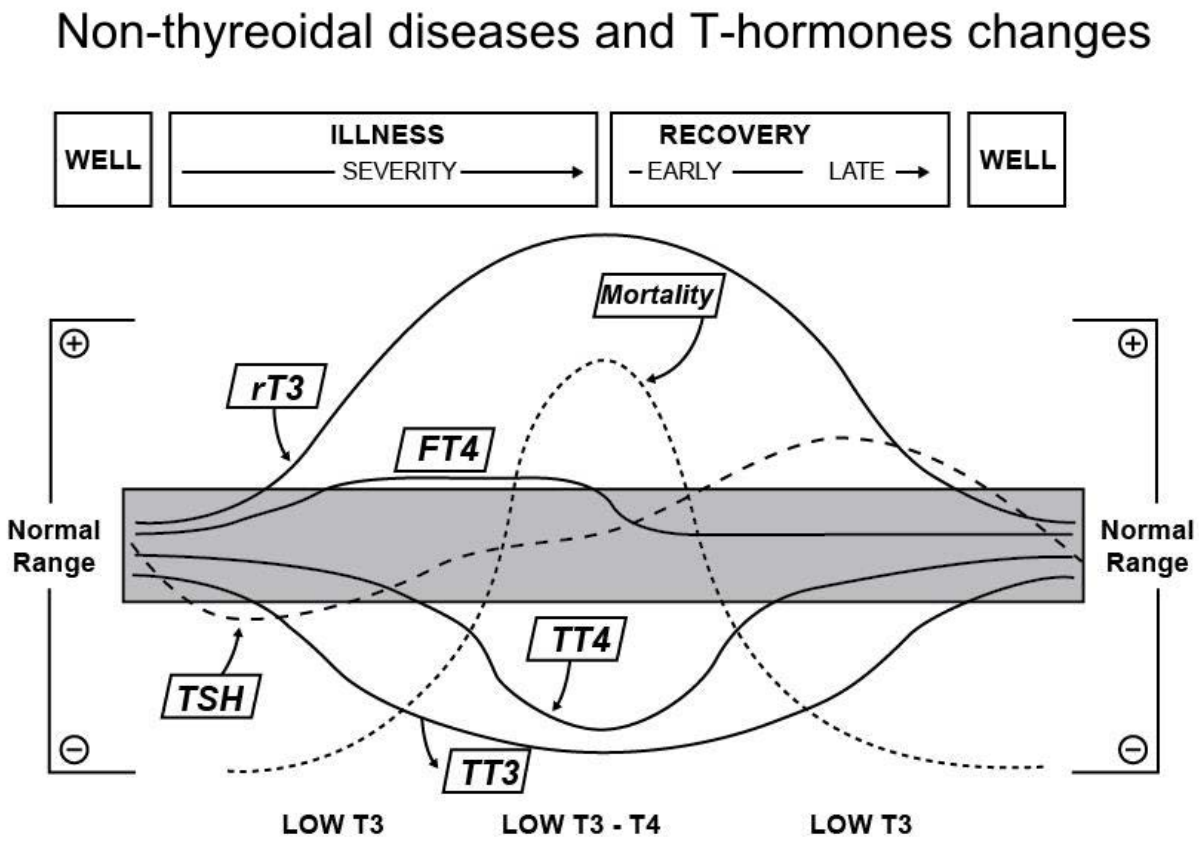


Figure 12.7. Non-thyroideal illnesses and T-hormones changes

Diagnosis and treatment of thyroid disorders in patients with serious non-thyroid illnesses are not at all easy, so it is recommended to collaborate with specialised endocrinologists, or, preferably, postpone the thyroid hormone testing. Levels 2 – 3 times as distant from the reference range do not typically indicate a serious thyroid illness.

12.3.4. Interpreting Laboratory Results of Thyroid Parameters

12.3.4.1. Reference Limits (Intervals)

The correct evaluation of laboratory results requires the determination of reference limits. Reference limits are generally used to diagnose a disease. Reference limits have to consider the incidence of thyroid disease, specific physiological conditions during pregnancy, age, gender and other aspects (selected population). When treating thyroid diseases, terms such as replacement or suppressive therapy are used, which means clinical conditions which have recommended target concentrations of thyroid tests regardless of reference limits refer to arbitrary limits. Furthermore, one should keep in mind that immunoassays are able to measure extremely low concentrations of free fractions (pmol/l), however their results depend on the method used, in particular as regards antibody quality, selection of standards and other factors, which may lead to marked differences of results between laboratories. These differences are relatively small for THS and free FT_4 and FT_3 (20 – 30 % on average), but very significant for autoantibodies (anti-TSH, Anti-TPO, anti-Tg) and thyroglobulin. Apart from the pressure on mutual standardization between manufacturers, the only solution is to interpret results using reference limits corresponding to the method used.

Two types of limits are used in laboratory diagnosis of thyropathies:

Reference limits for diagnosing thyropathies

These limits have been set based on the reference population sample using selection criteria with respect to the relatively high prevalence of thyropathies (autoimmune thyroid diseases or iodine deficiency) being over 5 %. The 95% confidence interval of the general "healthy" population cannot be used to set reference limits – really healthy individuals have to be selected from this population. Selection criteria are specific for the particular laboratory method. The following criteria have been recommended by consensus for setting reference limits in thyroid diagnosis:

Selection criteria for TSH, FT_4 and FT_3

Reference intervals should be set based on 95% confidence interval (for TSH log transformed) of values from 120 healthy euthyroid volunteers (60 males and 60 females) examined who:

- Had negative serum Anti-TPO and Anti-TgAb – immunoassays;
- Do not have thyroid dysfunction or other autoimmune illnesses in their personal or family history;
- Do not have any visible or palpable goitre;
- Do not use any drugs;
- Ideal age limit is 70 – 75 years, had ultrasonic thyroid examination.

Expected upper and lower reference limits:

TSH	0.3 – 0.5 and 3.5 – 4.5 mIU/l
FT_4	9.0 – 11.0 and 21.0 – 24.0 pmol/l
FT_3	3.2 – 3.6 and 6.1 – 6.5 pmol/l.

Selection criteria for Anti-TPO, Anti-Tg and Anti-TSHR

Reference intervals for thyroid antibody tests are significantly method-dependent. They should be set based on tests of 120 "normal" individuals without a history of thyroid illnesses. Such individuals should be selected so as to exclude AITD-predisposed persons. Ideally their thyroid should be examined on ultrasound. The following individuals are considered normal:

- Males – age 18 – 30 years, no visible or palpable goitre;
- Serum TSH concentrations 0.5 – 2.0 mIU/l;
- No thyroid disease in personal or family history;
- No non-thyroid autoimmune illness such as celiac disease, type I DM, lupus or other diseases.

Given the great variability of results between manufacturers, quantitative reference limits cannot usually be specified. The results are generally either positive or negative.

Selection criteria for Thyroglobulin (Tg)

The reference interval for Tg should be set depending on the locality as serum Tg concentrations are affected by iodine intake. In countries with sufficient iodine supply (including the CR) the reference interval for serum Tg for the euthyroid population (TSH, FT_4 within the reference range plus negative to Anti-Tg antibodies) ranges from 3 to 40 µg/l.

The following target values (arbitrary limits) are recommended for patients being treated for thyropathies (replacement therapy):

Replacement therapy with levothyroxine (LT_4):

Target TSH value: 0.3 – 3.0 mIU/l, near upper threshold values are tolerated in elderly (over 80 years) or on an individual basis. FT_4 levels are within the reference limits.

Suppressive therapy with levothyroxine (LT_4):

In the case of patients post thyroid surgery and radioiodine therapy for differentiated thyroid cancer, levothyroxine

(LT_4) doses are indicated in order to suppress TSH under the normal reference limit, thus reducing its proliferation--stimulating effect on thyrocytes. Target TSH values are arbitrarily recommended with FT_4 often being slightly over the upper limit.

- TSH under 0.1 mIU/l – high and medium-risk patients, within 5 years of operation;
- TSH: 0.1 – 0.5 mIU/l – for low-risk patients, or at 5 – 10 years from operation, providing the condition is favourable.

This recommendation does not apply to the preparation period for radioiodine therapy (targeted induction of hypothyroidism, TSH elevation) when it is desirable to reach TSH over 30 mlU/l either by discontinuing the levothyroxine treatment or by administering hrTSH.

Reference limits for paediatric population:

The hypothalamic-pituitary regulation axis of the paediatric population reacts to the changed environment and maturates after birth. That is why children under 1.5 years have markedly higher TSH levels, while FT_4 reaches its adult level at around 3 months of age. Figure 2 shows all ratios/multiples of TSH and FT_4 when compared with adults and demonstrates significant differences in TSH levels in the first 1.5 years of life, most markedly in the first months of life.

Usual values in children in proportion to adults

Age	TSH_{child}/ TSH_{adult}	TSH referential values (ml/U/l)	$FT4_{child}$/$FT4_{adult}$	FT4 referential values (pmol/L (ng/dL))
Foetus in the middle of pregnancy	2.41	0.7 – 11	0.2	2 – 4 (0.15 – 0.34)
Serum oft he umbilical vein	4.49	1.3 – 20	0.8	8 – 17 (0.64 – 1.4)
Newborn during the birth	4.28	1.3 – 19	1	10 – 22 (0.8 – 1.9)
3-day newborn	3.66	1.1 – 17	2.3	22 – 49 (1.8 – 4.1)
10-week infant	2.13	0.6 – 10	1	9 – 21 (0.8 – 1.7)
14-month child	1.4	0.4 – 7.0	0.8	8 – 17 (0.6 – 1.4)
5-year child	1.2	0.4 – 6.0	0.9	9 – 20 (0.8 – 1.7)
14-year child	0.97	0.4 – 5.0	0.8	8 – 17 (0.6 – 1.4)
Adult	1	0.4 – 4.0	1	9 – 22 (0.8 – 1.8)

*FT_4 assessed using the method of equilibration dialysis

Figure 12.8. Values for children vs. values for adults

Reference limits of pregnant women

Pregnancy has a major effect on the TSH level, which drops significantly with the growing hCG level; the greatest TSH suppression can be expected at the end of the first trimester. FT_4 rises slightly in the first trimester, and then, conversely, drops slightly under the average of normal population until the end of pregnancy. Trimester-specific reference limits should be used for TSH. Changes in free FT_4 near the limits are significant, while values for the normal population can be used with an individual assessment.

The following TSH values can be expected during pregnancy:

1st trimester	TSH 0.1 – 2.5 mIU/l
2nd trimester	TSH 0.2 – 3.0 mIU/l
3rd trimester	TSH 0.3 – 3.0 mIU/l

12.3.4.2. Interpretation of Thyroid Test Results

The patient's clinical condition, risk factors and history always have to be taken into account when interpreting results. Drug interferences and stress should be taken into account in hospitalized and multimorbid outpatients.

The following figure lists expected thyroid parameter values, however the possible effects of coincident diseases and drug interferences are not considered.

Diagnosis /Laboratory tests	**TSH**	**FT_4**	**FT_3**	**Tg**	**antiTPO**	**antiTg**	**antiTSHR**
Central disorders							
Central hypothyroidism	↑N↓	↓		↓	neg.	neg.	neg.
Central hyperthyroidism	↑	↑	↑	↓	neg.	neg.	neg.
Primary disorders - hypofunction							
Subclinical hypothyroidism	↑	N		N	neg./pos.	neg./pos.	neg./pos.
Hypothyroidism	↑↑	↓		↓N↑	neg./pos.	neg./pos.	neg./pos.
Deficiency in iodine hypothyroidism	↑↑	↓		↑↑	neg.	pos.	neg.
Postresection hypothyreosis	↑↑	↓		neg.			
Hypofunctional phase of AITD	↑↑	↓		↓	pos.	pos.	neg./pos.
Primary disorders – hyperfunction							
Subclinical hyperthyreosis	↓	N	N	N↑			
Grave's disease	↓↓	↑↑	↑↑	↑↑			pos.
Hyperfunctional adenoma	↓	↑	↑	↑N			neg.
T_3 - hyperfunction	↓	N	↑	↑N			neg.
Chronic overuse of T_4	↓↓	↑↑	↑↑	↓ negat.			
Transitory pregnancy toxicosis	↓↓	↑	↑	↑N			
Hyperfunctional phase of AITD	↓↓	↑↑	↑	↑↑	pos.	pos.	neg./pos.
Other disorders							
Eufunctional struma (diffuse, nodose)	N	N	N	↑ N			
Hashimoto's thyroiditis (AITD)	N	N	N	N	pos.	pos.	neg.
Circulating antiTSH	↑N↓	↑N↓	↑N↓	N↑	neg.	neg.	pos.
T_3 and T_4 resistance	↑N	↑↑	↑↑	↑↑			
Low T_3 syndrome	N↓	N↓	↓↓				

Figure 12.9. Basic interpretation of thyroideal tests

Specific Interpretations

Changes in thyroid hormone levels after initiating thyroid hypofunction or hyperfuction treatment:

As demonstrated in the following figure, the development of thyroid dysfunction is relatively slow, starting with the subclinical phase (change to TSH with free fractions within the reference limits), and gradually transferring to the clinical phase (periods 1 and 2). When treatment is started, free fractions and TSH gradually return to normal; this transitional period may last from some weeks to 9 – 12 months (see Fig. 11.4).

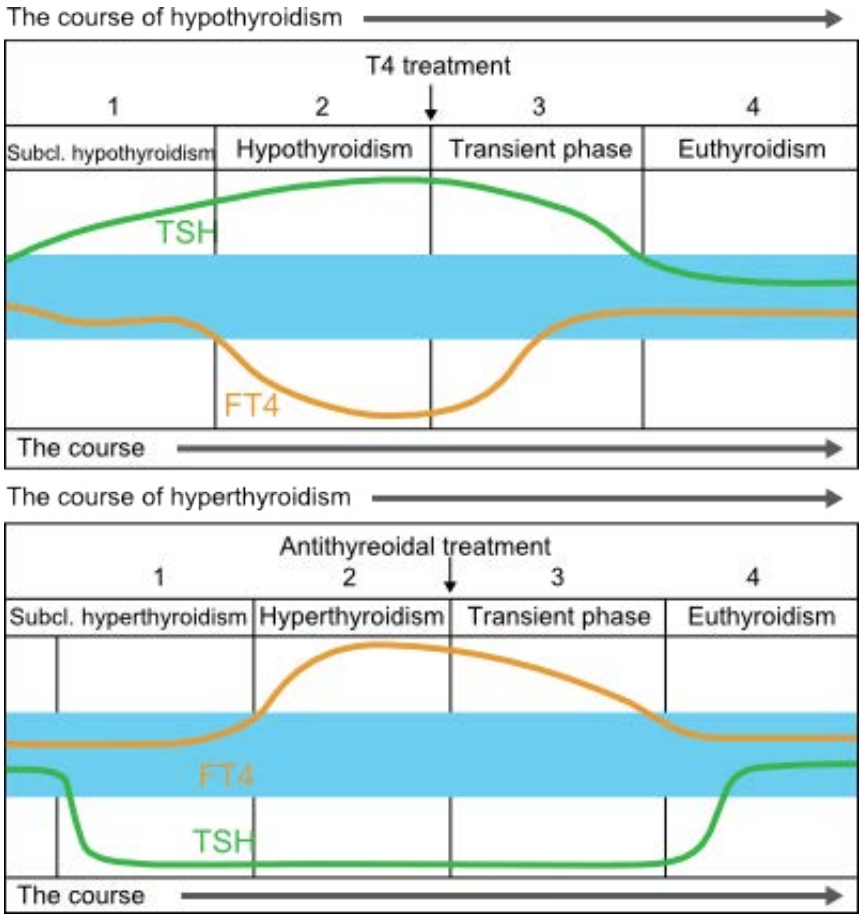


Figure 12.10. The course of hypothyroidism and hyperthyroidism

- **Primary hypofunction**, most commonly on an autoimmune basis, demonstrated by a positive anti-TPO or Anti-Tg, exceptionally positive blocking anti-TSHR.

The subclinical (latent) form: only TSH level changes however FT_4 or FT_3 remain within the reference interval; usually no or minimally expressed clinical symptoms. Diagnosis of subclinical thyropathy (pathological TSH, normal FT_4) should be confirmed by at least two pathological results of TSH at 6 – 8-week intervals (the follow-up result is normal in up to 25 % cases).

Manifest form – the disorder is fully developed: clinical symptoms, changed TSH and FT_4

- **Increased thyroid function (primary hyperfunction, thyrotoxicosis):** suppressed to an undetectable TSH level, elevated FT_4 and FT_3. Roughly 5 % of cases only indicate sporadically elevated FT_3 (**T_3-thyrotoxicosis**).

- **Graves' disease** – typically positive Anti-TSHR (an aetiological factor of the disease), but often only medium-positive Anti-TPO, Anti-Tg, undetectable TSH and high FT_4 and FT_3. There are also blocking anti-TSHR, which mostly induce clinically apparent hypofunctions with an orbitopathy. A change in the antibody response from stimulating to blocking antibodies (from hyperfunction to hypofunction) can occur in the patient on relatively rare occasions.

- **Amiodarone-induced thyrotoxicosis** occurs in about 15 % treated persons as a result of excess iodine in amiodarone, which causes thyrotoxicosis: present suppression of (undetectable) TSH, elevated FT_4 and slightly elevated FT_3, antibodies may or may not be identified – help distinguish between type I and II. Anti-TPO are positive in type 1; ultrasound examination helps distinguish between the two types.

- **Subacute thyroiditis** at the incipient stage – the clinical presentation (pain, high FW), laboratory signs of hyperfunction (thyrocyte causes damage – release of the reserve of produced T-hormones), negative antibodies, low ^{131}I accumulation, and uneven activity distribution on scintigraphy, high thyroglobulin level (marker of thyrocyte damage).

- **Autoimmune thyroid disease (AITD) – Hashimoto's thyroiditis**

Prevalence 10 %; fertile women are usually eufunctional; positivity grows with age; up to 40 – 50 % in women over 65 years; prevalence among men is several times lower.

They can be eufunctional; positive anti-TPO induces chronic lymphocytic inflammation, which, if prolonged, leads to thyroid destruction and ultimately to the development of subclinical and clinical primary hypofunction. The dysfunction development time is quite individual; the higher the anti-TPO concentration, the more likely it will develop quickly.

Acute exacerbation (example: postpartum thyroiditis) first takes place as a hyperfunction stage followed by a hypofunction stage, which often becomes permanent, or the function normalizes. A further antibody response may develop, for example anti-TSHT and the development of Graves' thyrotoxicosis.

The eufunction stage features TSH and FT_4 within the reference limits, positive Anti-TPO and/or Anti-Tg.

- **Postpartum thyroiditis** develops in the postpartum period in 50 % of women with positive Anti-TPO with a typical waveform as shown in the following figure (Figure 11.5). Anti-TPO and/or Anti-Tg antibodies increase after birth.

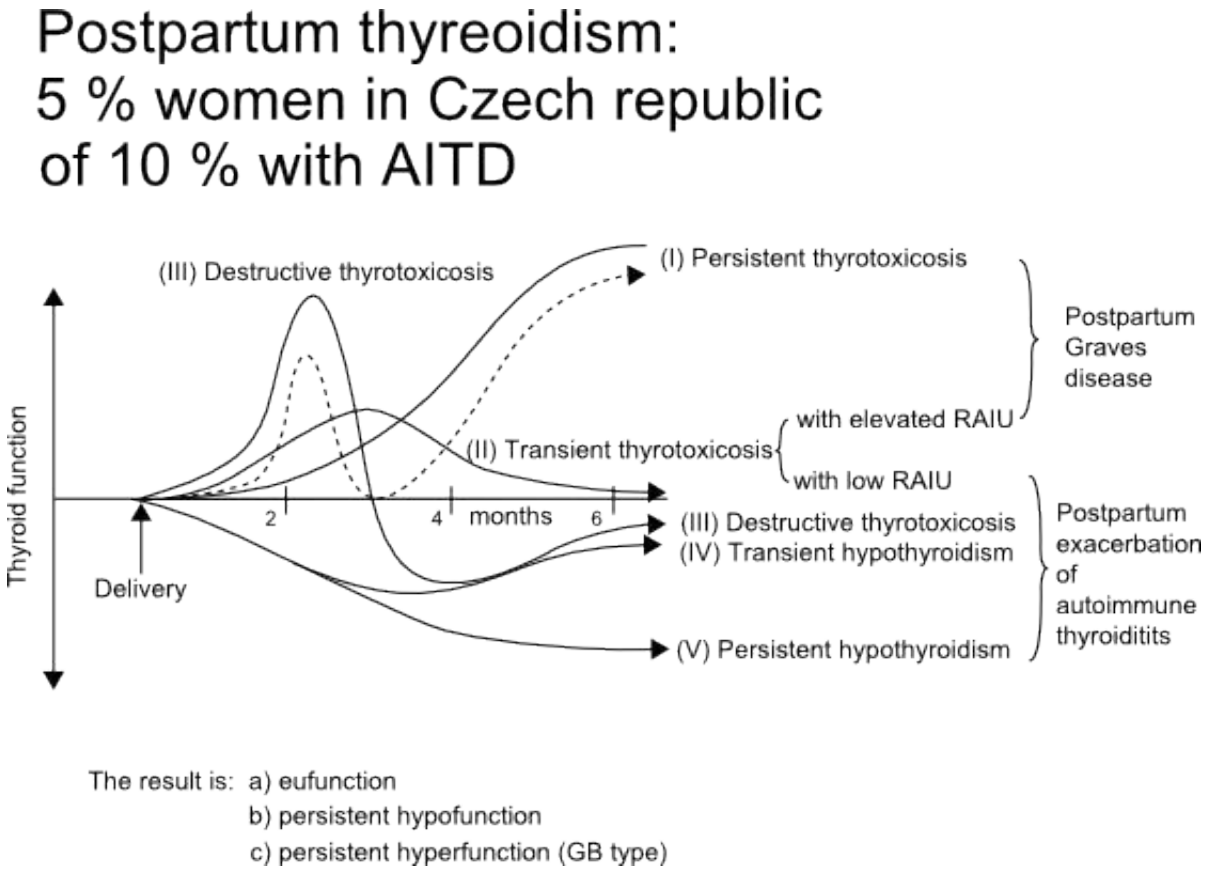


Figure 12.11. Postpartum thyreoidism

- **Non-thyroid disease**

Often observed in acutely hospitalised patients with a clinically serious condition (shock, sepsis), extensive surgeries, patients hospitalised at the ICU – any TSH, FT_4 and FT_3 outside the reference limits can be found. This is due to many factors: caloric deficit, hyposaturation of tissues with oxygen, the effect of cytokinins acting at the central level. The administration of dopamine and similar substances features TSH suppression or even undetectable TSH and normal FT_4.

- **Low T_3 syndrome:** primarily in patients with a metabolic disorder, shock conditions or other clinically very serious diseases. Develops on the basis of elevated T_4 conversion to (reversible) metabolically inactive rT_3; laboratory results include low FT_3 levels and normal or slightly decreased FT_4 and TSH. A critical decrease in the T_3 level is a poor prognostic indicator.

Thyroid parameters should always be treated with caution in non-thyroid diseases; slight laboratory imbalances of 2 – 3 times the normal values usually do not present serious thyropathy and are induced by non-thyroid diseases.

- **Thyroid hormone overdosing** (tyreotoxicosis factitia) – laboratory signs of hyperfunction but a very low thyroglobulin level and a low storage capacity of the radiopharmaceutical.

- **Central hypothyroidism** features decreased FT_4, while TSH may be slightly under the lower limit, normal or slightly elevated (if biologically inactive TSH is produced). TRH test results are pathological.

- **Central hyperfunction:** the most common cause is pituitary adenoma with TSH overproduction; TSH, FT_4 and FT_3 levels subsequently increase.

- **Thyroid hormone resistance syndrome**

Usually caused by thyroid hormone receptor beta mutations occurring in 1:50,000 live born children. Serum FT_4 and FT_3 levels are typically increased and are associated with a normal or slightly increased serum TSH level. However, TSH secretion is not disproportionate as the tissue response to the thyroid hormone is decreased and requires higher thyroid hormone levels to maintain a normal metabolism. TSH is markedly elevated in the TRH test. There are no symptoms of hyperthyroidism however tachycardia or dysrhythmia may occur.

12.3.5. Discordant Result

This is a situation when the measured laboratory results do not correspond with the patient's clinical condition. This is identified by the attending physician, who should notify the responsible laboratory about it.

Discordant results may have many causes: ranging from a technical error including wrong sample identification to an error in the laboratory procedure, or the presence of interfering agents or unusual isoforms such as TSH isoforms.

Recommended steps:

- Notify the testing laboratory about the discordant result;
- Compare the discordant result with possible interfering drugs which the patient is taking;
- Verify that the sample has been correctly identified;
- Request a repeated sample measurement;
- Take a new sample and measure the sample.

Request a sample measurement in another laboratory using another technology, if the inter-laboratory variability of results is over 50 %, interference is probable (does not apply to Anti-TPO and Anti-Tg measurements where variability between manufacturers is in the order of tens to hundreds in terms of percent).

The laboratory takes further steps in line with their own internal quality control system.

12.3.6. Summary – Key Information:

12.3.6.1. Indications for Testing

General:

- TSH tests with functional sensitivity of at least 0.02 mIU/l are recommended;
- Free T_4 and T_3 tests are preferred;
- The T_3 test is only indicated for thyroid hyperfunction;
- Anti-TPO and Anti-Tg tests are indicated when autoimmune thyroid disease is suspected, as well as for risk groups;
- Anti-TSH is indicated if Graves' thyrotoxicosis is suspected, to monitor the effect of its treatment, and to clarify the aetiology of exophthalmos;
- Thyroglobulin (Tg) test is indicated as a tumour marker in patients with differentiated thyroid cancer, always with Anti-Tg;
- Calcitonin tests are indicated if medullary thyroid carcinoma is suspected.

Specific Diagnosis and Monitoring of Thyropathies

- The first-choice method for outpatients is the TSH test since this is also capable of diagnosing subclinical disorders;
- TSH and FT_4 tests are carried out in patients undergoing treatment for thyroid hypofunction;
- TSH, FT_4 and FT_3 tests are carried out in patients undergoing treatment for thyroid hyperfunction, and Anti-TSH for Graves' thyrotoxicosis.

Specific Diagnosis and Monitoring of Thyropathy – Risk Groups

- Regular follow-up tests of TSH, Anti-TPO and Anti-Tg are indicated for risk groups every 2 years;
- Groups at risk of thyropathy include:
  - Pregnant women;
  - Patients with fertility disorders;
  - Women over 50 years;
  - Patients with other autoimmune diseases (type I DM, celiac disease, etc.);
  - Patients with IHD manifestation and tachyarrythmias;
  - Patients on medication or treatment inducing thyropathies (amiodarone, lithium, interferons gamma or other, radiotherapy of the neck and chest).

12.3.6.2. Collection:

- Samples may be taken at any time in the morning however always before using a drug containing T_4 or T_3 (FT_4 and FT_3 levels increase by 30 – 60 % for 6 hours);
- The sample may be stored in the fridge until the next day; refrigerated serum is stable for one week;
- **Reference limits do not vary with age and gender except for children under 5 years** (TSH levels are higher and the FT_4 and FT_3 range is wider in children under 5 years);
- Interferences of drugs and external influences such as stress primarily affect the TSH and total T_4 and T_3 (TT_4 and TT_3) assays;
- Iodineuria is measured in the first morning urine; store in the fridge; no urine stabilizers are required.

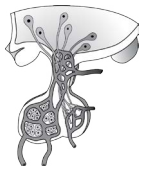
12.3.6.3. Interpretation of Results:

- Results have to be assessed considering the clinical condition, patient history and drugs taken;
- Reference limits, if used for diagnosing thyroid disorders, should be defined based on the selected population of healthy individuals according to the recommended selection criteria;
- Pregnant women have specific TSH reference limits for each trimester. The first trimester is characterized by a decrease in TSH with a slight elevation in FT_4 levels (thyroid stimulation by the hCG molecule);
- Patients on replacement therapy and patients undergoing treatment for differentiated thyroid cancer have arbitrarily recommended target TSH levels that differ from the reference limits;
- Subclinical disorders usually feature absent or minimal of clinical symptoms, while the TSH level is outside the reference limits, and free hormone levels are within the reference limits;
- Primary disorder means a disorder afflicting the thyroid gland. Hypofunction is characterised by low FT_4 and elevated TSH levels, while hyperfunction by elevated FT_4 and FT_3, and low or undetectable TSH concentrations;
- The most common cause of primary hypofunction is autoimmune thyroid disease (Hashimoto's thyroiditis, AITD) with typically positive Anti-TPO and/or Anti-Tg results. The most frequent cause of hyperfunction is thyrotoxicosis (Graves' disease) with typically positive Anti-TSH autoantibodies. Positive autoantibodies are not physiological, and their presence with normal TSH and FT_4 is indicative of eufunctional thyroiditis;
- Successful treatment of patients with thyropathies leads to the faster normalisation of free hormones (FT_4 and FT_3); TSH reverts to normal following some weeks up to 1 year of treatment;
- Hospitalised patients with a clinically serious condition (shock, sepsis), extensive surgeries, patients hospitalised in the ICU have imbalanced thyroid parameters, ranging from low T_3 syndrome to mere TSH suppression due to dopamine treatment. This may lead to great differences in results, which is usually not caused by thyroid disease but rather the acute condition and its treatment. Thyroid tests revert to normal after convalescence;
- A discordant result is a situation when the measured laboratory results do not correspond with the patient's clinical condition. This is identified by the attending physician, who should notify the responsible laboratory about it. Discordant results may have many causes: ranging from a technical error including wrong sample identification to an error occurring during a laboratory procedure, or the presence of interfering agents or unusual isoforms such as TSH isoforms.

12.4. References:

1. L. De Groot: *Thyroid Disease Manager*, , chapters 1 – 7 http://www.thyroidmanger.org

2. L.M Demers, C. Spenser: LABORATORY MEDICINE PRACTICE GUIDELINES Laboratory Support for the Diagnosis and Monitoring of Thyroid Disease;NACB 2002 , český překlad 2009, vydal AV Pron, Plzeň, ISBN 978-80-904492-0-6
3. Z. Límanová, Štítná žláza, kapitola 2 a 3, Trendy soudobé endokrinologie, Galén, Praha, 2006
4. L. J. DeGroot,J. L. Jameson *Endocrinology Volume 2*, W.B.Saunders Copany, 2001
5. L. Thomas: *Clinical Laboratory Diagnostics*, TH-Books, 199
6. Límanová Z., Pikner R., Springer D.: *Doporučení pro laboratorní diagnostiku funkčních a autoimunních onemocnění štítné žlázy*, Klin. Bioechem. Metab. 19(40),No1, 2010: 48-61
7. Stagnaro-Green A., Abalovich M., Alexander E. et al : *Guidelines of ATA for Diagnosis and managment of Thyroid Disease during preganncy na dpostpartum*: Thyroid, vol 21, No10, 2011: 1081-1125
8. Cooper D., Doherty G., Haugen B. et al: *Revised ATA management guidelines for patients with thyroid nodules and differentiated thyroid cancer*, Thyroid vol19, No 11, 2009 , 1167-1214



CHAPTER 13

13. Hormones of hypothalamus and hypophysis

Author: MUDr. Tomáš Franěk

Reviewers: doc. RNDR. Kristian Šafarčík, Ing. Vladimír Bartoš, Ph.D.

13.1. Prolactin

Prolactin is a polypeptide hormone produced in eosinophilic cells of the anterior pituitary, whose secretion is centrally inhibited from the hypothalamus through the action of prolactin-inhibiting hormone (PIH) and dopamine; conversely, its secretion is stimulated by the effect of serotonergic neurons, prolactin-releasing hormone and some drugs. Prolactin stimulates the transformation of lobuloalveolar epithelium in the mammary gland into the secretory type with subsequent breast milk formation. The action of prolactin on ovaries inhibits follicle maturation and aromatase activity with a subsequent decrease in oestradiol synthesis.

Indications for the prolactin assay include infertility, menstrual cycle disorders such as amenorrhoea, dysmenorrhoea and galactorrhoea, gynaecomastia in males, tumours and traumas in the sella turcica region, and signs of virilization.

Physiological elevated values can be found in pregnant women, postpartum lactation, extreme stress, prolactinomas and ectopic production. Furthermore, prolactin secretion is stimulated by the following drugs: chlorpromazine, haloperidol, phenothiazines, cimetidine, tricyclic antidepressants and reserpine, metoclopramide.

13.2. FSH – Follicle Stimulating Hormone

FSH is a glycoprotein secreted by the basophilic gonadotrophic cells of the anterior pituitary. The function of FSH is to stimulate the growth and maturation of ovarian follicles; FSH together with LH support oestrogen secretion and participate in endometrium transformation in the proliferative phase of menstrual cycle. The stimulation of Sertoli cells by the FSH helps maintain male spermatogenic epithelium, and affects the synthesis of inhibin and binding protein for angrogens (SHBG) in seminiferous tubules.

Indications for the FSH assay include amenorrhoea, dysmenorrhoea, oligomenorrhoea, infertility, primary and secondary hypogonadism, conditions following cytostatic treatment or radiation, atypical sexual maturation in children, pituitary traumas and tumours, gonadal dysgenesis, and hypothalamic-pituitary-gonadal axis disorders. The LH and FSH assay is also important for diagnosis in polycystic ovary syndrome where the FSH level decreases or is normal and the LH level increases, so that the LH/FSH ratio increases in contrast to healthy individuals.

Elevated values can be found in primary ovarian insufficiency, Turner syndrome, climacterium praecox, gonadotropinomas, and physiologically elevated values in the menopausal period and primary hypogonadism in males. Reduced levels can be found in pituitary traumas and endocrine-inactive tumours, anorexia nervosa, secondary ovarian insufficiency and pseudopubertas praecox.

13.3. LH – Luteinizing Hormone

LH is a glycoprotein secreted by the basophilic gonadotropic cells of the adenohypophysis. The function of LH is to act on ovarian theca cells where stimulation of steroid synthesis takes place. With the assistance of FSH, the steroids subsequently transform into oestradiol. In addition, LH is responsible for the final ripening of the ovarian follicles and also induces ovulation and the early development of the corpus luteum with subsequent progesterone secretion. In males, LH stimulates testosterone synthesis in Leydig cells in the interstitial tissue of the testicles.

Indications for the LH assay include amenorrhoea, dysmenorrhoea, oligomenorrhoea, infertility, primary and secondary hypogonadism, conditions following cytostatic treatment or radiation, atypical sexual maturation in children, pituitary traumas and tumours, gonadal dysgenesis, selected chromosome aberrations, and hypothalamic-pituitary--gonadal axis disorders.

Elevated values can be found in primary ovarian insufficiency, Turner syndrome, climacterium praecox, gonadotropinomas, and physiologically elevated values in the menopausal period and primary hypogonadism in males. Reduced levels can be found in pituitary traumas and tumours, anorexia nervosa, secondary ovarian insufficiency and pseudopubertas praecox.

13.4. Oxytocin

Oxytocin is a peptide hormone formed primarily in the paraventricular nuclei of the hypothalamus, from where it is transported and deposited in the granules of posterior pituitary nerve endings. The function of this hormone is to act on myoepithelial cells lining mammary gland outlets, which causes milk to be ejected from the mammary gland. Another important function is the effect on the smooth muscles of the uterus during pregnancy, by which it enhances labour at the end of pregnancy. Assays for this hormone are not common as their importance for diagnosis is low.

13.5. ADH – Antidiuretic Hormone

ADH or vasopressin is another hormone stored in the posterior pituitary. This hormone is composed of 9 amino acids. The basic physiological action of ADH is water retention in the body, and ADH thus participates in the homeostasis of the body. ADH is secreted as a result of an increase in blood plasma osmolality, and its function is to increase the permeability of distal tubules and collecting ducts in the kidneys so thatwater can enter the hypertonic renal interstitium. Secretion disorders, or disorders at the peripheral receptor level, will manifest themselves as diabetes insipidus with the inability of the kidneys to concentrate urine, with specific clinical impacts. The opposite situation occurs following some traumas or bleeding in the CNS when ADH overproduction occurs with normal blood plasma osmolality. This condition is referred to as inappropriate ADH secretion (SIADH). Indications for the ADH assay include diabetes insipidus and SIADH.

13.6. TSH – Thyroid-Stimulating Hormone or Thyrotropin

TSH is a peptide hormone synthesized by thyrotrope cells in the anterior pituitary gland. The function of this hormone is to stimulate thyroid cells to produce peripheral hormones, i.e. thyroxine and triiodothyronine. Secretion of this tropic hormone is partly regulated by direct inhibitory (negative) feedback caused by a high thyroid hormone level in the blood, which inhibits the adenohypophysis. A partial role is also played by neural mechanisms acting through the hypothalamus.

The assay for this hormone is essential for the diagnosis of thyroid gland disorders. Elevated levels can be found in hypothyroidism, subclinical hypothyroidism, following thyroidectomy, or pituitary adenoma producing TSH. Reduced levels can be found in hyperthyroidism, hypopituitarism and some non-thyroidal diseases such as febrile conditions, myocardial infarction, serious traumas and operations.

13.7. ACTH – Adrenocorticotropic Hormone

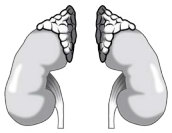
ACTH is a hormone synthesized in adenohypophysis cells in the form of a precursor, pro-opiomelanocortin (POMC). This precursor is subsequently cleaved by enzymes into several physiologically active peptides, one of them being ACTH. This hormone is the main hormone that regulates the production of glucocorticoid hormones in the adrenal cortex. Assays for this hormone are not common due to specific and complicated pre-analytical factors, and are solely used in the diagnosis of adrenal cortex function disorders, in particular to distinguish between a primary and secondary disorder of cortisol secretion.

Elevated levels can be found in primary hypercorticalism, Addison’s disease, adrenogenital syndrome, Cushing’s disease and adrenal damage. Reduced levels can be found in Cushing’s syndrome, when feedback ACTH inhibition occurs.

13.8. STH – Somatotropin (GH – growth hormone)

STH is a polypeptide hormone produced by acidophil cells in the adenohypophysis. Secretion is regulated from the hypothalamus by regulatory hormones such as somatoliberin and somatostatins. The function of this hormone is to control and stimulate growth due to different effects on metabolism. Major effects include the stimulation of proteosynthesis and inducing a positive nitrogen balance; furthermore, STH has anti-insulin effects and increases the content of free fatty acids in the blood plasma. Peripheral effects of this hormone depend on the production of peptide growth factors known as somatomedins (IGF-1 and IGF-2).

Indications for the STH assay include growth disorders, nanism or acromegaly. Elevated levels can be found in gigantism, endocrine-active pituitary tumours and acromegaly. Reduced levels can be seen in nanism.



CHAPTER 14

14. Adrenal Cortex Hormones

Author: Prof. MUDr. Richard Průša, CSc.

Reviewers: doc. RNDR. Kristian Šafarčík, Ing. Vladimír Bartoš, Ph.D.

14.1. Biochemistry Histology, Secretion Regulation and Effects of Adrenal Cortex Hormones

The adrenal gland of a newborn child weighs twenty times as much in proportion to their total weight than the gland of an adult. The foetal zone, considerably wide in the foetus, is found in the place of the reticular zone, although it does not form in anencephali. Maternal cortisol transfers to the circulation of the foetus and affects feedback ACTH production in the foetus's pituitary. Maternal hormones ensure the balance of electrolytes. Foetal zone involution begins 3 to 4 days after birth and proceeds very fast, so the basic adrenal cortex remodelling is finished in the third week after delivery. The newborn child is able to maintain internal homeostasis in normal conditions, although the adrenal cortex undergoes rapid remodelling. Exceptions to this are inherited defects of steroidogenic enzymes in the adrenal cortex, which may lead to, for example, congenital adrenal hyperplasia.

From the histological point of view, the adrenal cortex is differentiated into three zones, which also differ in the gene expression of enzymes that take part in the synthesis of corticoids. Zona glomerulosa primarily produces aldosterone (the zone does not have 11 beta-hydoxylase); zona fasciculata produces cortisol (the zone does not have P450--aldosynthase, i.e. 18-hydroxylase/18-oxidase); zona reticularis produces corticoids with an androgenic effect such as dehydroepiandrosterone (DHEA). DHEA levels are high at a young age and decrease with growing age. Foetal DHEA is converted into oestrogen in the placenta. Steroidogenic factor 1 (SF-1) is a special protein that takes part in corticoid synthesis. It is a transcription factor bound to the TGG CTA motif of the gene regulation domain. If absent, the adrenal gland and gonads are lost.

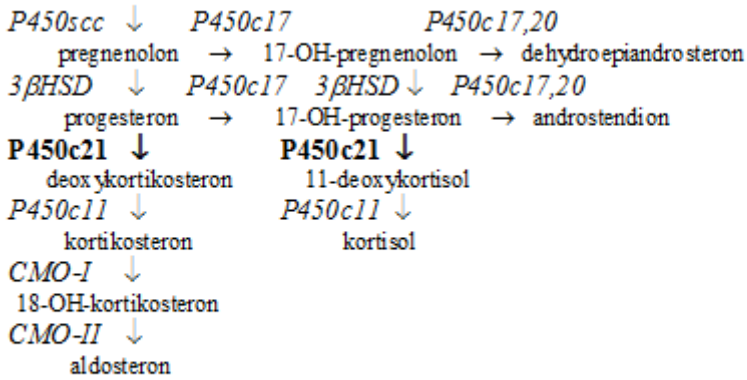


Figure 14.1. Synthesis of adrenal cortex hormones

The basic glucocorticoid is cortisol, whose secretion is regulated by corticoliberin or corticotropin-releasing hormone (CRH) from the hypothalamus, and by adrenocorticotropin (ACTH) from the adenohypophysis. Approximate amounts excreted per day are: cortisol 10-20 mg, corticosterone 3 mg, aldosterone 0.3 mg. The biological half-life of cortisol is about 100 minutes. This process typically involves feedback through glucocorticoid receptors in the hypothalamus and the pituitary gland. Disorders may also occur at a non-steroidogenic level: a gene defect for the glucocorticoid receptor and a gene mutation for the ACTH receptor have been described (they lead to familial glucocorticoid deficiency syndrome).

Essential for hypothalamic-pituitary-adrenal axis function is the gene for the nuclear hormone receptor (DAX-1), mutations or deletions of which lead to congenital hypoplasia linked with the X chromosome. A high level of cortisol induces a drop in CRH and ACTH production within minutes. If the cortisol concentration is high for a long time, the pro-opiomelanocortin (POMC) precursor synthesis is also lower. An important stimulus for cortisol secretion increase is stress situations, including hypoglycaemia or anxious reactions (important in paediatrics). The pronounced circadian rhythm of cortisol is used in the laboratory diagnosis. Lowest concentrations are usually at 4 a.m., while maximum levels of cortisol and ACTH are at 8 a.m. Cortisol in the blood is bound to proteins (about 90%) transcortin (cortisol--binding globulin – CBG) and albumin. For the most part, cortisol is metabolized in hepatocytes to 17-hydroxysteroids eliminated in the urine. More than 95% of metabolites conjugate in the liver with glucuronic acid (3-alpha-hydroxy group), and a smaller part is sulfatized at the 21-hydroxy group. The transformation of cortisone (a biologically inactive form that does not bind to glucocorticoid receptors) to cortisol also takes place in the liver. Only a small part is excreted in the urine as free cortisol.

Cortisol affects the transformation of basic metabolites: it stimulates the conversion of amino acids into glucose (gluconeogenesis) in the liver, stimulates proteolysis in muscles, and lipolysis in fatty tissue (through hormone-sensitive lipase). It has a significant effect on the immune reaction of the body, and has anti-allergic, anti-oedematous, anti-inflammatory and anti-exudative effects. Cortisol acts through nuclear steroid receptors. Cortisol or aldosterone first penetrates into the cytoplasma, binds to the steroid receptor, and releases the heat shock protein (Hsp90) that was previously bound to this structure at the carboxy-terminus. The receptor to which the hormone is not bound is prevented by hsp from translocation to the nucleus. Receptors for steroid hormones are dimeric transcription factors. After binding to the steroid, the complex penetrates into the nucleus, the conformation changes and the DNA binding site (with “zinc fingers”) is created together with the structure responsible for affecting promotor areas with negative and positive hormone response elements (HRE). These are short (15 bp) palindromic sequences.

14.2. Laboratory Diagnostics

14.2.1. Total Serum Cortisol

This test is used mainly to diagnose Cushing’s syndrome, in which the fluctuations in diurnal cortisol levels are suppressed. However, this test is less reliable than the 24-hr urinary free cortisol test. The test is made on serum or plasma collected between 8 and 9 a.m. or as a daily profile, always before meals, and haemolysis should be prevented. The patient must rest for at least 30 minutes before collection. The reference range depends on the method. The table shows guidance values only:

	Cord blood	Children 1-15 years	Adults 15-60 years
	138 - 773 nmol/l		
8:00 a.m.		414 - 690 nmol/l	138 - 773 nmol/l
5:00 p.m.		138 - 276 nmol/l	55 - 386 nmol/l
8:00 p.m.		Less than 50% of morning value	Less than 50% of morning value

Conversion factor: ug/dl x 27.6 = nmol/l

Table 14.1. The reference ranges for serum cortisol

Heterophilic antibodies in the human serum may react with immunoglobulins in reagents and interfere; cross reactions have not been proved, however.

Elevated cortisol levels are tested for the following diagnoses: burns, Crohn’s diseases, Cushing’s disease, Cushing’s syndrome, ectopic ACTH production, eclampsia, severe hepatopathies, hyperpituitarism, hypertension, hyperthyroidism, severe infections, obesity, osteoporosis, acute pancreatitis, pregnancy (elevated CBG, cortisol levels may be up to 2.5 times higher in the third trimester), severe renal diseases, shock conditions, increased stress (heat, cold, traumatic or mental stress), surgery, virilization. Cortisol levels increase following the administration of oestrogens, oral contraception, yohimbin, vasopressin, also after amphetamines, metamphetamines, nicotin, alcohol, spironolactone, interferon-gamma and metoclopramide.

Reduced cortisol levels are typical of the following diagnoses: Addison’s disease, adrenal insufficiency, adrenogenital syndrome (congenital adrenal hyperplasia), congenital adrenal hypoplasia, chromophobe adenoma, craniopharyngioma, hypophysectomy, hypopituitarism, hypothyroidism, hepatopathy, rheumatoid arthritis, Waterhouse-Friedrichsen syndrome, and following administration of dexamethasone and derivative preparations, ketoconazole, morphine, the inhalation of glucocorticoids for asthma bronchiale treatment and grapefruit juice ingestion.

14.2.2. Urinary Free Cortisol

Free cortisol is unconjugated cortisol filtered through the glomeruli to the urine. Free cortisol represents only about 5% of the total circulating cortisol, but the amount filtered to the urine accurately follows the secretory pattern of the adrenal cortex. This test is important because 24-hour urine collection eliminates the influence of diurnal variations. The main indication for the assay is Cushing's syndrome and the differential diagnosis of obesity; the assay is not suitable for a diagnosis of Addison's disease. Cautious interpretation is required in patients with diminished glomerular filtration (renal insufficiency). The patient should not be subject to physical strain or stress. Urine should be collected in a glass or plastic container (refrigerated container), and urine preservation with boric acid (10 g in a 3-litre container) or acetic acid (33%) is advisable. If the patient is catheterized, the collection bag must be placed on ice and emptied into a refrigerated container every hour. The reference range depends on the method. The table shows guidance values only:

Children 1-12 years	6 - 74 nmol/24 hrs
Both sexes 12-60 years	6 - 138 nmol/24 hrs

Conversion factor: ug/24 hrs x 2.76 = nmol/24 hrs

Table 14.2. The reference ranges for Urinary Free Cortisol

14.2.3. Hydroxyprogesterone

17-hydroxyprogesterone is derived from progesterone and is a metabolic precursor of 11-deoxycortisol in the pathway for the synthesis of cortisol. It has a pronounced diurnal rhythm (highest values in the morning). Elevated levels are usually due to steroid 21-hydoxylase or 11-hydroxylase deficiency. 17-hydroxyprogesterone is converted and excreted as pregnantriol. The test is made on the serum or plasma, or a dried blood spot for neonatal screening. Collection should be made in the morning, women at the follicular phase. Very young children, particularly infants under 2-3 months, have very high 17 OH-pregnenolone sulphate levels. Samples from infants should be tested after extraction. The test is used especially for the diagnosis of congenital adrenal hyperplasia. It is further used for the ACTH stimulation test and to monitor patients on glucocorticoid substitution therapy. Elevated 17-hydroxyprogesterone is tested to prove diagnoses such as: Antley-Bixler syndrome, congenital adrenal hyperplasia (levels 120 – 700 nmol/l), germinoma, hirsutism, ovarian cysts and tumours, polycystic ovary syndrome, virilization, steroids and metformin. On the other hand, reduced levels are typical of congenital adrenal hypoplasia, male pseudohermaphroditism (17 beta-hydroxylase deficiency) and Addison's disease. The reference range depends on the method. The table shows guidance values only:

Cord blood	27 – 150 nmol/l		
0 – 6 weeks	0.2 – 21 nmol/l		
6 weeks – 1 year	0.2 – 9.1 nmol/l		
1 – 10 years	0.1 – 2.7 nmol/l		
10 – 15 years	0.6 – 8.0 nmol/l		
over 15 years	Males	Females	
	1.5 – 7.2 nmol/l	0.4 – 9.4 nmol/l	
		Follicular phase	0.45 – 3.3 nmol/l
		Luteal phase	2.1 – 9.4 nmol/l
		Pregnant	6 – 36 nmol/l
		Post-menopausal	0.24 – 3.9 nmol/l

Conversion factor: ng/dl x 0.03 = nmol/l

Table 14.3. The reference ranges for 17-Hydroxyprogesterone

14.2.4. ACTH, Aldosterone, Renin

The ACTH test is intended to distinguish between a primary and secondary disorder of cortisol secretion. It is a special test made only by a few laboratories, and requires adherence to special pre-analytical conditions (e.g. sample delivered on ice and fast processing). Guidance reference limits are 5–41 ng/l (1–10 pmol/l). Tests for aldosterone and renin, which is closely related to hypoaldosteronism, for example, are dealt with in another chapter. Guidance reference limits for aldosterone are 0.08 – 0.61 mmol/l in the prone position, and 0.19 – 0.83 mmol/l in the sitting position.

14.3. Functional Tests

14.3.1. CRH Test

Indications: Cushing's syndrome – to distinguish between a primary (peripheral) and a secondary (central) origin. To determine cortisol and ACTH, 1 ml of anticoagulated blood with K2EDTA is taken (the patient must rest for at least 2 hours before sampling). Following the determination of basal values, 100 µg (human) CRH is administered intravenously. Samples should then be taken at intervals of 15, 30, 45 and 60 minutes. Caution – flush symptoms may occur! Constant cortisol and ACTH levels are indicative of pituitary ACTH deficiency. An excessive rise in cortisol and ACTH points to hypothalamic-pituitary Cushing's syndrome. (The elevated ACTH may persist if the hypothalamic cause has been present for a long time.) Elevated ACTH with normal cortisol levels is indicative of an adrenal tumour.

14.3.2. ACTH (Synacthen) Stimulation Test

This is a diagnostic test to evaluate the adrenal cortex function used for suspected cortisol production deficiency. The administration of exogenous ACTH stimulates the adrenal cortex. The test points to the secretory capacity of the adrenal cortex. Contraindications include fresh ulcerative diseases of the stomach and duodenum, decompensated diabetes mellitus, acute infections, acute psychosis or acute glaucoma.

Short test:

A sample of non-anticoagulated venous blood for a serum cortisol level assay is taken at 7:00 a.m. fasting, before drug administration (S-cortisol "before"). Then 0.25 mg of a synthetic ACTH analogue such as Synacthen inj. Novartis; 1 vial = 0.25 mg is applied i.v. The medication is administered under medical supervision due to the risk of anaphylactic shock. Additional samples of non-anticoagulated blood for the serum cortisol assay are taken 1 and 2 hours after the administration. Evaluation:

- An increase in S-cortisol over 690 nmol/l rules out the diagnosis of insufficiency;
- A very pronounced rise is indicative of Cushing's syndrome due to bilateral adrenal cortex hyperplasia;
- A normal or elevated response occurs in 50% of autonomous adrenal tumour cases; the other 50% of cases do not show any adrenal cortex response.

Prolonged test:

A sample of non-anticoagulated venous blood for a fasting serum cortisol assay is taken at 7:00 a.m. Thereupon 1 mg of ACTH-depot is administered intramuscularly, and additional samples of non-anticoagulated blood (2 ml) for the serum cortisol level assay are taken in 1, 2, 4, 8, 12 and 24 hours.

Evaluation:

- Normal response – more than 900 nmol/l (maximum in 8 hours);
- Insufficient response at all phases after stimulation is indicative of the primary form of adrenal failure;
- Insufficient or no response in the first hour, then a slow rise in cortisol level with a peak in 24 hours, is indicative of secondary adrenal insufficiency.

14.3.3. Metyrapone Test (Metopirone Test)

The test indicates pituitary ACTH reserve and is indicated in the case of suspected primary or secondary hypocorticism. Metyrapone inhibits 11-b-hydroxylase, thereby inducing the conversion of 11-deoxycortisol into cortisol and other metabolites. The decrease in cortisolaemia stimulates ACTH secretion. The increased ACTH level leads to an increased production of 11-deoxycortisol, which cannot be metabolized further, and its level rises.

Blood for the S-cortisol assay is taken at 8 a.m. Metyrapone (Metopirone 250 mg cps Novartis) is administered at 12 p.m.: 2 g to patients under 70 g of body weight, 2.5 g to patients from 70 to 90 kg and 3 g to patients over 90 kg. Sometimes it is advisable to do the test with a lower inhibition dose of 750 mg. Evaluation: Normal response – decrease in S-cortisol under 222 noml/l.

14.3.4. Dexamethasone Suppression Test

This is a diagnostic test of the suppressibility of the hypothalamic-pituitary-adrenal axis used for suspected hypercortisolism. The test determines the ability of the hypothalamic-pituitary system to decrease ACTH, and thereby also cortisol secretion following an increase in the amount of circulating glucocorticoids. Contraindications are the same as those for the ACTH test.

Standard test: The test starts by finding baseline levels of free cortisol in a sample from three 24-hour urine collections. Then 0.5 mg dexamethasone (Dexamethazon tbl, 1 tbl. = 0.5 mg) is administered orally for two days every 6 hours, starting at 6 a.m. (daily dose 2 mg). On the second day of administration, free cortisol is tested in the 24-hour urine collection. The test continues with an increased dose of 2 mg every 6 hours for the next two days (daily dose 8 mg). On the second day of this stronger blockade, free cortisol is once again tested in the 24-hour urine collection.

Evaluation:

Weaker blockade (2 mg): Free cortisol excretion drops below 60 nmol/day in healthy people. Patients with Cushing's syndrome do not show any decrease.

Stronger blockade (8 mg): Urinary free cortisol drops below 50% of the baseline concentration (average from three baseline values) in central Cushing's syndrome. There is no decrease in patients with an adrenocortical tumour producing cortisol, or with paraneoplastic Cushing's syndrome. Insufficient suppression of increased baseline values of urinary free cortisol at a weaker blockade is indicative of Cushing's sydrome, and a stronger blockade differentiates the central Cushing's syndrome from the other forms.

Fast (short, overnight) test: A 2-ml sample of non-anticoagulated venous blood for the fasting serum cortisol assay is taken before medication at 8 a.m. At 11 p.m. 1 mg (2 tbl) dexamethasone (Dexamethazon tbl.; 1 tbl. = 0.5 mg) is given p.o. The tablets are given with a small quantity of food. Dexamethasone is a synthetic fluorinated glucocorticoid with a slight mineralocorticoid effect; the strongest hypothalamic-pituitary-adrenal axis inhibition activity occurs if dexamethasone is administered at night. A single administration of 1 mg dexamethasone at night will sufficiently suppress ACTH secretion for 24 hours in most healthy people. The next day, once again at 8 a.m. fasting, 2 ml of non-anticoagulated venous blood for a serum cortisol level assay are taken before medication. Evaluation: S-cortisol will drop below 100 nmol/l in fasting healthy people in the morning. Patients with Cushing's syndrome will have insufficient suppression (over 100 nmol/l).

Stronger overnight dexamethasone test: 2 ml of non-anticoagulated blood are taken for the fasting serum cortisol assay at 8 a.m., and 8 mg dexamethasone is given at 11 p.m. A 2-ml sample is taken for the cortisolaemia on the next day at 8 a.m. Evaluation: A drop in cortisolaemia under 50% of the baseline concentration will occur in patients with central Cushing's syndrome. Cortisolaemia will not drop to this extent in the case of adrenocortical tumours producing cortisol or paraneoplastic Cushing's syndrome.

14.3.5. Renin-Aldosterone-Orthostatic Test

Indications for this test include a differential diagnosis of idiopathic primary hyperaldosteronism due to Conn's disease (trias: hypokalaemia, hypernatraemia, hypertonia), or proof of isolated hypoaldosteronism. It is not indicated if adrenal adenoma has been confirmed.

Test procedure: The patient rests in bed from 12 p.m., baseline aldosterone and renin values are determined in the prone position at 8 a.m., and the collection is repeated after 2 hours of orthostasis.

Evaluation: An increase in the aldosterone and renin levels of between 0.5 times and twice is normal.

14.3.6. Insulin Test

Hypoglycaemia is a strong stress impulse for ACTH and STH secretion. Blood for blood sugar, cortisol, ACTH and STH assays is taken in the morning. I.v. insulin is administered: 0.05 – 0.1 U/kg of body weight for suspected hypothalamic-pituitary insufficiency, 0.15 U/kg for an anticipated normal response, 0.20 – 0.30 U/kg in obese patients with Cushing's syndrome. Blood is taken in 30, 60, 90 and 120 minutes. Evaluation: Blood sugar in healthy people drops under 2.2 mmol/l and cortisol rises over 320 nmol/l, or reaches a value 1.5 times higher than the baseline value. ACTH reaches at least twice as high against the baseline value. Contraindications for the test include ICHS, cerebrovascular alterations, gestosis or epilepsy.

14.4. Hypercorticalism

An overproduction of glucocorticoids manifests itself with clinical signs known as Cushing's syndrome. The patients are obese, have hypertension and a loss of muscular tissue, they may sometimes have steroid diabetes (hyperglycaemia), hypernatraemia and hypokalaemia. Conn's syndrome is a consequence and sign of isolated overproduction of aldosterone.

The primary cause is usually an adrenal cortex adenoma or adenocarcinoma. The circadian rhythm is missing, and the ACTH concentration is low. Cortisol secretion is autonomous; secretion does not rise following stimulation (ACTH or insulin test) and does not decrease following administration of suppressants (dexamethasone).

The secondary (central) cause is adenohypophyseal adenoma with higher ACTH concentrations (Cushing's disease). The adrenal cortex reacts to functional tests as expected (stimulation, suppression). If CRH overproduction in the hypothalamus is the cause, the condition is referred to as hypercorticalism.

Ectopic ACTH production is relatively frequent in some tumour diseases such as small-cell or bronchogenic lung cancer, sometimes with concurrent ADH production. The metabolic effects of mineralocorticoids may sometimes dominate the clinical picture. Cortisol does not react in functional stimulation and suppression tests.

Glucocorticoid therapy (long term, high doses) leads to the development of Cushing's syndrome, i.e. hypokalaemia, hyperglycaemia, etc. CRH and ACTH production is suppressed and adrenal cortex atrophy occurs. Abrupt discontinuation may be life-threatening. The most efficient artificial glucocorticoids such as dexamethasone do not interfere with the cortisol assay.

14.5. Hypocorticalism

The cause of primary hypocorticalism is the destruction of the adrenal cortex. Conditions include autoimmune adrenalitis (assay for autoantibodies against steroid 21-hydroxylase) and infectious adrenalitis (tuberculosis, meningococcal infection, cytomegalovirus infection, HIV and other infections). Adrenal involvement in tumour metastases and congenital adrenal hypoplasia (AHC, DAX gene mutation or deletion, or microdeletion syndrome connected with glycerolkinase and the gene for dystrophin defect – DMD) in male patients are rare. If the glucocorticoid deficiency is accompanied by a mineralocorticoid deficiency, adrenal insufficiency (Addison's disease) will develop (weakness, hypotension, hyponatraemia, hyperkalaemia) that may precipitate an Addisonian crisis. An increase in the ACTH concentration is confirmatory. The cortisol level is decreased or normal. The adrenal cortex does not react to stimulation tests. Increased pigmentation is the consequence of increased POMC (pro-opiomelanocortin) production.

The cause of secondary hypocorticalism is usually the inability of the pituitary to produce ACTH or damage to the hypothalamus. These conditions occur after traumas (of the brain), bleeding, radiotherapy, hypophysectomy, neurosurgical interventions, if a tumour is growing from the other cells (prolactinoma), after bleeding into the pituitary following birth, or because of infection (tuberculosis). The ACTH level is low. The adrenal cortex reacts to the stimulation test with ACTH, but not to CRH or hypoglycaemia (insulin test). Concomitant deficient production of other adenohypophyseal hormones is common.

Adrenal insufficiency may also be caused by a long-term glucocorticoid therapy, which may lead to adrenal cortex atrophy. Abrupt discontinuation of therapy (or a dramatic dose reduction) exposes patients to risk of adrenal crisis.

Congenital adrenal hyperplasia (CAH) is a group of genetic diseases characterized by a hyperplastic growth of the adrenal cortex (influence of ACTH) and an impaired synthesis of adrenocortical hormones as a result of enzymatic defects. Clinical signs are the consequence of deficient cortisol, or aldosterone in some cases, and excessive precursors synthesized prior to the defective enzymatic reaction. This is the most frequent autosomal recessive genetic disease (1:100) with a wide spectrum of phenotype signs, from life-threatening salt disorders (hyponatraemia, hyperkalaemia) with severe external genitalia malformations in newborns, to less conspicuous forms of hirsutism or virilization at an older age.

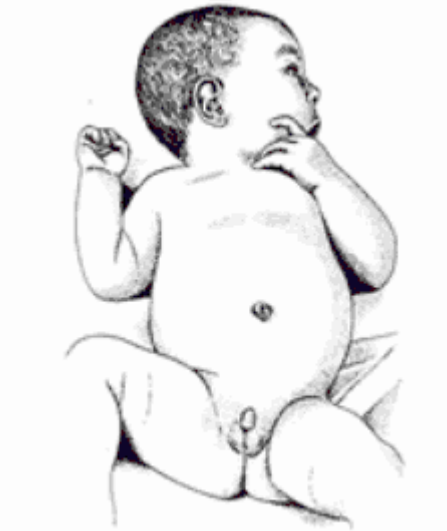


Figure 14.2. A newborn baby with CAH

Cornelius

27 June 1670 - 19 July 1670

"On 27 June 1670, a baby was born in Delft. Due to the malformation of the sex organs, the parents did not want to decide on a girl's or boy's name without expert advice. Experts concluded that it was a boy from the apparently perfect penis and a formation very similar to a scrotum. Therefore, the child was named Cornelius. Overcome by a serious disease, Cornelius departed this life on 19 July. The autopsy revealed, however, that Cornelius was a girl with an oversized clitoris."

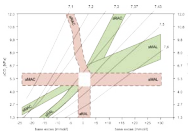
Regnier de Graaf (1672):

De mulierum organis generationi inservientibus tractatus novus

Laboratory findings of the severest forms are dominated by high ACTH, renin and 17-hydroxyprogesterone levels, hyponatraemia and hyperkalaemia.

Neonatal screening for CAH is based on the 17-hydroxyprogesterone assay on a dried blood spot. It is important for early corticoid substitution therapy. A prenatal diagnosis enables the use of intrauterine treatment that prevents, in particular, female foetus virilization. The commonest cause of CAH is mutations in the gene for steroid 21-hydroxylase (about 90% of cases).

Another cause of CAH (about 5%) is a deficiency of the mitochondrial enzyme 11-beta-hydroxylase. A 17-alpha--hydroxylase/17,20-lyase deficiency is special (it is one gene, one enzyme, but two enzyme activities dependent on post-translational modification). This enzyme deficiency is usually present in the adrenal gland and gonads and manifests itself in CAH, hypertension, Biglieri syndrome in females, and New syndrome in males. Isolated 17,20-lyase deficiency has been observed in rare cases. Another enzyme the deficiency of which causes classical forms of CAH with salt losses is 3-beta-hydroxysteroid dehydrogenase. A final cause of CAH can be defects in the StAR protein (steroidogenic acute regulatory protein) co-responsible for the mitochondrial conversion of cholesterol to pregnenolone (this disease is sometimes referred to as lipoid congenital adrenal hyperplasia).



CHAPTER 15

15. Disorders of Acid-Base Balance

15.1. Metabolic and Respiratory Disorders of Acid-Base Balance

Authors: prof. MUDr. Jaroslav Racek, DrSc., MUDr. Daniel Rajdl, Ph.D.

Reviewer: prof. MUDr. Antonín Kazda, DrSc., MUDr. Helena Brodská, Ph.D.

Acid-base balance (ABB) disorders are divided into two groups, depending on their cause: Metabolic and respiratory; acidosis and alkalosis is distinguished based on whether acids or bases are prevalent. From the above it follows that four main groups of ABB disorders can be distinguished:

- Metabolic acidosis (MAc);
- Metabolic alkalosis (MAI);
- Respiratory acidosis (RAc);
- Respiratory alkalosis (RAI).

The next paragraphs examine these basic disorders, taking note of their causes, clinical signs, laboratory alterations and means of compensation. Methods for their treatment will be mentioned in brief.

15.1.1. Metabolic Acid-Base Disorders

Metabolic acid-base disorders are characterized in laboratory results by changes in hydrogen carbonate concentration. They are the commonest of all acid-base balance disorders.

15.1.1.1. Metabolic Acidosis

Metabolic acidosis (MAc) can be due to the accumulation of acidic metabolites or loss of hydrogen carbonates.

The clinical presentation of MAc is typified by deep acidotic Kussmaul breathing caused by stimulation of the respiratory centre by a high concentration of hydrogen ions. The drop in cerebrospinal fluid pH is manifested in headaches, vision disturbances, disorientation, somnolence or even coma. GIT manifestations take the form of nausea and vomiting.

In laboratory assays, MAc is manifested by negative BE value, reduced HCO_3^-, a drop in pH and later, as a result of lung compensation of the defect, by a drop in p_aCO_2. The low p_aCO_2 leads to mental state alterations or even coma. Urine is usually acidic, and the level of excretion of phosphates and ammonium ions is higher. HCO_3^- excretion decreases.

Metabolic Acidosis with Elevated AG Value

This type of MAc may be due to an accumulation of acidic endogenous metabolites or an intake of acids or substances transformed into acids in the body. This occurs in:

- *Ketoacidosis* found in decompensated diabetics, during fasting, in alcoholics, following prolonged physical strain;
- *Lactate* (*lactic*) *acidosis* due to, for example, insufficient oxygenation of the blood and perfusion of tissues when lactic acid accumulates in the body; similarly as in the case of a disorder affecting the elimination of this relatively strong acid – see below;
- *Renal acidosis* in patients with renal failure when endogenous anions (phosphates and sulphates) accumulate;

- Accumulation of exogenous anions in *intoxications* where the toxic substance is an acid itself (e.g. salicylic acid) or it is metabolized to an acidic catabolite (lactate as a product of ethanol metabolism, glyoxylic and oxalic acids produced from ethylene glycol, formic acid from methanol, etc.);
- *Inadequate parenteral nutrition*, for example, following the fast administration of sorbitol, xylitol, fructose or protein hydrolysates; this is the reason that carbohydrates other than glucose are practically not used today in parenteral nutrition.

Lactic acidosis is probably the most frequent of the disorders listed above. It can be divided into types A and B based on the cause.

Type A lactic acidosis is connected with *insufficient blood oxygenation and tissue hypoxia* – hypoxic, anaemic, stagnant or histotoxic hypoxia.

Hypoxic hypoxia may be caused by low pO_2 in the inhaled air, hypoventilation, pulmonary diseases connected with a limitation of the respiratory area, alveo - capillary blocks, etc.

Anaemic hypoxia is caused by a red blood cell or haemoglobin defect, or anaemia.

The cause of *stagnant hypoxia* can be the failure of the heart's function as a pump, thrombosis, sepsis, a decrease in the circulating blood volume and even shock.

The cause of *histotoxic hypoxia* is usually damage to enzyme cell systems by radiation, heat, infection, hypoxia, carbon monoxide or cyanide poisoning, etc.

Type B lactic acidosis *is not accompanied by tissue hypoxia*. It is caused by impaired lactate elimination in hepatic and renal diseases, leukaemia, oral antidiabetic treatment (biguanide type – metformin), the fast administration of fructose, sorbitol or xylitol, in congenital metabolism disorders such as type I glycogenosis, methylmalonic aciduria etc. Lactate elimination disorder may also be due to insufficient supply of thiamine (vitamin B1) which is needed for the oxidative decarboxylation of α-keto acids (and pyruvic acid as well). In reality, however, both lactic acidosis types are often combined – for example, a rise in the lactate concentration from hypoxia in a patient in shock is potentiated by a lactate elimination disorder in the hypoxic liver.

A useful parameter for evaluating hypoxia in a certain organ is the determination of *the difference in arterial and venous lactate levels*. And finally, another option for hypoxia evaluation is establishing the *lactate/pyruvate ratio*, which varies with the supply of oxygen to tissues. Since glycolysis takes place in the cytoplasm, it is sometimes recommended to use a similar *β-hydroxybutyrate/acetoacetate ratio* expressing the redox processes in mitochondria.

Lactic acidosis can also be seen physiologically when muscles work with an "oxygen debt". This occurs during all sprinting-type physical activities, whereas extended physical exertion (endurance) must be performed in an aerobic mode.

The degree of physical training can be evaluated from the **anaerobic threshold**. Bicycle ergometrics compprises a gradual increase of physical exertion while the lactate level is monitored at regular intervals. Although the level increases slightly at first, a turning point then occurs after which the lactate concentration starts growing dramatically; the value of exertion at the turning point is called the anaerobic threshold (Figure 1).

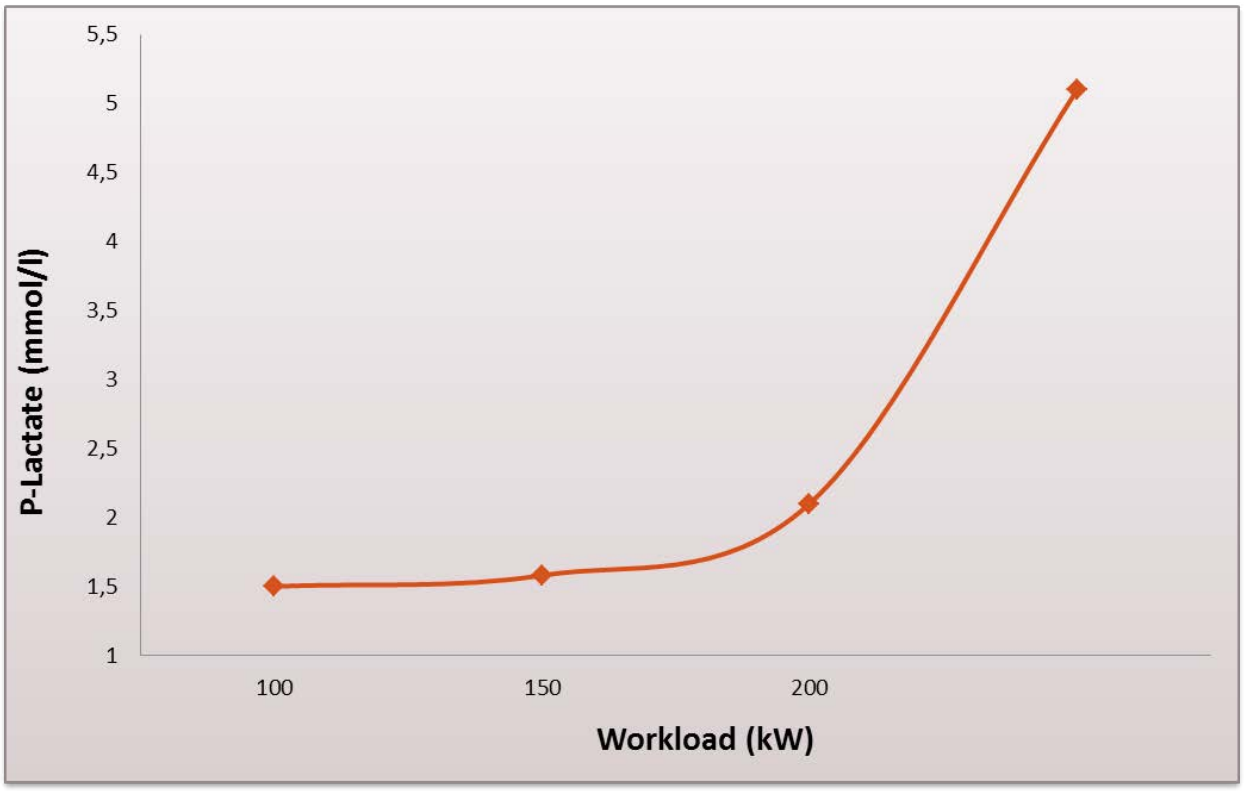


Figure 15.1. Determination of the anaerobic threshold

The body is unable to do any more intensive exercise in the aerobic mode. Regular training can move the anaerobic threshold up to higher values – an individual can achieve better performance aerobically. This is important not only for sportspeople but also in patients after myocardial infarction. It is true, however, that the ability to cope with strain in the aerobic or anaerobic mode is to some degree innate, and is given by the ratio of white (anaerobic) to red (aerobic) muscle fibres.

Metabolic acidosis with normal AG value

This is caused by excessive loss of *hydrogen carbonates*. Such losses occur in:

- Severe diarrhoeas, losses of intestine content through GIT drains and fistulas distal to the duodenum;
- Hypoaldosteronism;
- Ureterosigmoidostomy (intestinal mucosa absorbs Cl^-);
- Type 1 (distal) renal tubular acidoses, where distal tubule cells are not able to secrete H^+);
- After NH_4Cl and arginine hydrochloride administration in treatment for MAI;
- After acetazolamide (carbonic anhydrase inhibitor) administration;
- In dilutional acidosis with a relative loss of hydrogen carbonates following infusions lacking the buffer system (isotonic NaCl solution); the concentration of the bicarbonate buffer's acid component is normal.

Metabolic acidosis is *compensated* by hyperventilation leading to a decrease in p_aCO_2, and therefore an increase in pH. The maximum degree of compensation is reached 12 – 24 hrs from the occurrence of the disorder, and if the disorder is removed, the compensation persists for the same time. The compensation effect depends on respiratory apparatus efficiency: while it may be very effective in a young, otherwise healthy person, the effect will be limited in a person with severe emphysema.

The guiding principle when *treating MAc* (as well as other ABB disturbances) is that the *underlying disease* should be *treated first*: hypoxia by oxygen therapy, diabetic ketoacidosis by insulin, fasting ketoacidosis by glucose supply, etc. *Alkali treatment* should be used only in exceptional cases; the objective is to adjust the pH not the BE value.

Usually the 8.4% $NaHCO_3$ or 10% $KHCO_3$ solutions for coincident hypokalaemia are used; both solutions contain 1 mmol HCO_3^- in 1 mL. Doses do not usually exceed 50 mL (20 mL of the solution are often sufficient). Administration of the $NaHCO_3$ solution has its limitations too. The $NaHCO_3$ solution is suitable only for non-hypoxic types of metabolic

acidosis (uraemia, diabetes, diarrhoea). In lactic acidoses, it leads to an overproduction of CO_2 in the ECF with the development of respiratory acidosis; at the same time, as a result of proton accumulation in cells, cell metabolism becomes impaired (further decrease in intracellular pH).

The sodium lactate solution is suitable only for acidoses with a good liver function, without the risk of lactic acid retention.

Organic acids (lactic acid, acetoacetic acid, β-hydroxybutyric acid) are metabolized to bicarbonate when the cause of their accumulation has been removed. Prior administration of $NaHCO_3$ could therefore lead to a shift in blood pH to the alkaline side with all the disadvantages of alkalaemia.

Many unexplained lactic acidoses are caused by the depletion of thiamine (vitamin B_1), needed for oxidative decarboxylation and also for aerobic glycolysis and the proper function of the Krebs cycle. This may occur in patients on long-term parenteral nutrition. The disturbance will quickly disappear following administration of vitamin B_1.

15.1.1.2. Metabolic Alkalosis

This serious disorder is divided into two groups depending on the cause and method of treatment.

Metabolic alkalosis responding to chloride treatment

This type of MAI is caused by excessive loss of chlorides, either in the acidic gastric juice or urine. Such losses occur in:

- Loss of gastric juice;
- Prolonged vomiting (considerable amounts of Na^+ and K^+ are lost);
- Protracted diarrhoeas with a loss of chlorides (chloridorrhoea);
- Overdose of alkaligenous diuretics (furosemide); Cl^- losses are greater than Na^+ losses; in addition, a considerable amount of K^+ is excreted in the urine.

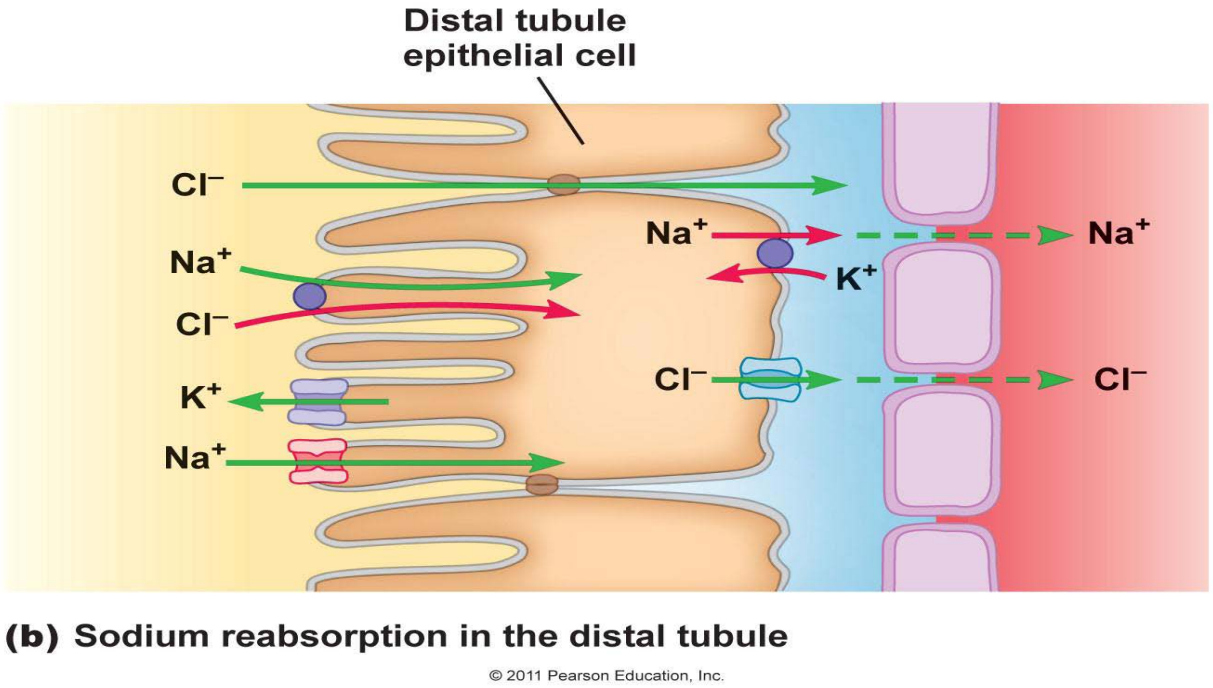


Figure 15.2. The pathogenesis of this MAI type involves a lack of chlorides in the proximal renal tubule, where they are normally absorbed together with Na^+. Consequently, most Na^+ ions come to the distal tubule where they are exchanged for K^+ and H^+: Characteristic for this type is a very low concentration of chlorides in the urine (U-Cl^- < 20 mmol/L).

Metabolic alkalosis resistant to chloride treatment

Causes of this condition are diverse and may include:

- Primary hyperaldosteronism (Conn's syndrome);
- Secondary hyperaldosteronism (in nephrotic syndrome, cardiac insufficiency, chronic hepatopathy or renal artery stenosis);
- Prolonged administration of corticoids with manifested mineralocorticoid effect;
- Excessive administration of antacids such as $NaHCO_3$;
- Catabolism with K^+ discharge from cells in exchange for H^+ and Na^+; the result is ICF acidosis and ECF alkalosis;
- Fast correction of ketoacidosis when ketone bodies are metabolized to hydrogen carbonates;
- Dehydration with a concentration of the internal environment, a relative increase in HCO_3^- and secondary hyperaldosteronism (concentration alkalosis).

Na^+ is absorbed in the distal renal tubule and excessive amounts of H^+ and K^+ are excreted in hyperaldosteronism, which is the commonest cause of this type of MAI; HCO_3^- is retained at the same time. The chloride concentration in urine, however, is usually higher than 20 mmol/L in this case.

In alkalaemia, the concentration of ionized calcium in the plasma decreases and neuromuscular irritability increases, while the tendency to spasms is increased. The drop in K^+ in the plasma may lead to heart rhythm disorders. The release of oxygen from its bond to haemoglobin worsens and the haemoglobin dissociation curve shifts to the left. Additional signs are anoxia, vomiting, subileus, muscular weakness, paresthaesia and confusion.

Acidic urine reaction sometimes occurs in MAI – this is referred to as *paradoxical aciduria*. This is caused by hypokalaemia, which often occurs with metabolic alkalosis: due to shortage of K^+, Na^+ has to be replaced with H^+ in the distal renal tubule. The result is acidic urine formation. Since HCO_3^- ions are absorbed along with Na^+, metabolic alkalosis is intensified. *Circulus vitiosus* should be interrupted by the supply of K^+ (usually in the form of KCl, which also acidifies the internal environment).

The respiratory compensation of metabolic alkalosis (hypoventilation) is *imperfect*, *and leads not only to hypercapnia but also to hypoxaemia.* A drop in p_aO_2 under 8.0 kPa is registered by chemoreceptors with subsequent stimulation of the respiratory centre. This mechanism does not allow p_aCO_2 to rise over 8.0 kPa even if the disorder lasts for several days. An exception is those conditions when the last stimulus provoking escalation of the respiratory centre function is removed by increased p_aO_2 during oxygen therapy.

While metabolic acidosis (due to fasting, physical strain, etc.) has been common in the course of human evolution, and humans can tolerate and compensate it quite well, metabolic alkalosis is compensated and thereby tolerated much worse; this underlines the need to distinguish the risk of alkalosis in time.

In laboratory results, MAI will manifest itself by an increase in HCO_3^-, BE and pH, and a compensatory increase in p_aCO_2. Hypokalaemia and hypochloridaemia are concurrent; the ratio of ionized calcium to the total is reduced. The Cl^- concentration in the urine varies depending on the cause of MAI (see above). As mentioned before, urine may be acidic in paradoxical aciduria.

Treatment for MAI depends on the type of disorder. Arginine hydrochloride is administered for MAI due to chloride loss; NaCl and KCl solutions have lower acidification ability. Another option is to administer 0.1 mol/L hydrochloric acid using a slow infusion into the central vein. This type of MAI should not be considered cured until normalization of chloride losses in the urine occurs. When treating the second type of MAI, the objective is to affect the primary cause; lacking potassium is usually replenished in the form of KCl.

Threatening disorders may be optionally treated by haemodialysis, a fast and effective method for adjusting the metabolic ABB component.

When treating metabolic alkalosis and other acid-base disorders one should consider the dependence of K^+ concentration in the extracellular fluid on the pH, and avoid violent kalaemia fluctuations with potential cardiac arrhythmia as a result of changes in pH.

15.1.2. Respiratory Acid-Base Disorders

These disorders are characterized by primary changes in p_aCO_2.

15.1.2.1. Respiratory Acidosis

Respiratory acidosis (RAc) is relatively rare in patients receiving intensive care, unless the acute disease or injury is preceded by chronic pulmonary disease.

Causes of respiratory acidosis may include:

- *Central respiratory depression* (injuries, inflammations, tumours in the region of the respiratory centre or as an effect of tranquilizers or hypnotic drugs, sleep apnoea syndrome, etc.);
- *Ventilation disorder* (chronic obstructive pulmonary disease, pulmonary oedema, extensive pneumonia, pneumothorax, extensive hydrothorax, status asthmaticus, chest deformities, chest injuries with breathing hampered by pain, extreme obesity [Pickwickian syndrome], etc.);
- *Cardiac and respiratory arrest* where RAc is combined with metabolic lactic acidosis.

Laboratory results will show RAc as an increase in p_aCO_2 and a decrease in pH; decreased p_aO_2 with increased lactate concentration are indicative of RAc combined with MAc. Increased p_aCO_2 (hypercapnia) normally forces the respiratory centre to increase respiratory activity. If p_aCO_2 exceeds 8.0 kPa, the respiratory centre receptivity to this stimulus decreases, and low p_aO_2 (hypoxaemia) remains the only stimulus. K^+ and Ca^{2+} concentrations in the blood increase. The haemoglobin dissociation curve shifts to the right. Vasodilatation occurs in the peripheries.

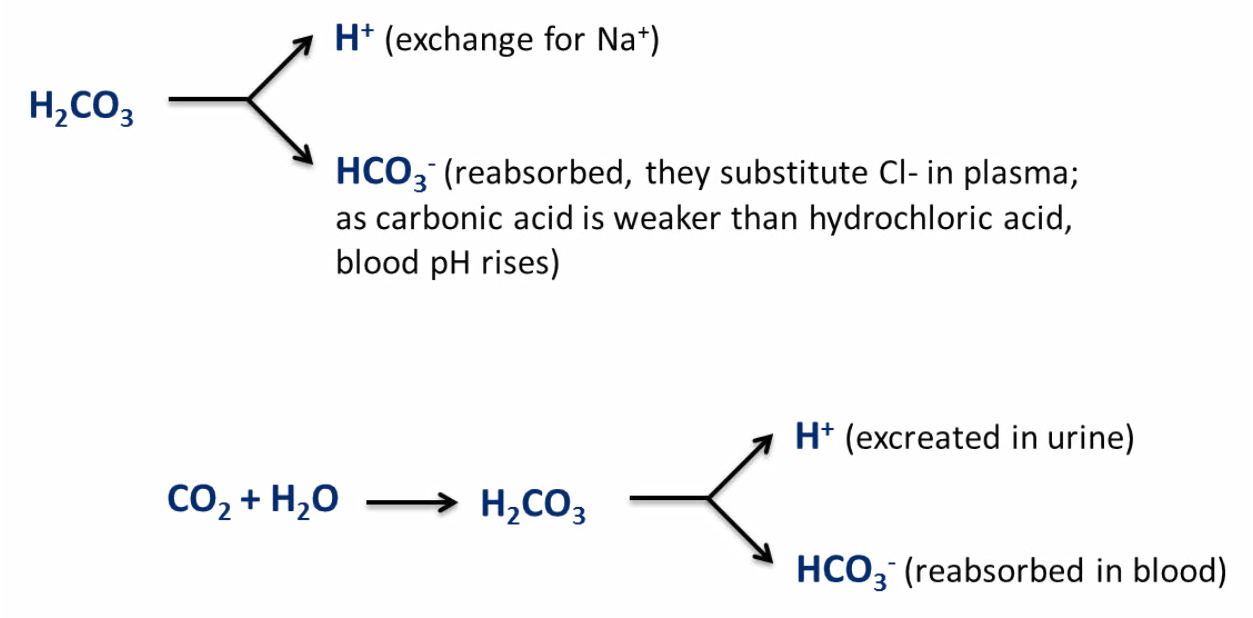


Figure 15.3. RAc compensation takes place in the kidneys. The exchange of Na^+ for surplus H^+ is accelerated in the tubules. H^+ and Cl^- losses increase, while Na^+ is retained. Hydrogen carbonates are also retained and their concentration in the blood increases.

Clinical signs of RAc include neurological symptoms of acidaemia and cerebral vasodilatation – headache, tremor, spasms, motor skills disorders, psychoses, sleep disorders, somnolence, fatigue or even coma.

The *therapy* for acute conditions of RAc consists in the removal of the cause in order to improve alveolar ventilation. Medication can also be used (bronchodilatants, sympathomimetics). When using oxygen therapy for chronic conditions, special caution should be exercised to ensure against removing the last stimulus, hypoxaemia, for the respiratory centre through a rapid increase in p_aO_2. Chronic disorders are compensated by an increased concentration of hydrogen carbonates; if hypercapnia is removed by controlled ventilation, renal compensation will persist for many days, which would put the patient at risk of alkalaemia with all its adverse effects. Therefore, the rate of decrease in p_aCO_2 should not exceed 0.3 – 0.7 kPa/hr during controlled ventilation.

Respiratory Insufficiency

The condition when the lungs are unable to ensure sufficient exchange of oxygen and carbon dioxide between atmospheric air and blood is referred to as *respiratory insufficiency*. In terms of laboratory results, it is characterized by pathological levels of blood gases (p_aO_2 and p_aCO_2).

A fall in p_aO_2 and haemoglobin saturation with oxygen is typical of *partial respiratory insufficiency (hypoxaemic*

failure). Initially it can be only found following exercise, later even in resting patients.

The laboratory picture of *global respiratory insufficiency (hypercapnic failure)* also shows a rise in p_aCO_2, i.e. respiratory acidosis. The later onset of hypercapnia is caused by the properties of carbon dioxide, which diffuses 30 times faster than oxygen.

Causes of respiratory insufficiency may be at the level of the respiratory centre, lungs and airways, or it involves a chest wall disease with the same causes as those of respiratory acidosis.

Hypoxaemia supports anaerobic tissue metabolism with lactate accumulation and the development of metabolic *lactic acidosis*.

The clinical presentation of respiratory insufficiency is given by the underlying disease and the combined effect of hypoxaemia, hypercapnia and RAc. A drop in p_aO_2 under 6.7 kPa may lead to a manifestation of *latent perfusion disorders* such as angina pectoris, intermittent claudication, transient ischemic attacks in the CNS, etc. The patient suffers from headaches, depression, anxiety, somnolence or, conversely, motor restlessness. Hypercapnia causes vasodilatation and increased permeability of cerebral veins, short naps, insomnia at night, and somnolence during the day.

15.1.2.2. Respiratory Alkalosis

Respiratory alkalosis (RAl) develops in conditions accompanied by respiratory centre irritation with consequent hyperventilation. Respiratory centre irritation occurs in the following situations:

- Voluntary hyperventilation;
- Stress, and in hysterical patients;
- Respiratory centre disorders (ictus, trauma, tumour, inflammation);
- Metabolic disorders (fever, hepatic encephalopathy, effect of drugs – salicylates);
- Inappropriately set controlled ventilation parameters; a slight RAl may be deliberately induced in controlled ventilation to prevent cerebral oedema (a decrease in p_aCO_2 causes cerebral vasoconstriction with decreased cerebral blood flow).

Respiratory centre irritation may also be caused by local acidosis in anaemia or cardiac failure, for example.

RAl is manifested in *laboratory results* as a fall in p_aCO_2 and a rise in pH. Compensation takes place in the kidneys: decreased exchange of Na^+ for H^+ in renal tubules, hydrogen carbonates are lost, increased H^+ and Cl^- levels in the blood (Cl^- ions replace the missing hydrogen carbonates in the anion column).

The *clinical presentation of RAl* is characterized by paraesthaesia, sweating, dizziness, weakness, fatigue and disorders of consciousness (cerebral perfusion and function are considerably limited at p_aCO_2 of 2.7 – 3.4 kPa). A decrease in the ratio of ionized calcium to the total calcium in plasma may occur with proneness to tetany.

The *treatment for RAl* should once again focus on the cause of the disorders. In psychogenic RAl, the patient should be sedated and the dead space in the lungs extended with a tube or bag so that the patient inhales part of the already exhaled carbon dioxide. The objective of controlled ventilation is to increase p_aCO_2 gradually up to an optimum value of about 4.7 kPa.

Respiratory alkalosis may be the first sign of hypoxaemia (in latent cardiac failure, successive pulmonary artery embolism, in patients with severe anaemia, etc.).

15.1.3. Acid-Base Compensation (Diagnostic) Diagram

The acid-base compensation diagram allows for a fast, approximate evaluation of the acid-base balance status, and a classification of the disorder depending on p_aCO_2 and BE. In addition, the diagram shows the estimated development of the disorder during compensation.

This only provides an initial outline. Patient history, clinical signs and the chronological characteristics of the disease, the results of assays for basic minerals and other analytes in the serum or urine (lactate, ketone bodies) and other information, not only the results of laboratory examinations, must be known in order to establish an accurate diagnosis.

Figure 4 shows the acid-base compensation diagram. As is evident from the diagram, we can recognize within it not only acute and compensated acid-base disorders, but also mixed disorders; one exception to the latter group is combined MAc and MAl.

The involvement of metabolic disorders is usually assessed on the basis of the **anion gap** (AG). It can be calculated accordint to the formula: AG = $(Na^+ + K^+) – (Cl^- + HCO_3^-)$. AG usually ranges from 10 to 20 mmol/L, 15 mmol/L on average. Elevated AG values are usually in MAc due to the accumulation of anions other than chlorides, and reduced AG values in MAl are due to chloride losses.

Stewart and Fencl's methods may also be used to assess MAc and MAl combinations. This method uses the fact that plasma cation and anion columns must be of equal height. These authors distinguish strong ions such as Na^+, K^+, Ca^{2+} and Mg^{2+} cations and the Cl^- anion. Their difference is referred to as **SID (strong ion difference)** = $(Na^+ + K^+ + Ca^{2+} + Mg^{2+}) – Cl^-$. Also involved are weak anions such as albumin and inorganic phosphate, and "unmeasured" anions such as lactate, β-hydroxybutyrate, anions produced by alcohol oxidation in intoxication, and, finally, bicarbonate, which plays a special role: it may be formed from CO_2 in the body, or, in contrast, be decomposed and replaced with stronger acid anions, hence its being regarded as a dependent variable. It is the evaluation of this ion ratio that enables us to estimate the role of each metabolic disorder in acid-base alterations.

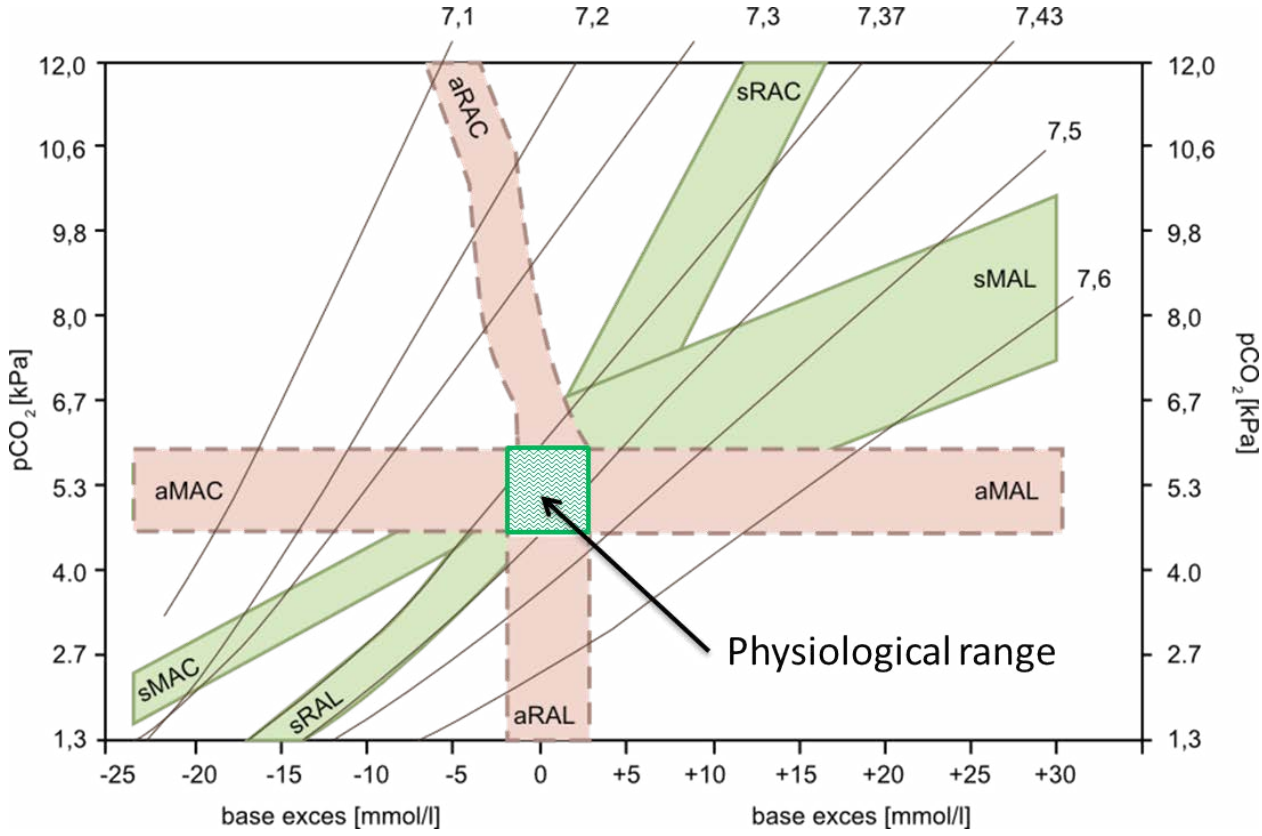


Figure 15.4. Acid-base compensation (diagnostic) diagram; the green area with waves shows the range of physiological values; indexes in disorder abbreviations: a = acute, s = stabilized (maximum compensation)

15.1.4. Acid-Base Balance (and Blood Gases) Measurement

15.1.4.1. Collection of Blood

Arterial blood, usually taken from a. brachialis or a. radialis, is best suited for partial blood gas pressure determination. Collection is made using a syringe with a plunger (Figure 5) or a special capillary collection system filled by the free flow of blood until the system is completely full.

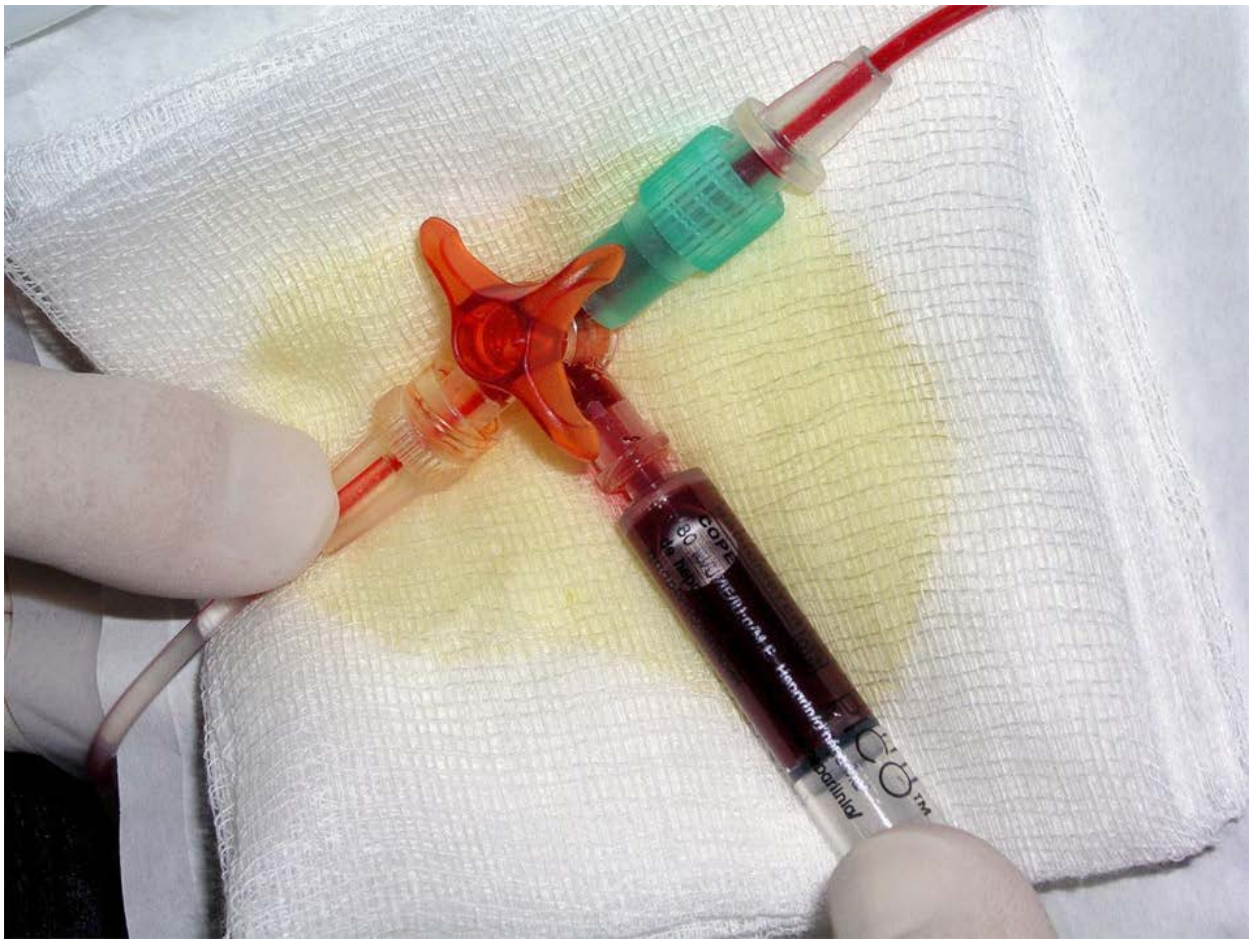


Figure 15.5. Arterial blood collection for blood gas assay

For repeated sampling or acid-base level determination, *arterialized blood* is usually obtained from a hyperaemized earlobe, i.e. after it has been thoroughly warmed (heated in hot water or using hyperaemizing ointment). This will accelerate blood flow through the capillaries and the blood composition will be similar to that of arterial blood. In this case, blood is taken into a glass capillary tube, which is sealed with a plastic cap when full. Fingertip blood is less suitable. This is due to insufficient perfusion of peripheral parts of the body where pO_2 is 2 – 3 kPa lower than pO_2 in arterial blood. Blood taken from the earlobe is not suitable during blood circulation centralization or oxygen inhalation treatment.

All collection syringes or capillary tubes contain *heparin as an anticoagulant*, which is the reason blood should be thoroughly mixed after collection. Since many analyzers measure other analytes (minerals, lactate, glucose, urea, creatinine, etc.) in addition to pH and partial blood gas pressure, it is advisable to use collection syringes containing lithium salt of heparin which is saturated with calcium to avoid binding calcium and magnesium from the blood. A metal agitator (wire) is inserted into glass collection capillary tubes before sealing, and blood is mixed using a magnet applied to the outer surface of the capillary. Care should be taken to always ensure *anaerobic collection*. Blood must not flow down the finger, and no air bubble must be entrained in the syringe or capillary tube. Air would completely impair the oxygen assay; an air bubble is known to increase pO_2 by up to several kPa! The syringe with the blood must be hermetically sealed.

Following collection, the syringe or capillary tube should be placed horizontally and the analysis should be made as soon as possible. Samples should be transported on melting ice, which prevents any decrease in pO_2 that would otherwise occur due to metabolic processes in the collected blood (oxygen is consumed by white blood cells and thrombocytes, not red blood cells). The test should be made within 2 hours of collection, even when the blood is treated as described above.

15.1.4.2. Acid-Base Balance and Blood Gases Measurement Procedure

Analyzers take measurements of acid-base balance and blood gases in a cell heated to 37°C; measured in this way are pH, p_aCO_2 and always p_aO_2. In addition to pH and partial blood gas pressure, these analyzers also offer all other

related calculations, such as HCO^{3-} and BE, that characterize the acid-base balance and oxygen metabolism. These instruments also measure haemoglobin concentration, and many of them measure oxygen saturation using the direct method (Figure 6) as well as the content of other haemoglobin derivatives. The instruments are sold by many manufacturers such as Radiometer, Roche, Instrumentation Laboratory, Nova Biomedical, Siemens, etc.

Special instruments measuring blood gases, acid-base balance and other acute parameters are designed to enable their use right at the patient's bedside in intensive care units, departments of anaesthesiology and resuscitation, and operating theatres. Their small size ensures portability and ease of operation allows assays to be performed by trained clinical personnel. Examples: Instrumentation Laboratory, Roche, Radiometer or Abbott analyzers.

Figure 15.6. Measurement of blood gases, acid-base balance and some other analytes on the analyzer

15.2. Combined Metabolic Acid-Base Balance Disorders

Summary:

Combined acid-base balance disorders are caused by a combination of several simple disorders. Examples of common combinations:

- Vomiting (loss of chlorides → metabolic alkalosis) + fasting (accumulation of ketone bodies → metabolic acidosis)
- Diabetic ketoacidosis + dehydration (poor tissue perfusion → lactate formation → metabolic acidosis)
- Salicylate intoxication (organic acid → metabolic acidosis + respiratory centre irritation by salicylate → respiratory alkalosis)
- Respiratory insufficiency with hypercapnia (e.g. respiratory acidosis in COPD - Chronic obstructive pulmonary disease) is also connected with hypoxia and lactic acidosis

The treatment for combined disorders focuses on removing the cause (all of the causes if there are more than one); the target laboratory parameter is pH (not BE).

If *two or more simple disorders* are present at the same time, the condition is referred to as a **combined (mixed) acid-base balance disorder**. The disorders either develop independently of each other, or one disorder conditions the

other, which, however, *never develops in order to compensate the former*.

The disorders may counteract one another (a combination of acidosis and alkalosis), so the resulting pH does not deviate much from the norm and blood pH and other ABB parameters may even be in the physiological range. In other cases, both disorders shift the pH value in the same direction and potentiate each other. The paragraphs below describe some examples of combined disorders.

15.2.1. Combination of Counteracting Disorders

15.2.1.1. Combined MAc and MAl

The patient vomits and develops *hypochloraemic MAl*. Since the patient is not taking in any food, another disorder, metabolic ketoacidosis, gradually develops as a result of fasting. One disorder may cancel out the other, so acid-base assays do not show any pathological deviation. In such instance it is very important to examine the basic ion concentration in the serum (*Figure 16.1*).

The first column represents anions in healthy individual's serum. The second column shows changes caused only by the loss of chlorides through vomiting. In the third column accumulating ketone bodies (included within unmeasured anions) will cause a decrease in HCO_3^- concentration to the normal level. The necessity to consider patient history and the clinical presentation of the disease is beyond dispute.

Another example of a MAc and MAl combination is a patient with uraemia (renal MAc) who vomits repeatedly (hypochloraemic MAl is developing).

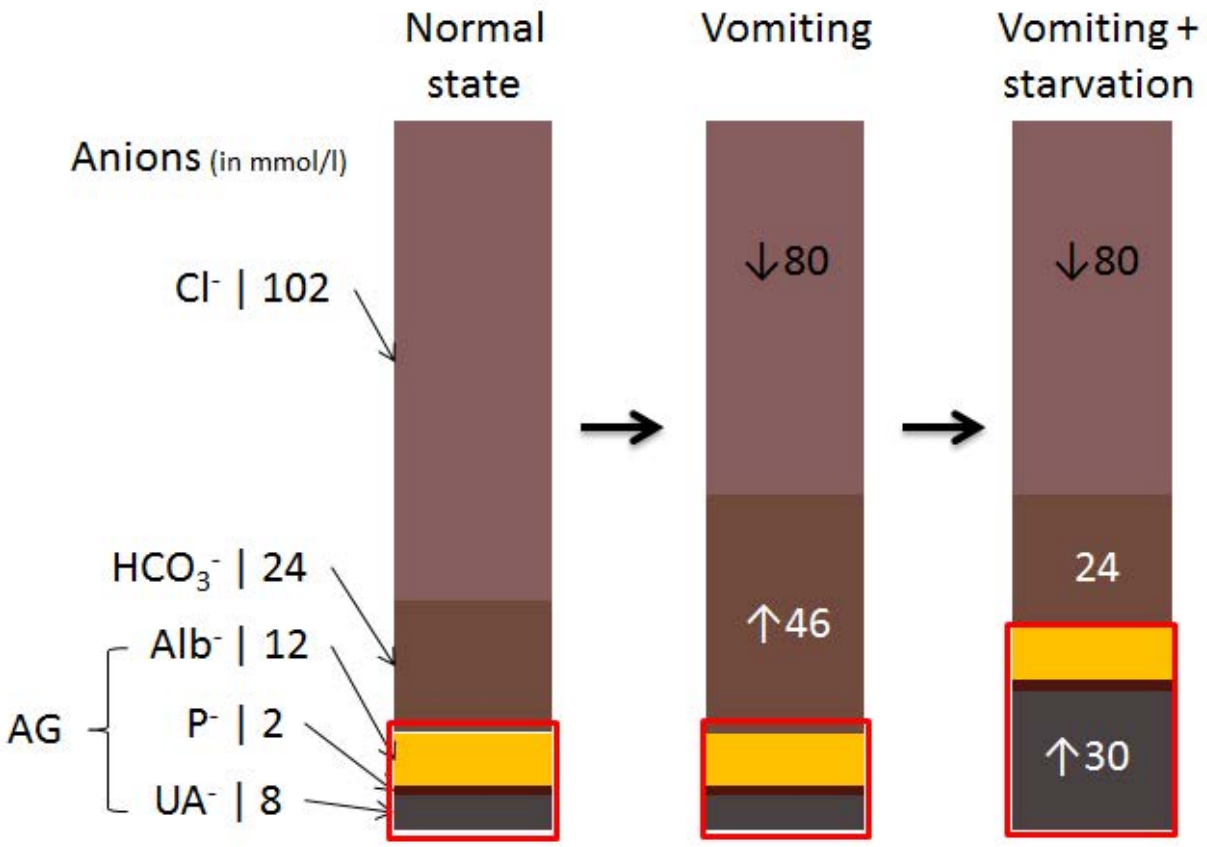


Figure 15.7. The development of a combined acid-base balance disorder with concurrent vomiting and fasting as shown by serum anions; figures show ion concentration in mmol/l. Lost chlorides are replaced with hydrogen carbonates during vomiting (and there will be no change in the anion gap). An increase in ketone bodies (organic acids) during fasting leads to a rise in unmeasured anions (and the anion gap) – refer to the preparation for the relevant calculations.

15.2.1.2. Combined MAc and RAl

In salicylate intoxication, the presence of relatively strong salicylic acid in the blood will induce MAc at first. As soon as the acid crosses the blood-brain barrier, RAl will develop as a result of respiratory centre irritation.

15.2.2. Combination of Disorders Acting in the Same Direction

15.2.2.1. Combination of Two Different MAcs

Example: A decompensated diabetic with ketoacidosis and hyperglycaemia; this will cause osmotic diuresis with polyuria and dehydration. Hypovolaemia will lead to tissue hypoxia and another MAc (this time lactic MAc) will develop.

15.2.2.2. Combined RAc and MAc

A typical example is the situation occurring in a patient with severe global respiratory insufficiency or even with cardiac and respiratory arrest. Carbon dioxide accumulates in the body (RAc) and tissues demanding oxygen produce a large amount of lactate (MAc). The result is a dramatic decrease in the internal environment pH.

Additional examples of the RAc and MAc combination, developing from a compensated simple disorder, are shown in the diagram in *Figure 15.8*.

Patient A is a diabetic with acute developing metabolic ketoacidosis (Position A_1). As the disorder has lasted for some time, it is compensated by hyperventilation, at first partially (A_2), later in full (A_3). Subsequently, the condition was complicated by respiratory muscle fatigue, which not only prevented pulmonary compensation, but also led to the development of respiratory acidosis (A_4).

The second patient B suffers from respiratory acidosis, at first acute (B_1), later partially (B_2) and fully compensated (B_3) by renal function. Extensive pneumonia has led to further worsening of gas exchange in the lungs with a further increase in hypercapnia and hypoxaemia, resulting in hypoxic lactic MAc (B_4).

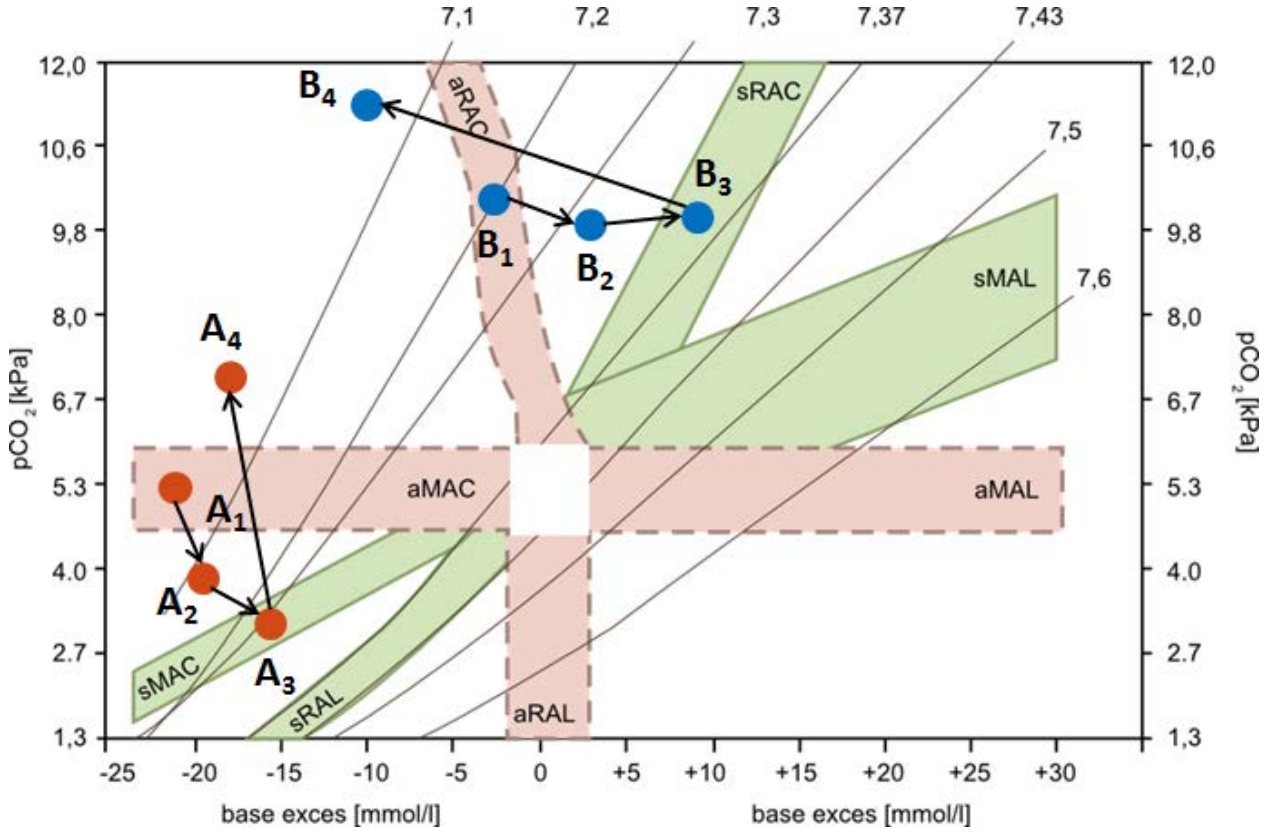


Figure 15.8. The development of a combined disorder from the primary simple compensated disorder (see text for details); a = acute, s = stabilized.

15.2.3. Combination of Multiple Disorders

Perhaps the best example of the whole range of potential disorders is a patient with severe hepatic disease. The following acid-base disorders may occur in such patient:

- *Lactic MAc* as a result of impaired lactate elimination in the liver;
- *Metabolic alkalosis* in hypoalbuminaemia (a decrease in the anion column is compensated by a rise in hydrogen carbonate concentration); decreased oncotic pressure due to hypoalbuminaemia may lead to reduced blood stream with overproduction of renin and subsequently aldosterone, which supports the development of MAl;
- *Respiratory alkalosis* is the result of respiratory centre irritation by accumulating toxic substances;
- If the hepatic failure is complicated by hepato-renal syndrome, *renal MAc* will develop;
- Additional disorders may be induced by pharmacotherapy with diuretics (e.g. hypochloraemic metabolic alkalosis due to furosemide administration);
- *Alcoholic ketoacidosis* also exists – this occurs mainly in alcoholics or following a one-time ingestion of a large amount of ethanol. It is caused by a combination of fasting (increased elimination of fatty acids and formation of ketone bodies), dehydration from vomiting (reduced excretion of ketone bodies through the kidneys), and the metabolic effects of ethanol (mainly an increase in the NADH + H^+/NAD^+ ratio and the resulting decrease in gluconeogenesis and production of ketone bodies with prevalent β-hydroxybutyrate – undetectable by urine dip sticks).

15.2.4. Summary of Mixed Acid-Base Disorder Diagnostics

A simplified procedure that offers the fastest diagnosis of acid-base disorder is as follows:

1. **Clinical approach** – detailed patient history and physical examination often provide sufficient information for correct diagnosis. For example, information about vomiting, diarrhoea, polyuria, exposure to toxic noxa, alcoholism, use of diuretics, concomitant diseases or Kussmaul breathing, acetone breath or signs of cardiac failure can lead us to a correct diagnosis very fast.
2. **Compensation** (diagnostic) **diagram** – see *Figures 2* and *3*. A good tool since it allows us to easily identify the development of a disorder over time (compensation, correction), to differentiate the respiratory and metabolic components and to estimate therapeutic doses. The diagram does not work when used for mixed disorders (counteracting disorders in particular), nor does it allow the cause of the disorder to be revealed and the zone of compensated MAl is too broad (due to the limitation of hypoventilation compensation by hypoxia).
3. **Auxiliary calculations** – based on the model of electroneutrality. Corrected chloride concentration and the anion gap (or unmeasured anions for greater precision) are particularly useful. The corrected chloride concentration helps reveal any effect of the change of chlorides on the acid-base balance, the anion gap helps in the case of increased production of acids (mainly in poisonings – methanol, ethylene glycol, ethanol, salicylates, less in lactic acidosis and ketoacidosis – the clinical picture is quite characteristic here).

15.2.5. Notes on Acid-Base Disorder Treatment

Treatment for acid-base disorders should always be based on patient history and the clinical presentation of the disease. The objective is to *remove or mitigate the root cause of the disease* (vomiting in hypochloraemic MAl, tissue oxygenation disorders in lactic MAc, bronchopneumonia in RAc, diabetic decompensation in diabetic ketoacidosis, etc.).

If acidifying or alkalizing solutions are administered to treat for metabolic acid-base disorders, the aim is to reach the **target pH**; this is 7.2 in acidoses and 7.4 in alkaloses. *Adjusting the BE value to zero is not advisable* and will not normalize pH, which should be the ultimate goal of the treatment.

Care should be taken during infusion correction of metabolic acid-base disorders, in particular compensated acidoses or mixed disturbances with acidosis and alkalosis. Most disorders will resolve spontaneously if the underlying cause is eliminated. In any case, **persistence of compensatory mechanisms following elimination of the cause of disorder** *has to be taken into account in compensated disorders*. The same certainly holds true for compensated respiratory disorders, where renal compensation persists even for several days. The principle of **concurrent treatment for both disorders** must be adhered to in *mixed disorders* so that the other one does not become prevalent. *Figure 3* shows several examples of inappropriate approach.

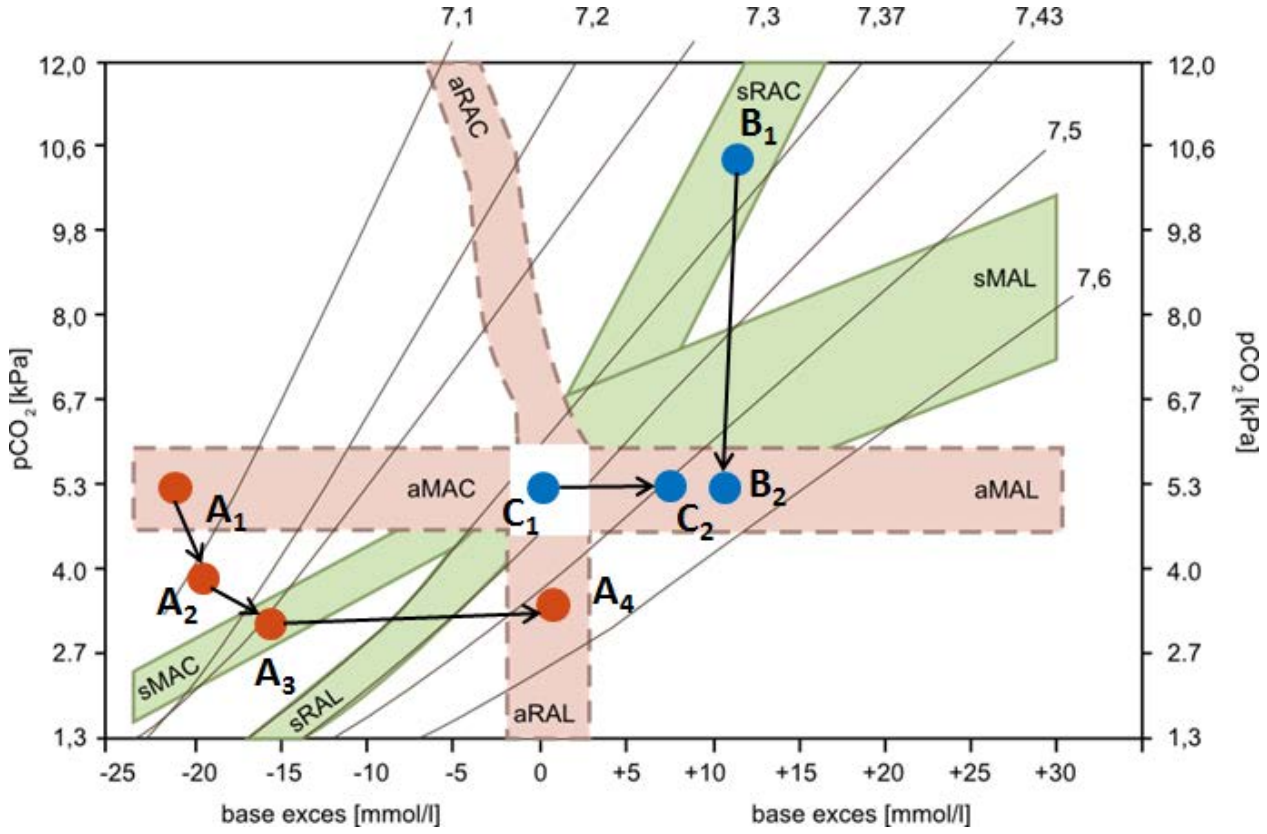


Figure 15.9. Examples of inappropriate treatment for compensated and combined acid-base balance disorders, leading to dangerous alkalaemia.

Patient A is a diabetic in ketoacidosis (A_1), which is gradually compensated by hyperventilation (A_2, A_3). If the primary disorder, MAc, is eliminated by insulin administration, persisting hyperventilation will induce severe alkalaemia in the patient, albeit with BE = 0 (A_4).

Patient B has chronic respiratory acidosis fully compensated by renal function (B_1). If normocapnia is reached in the patient by controlled ventilation, persisting renal compensation of the respiratory disorder will once again induce alkalaemia (B_2).

Position **C_1** represents a patient *with hypochloraemic MAl (vomiting) combined with metabolic ketoacidosis (fasting)*. If only MAc is eliminated, for example by glucose infusion, this would lead to a shift back to the alkalaemia region (C_2). Therefore, both disorders must be treated at the same time in combined disorders, i.e. to supply missing chlorides in the form of an isotonic NaCl solution infusion, for example, in addition to glucose.

In all cases mentioned above, the patients were at risk of *alkalaemia*, with all its adverse effects such as:

- Hypokalaemia;
- Decrease in ionized calcium and magnesium concentration with proneness to tetany;
- Oxyhaemoglobin dissociation curve shift to the left with decreased release of oxygen from the bond to haemoglobin;
- Increased toxicity of some drugs such as digoxin.

CHAPTER 16

16. Importance of Oxygen Assays

Authors: doc. MUDr. Milan Dastych, CSc.

Reviewer: prof. MUDr. Antonín Kazda, DrSc.; MUDr. Helena Brodská, Ph.D.

16.1. Role of Oxygen in the Body

The role of oxygen in the energy metabolism is essential for all life. Electron pairs produced by the oxidation of NADH and $FADH_2$ are transported to oxygen through respiratory chains in the mitochondria, by which energy is released for ATP synthesis by ADP phosphorylation. The resulting high-energy phosphate bonds are a direct source of energy for cell, tissue and organ functions in the body.

Because this source of energy is limited, the human body relies on a continuous inflow of oxygen from the atmosphere and effective oxygen supply to cellular mitochondria.

Oxygen transport is greatly facilitated by the oxygen pressure gradient between the atmosphere and the intracellular space.

There are some places on the oxygen pathway to cells that may adversely affect oxygen transport and have an ultimate impact on the development of tissue hypoxia.

16.2. Partial Oxygen Pressure along the Oxygen Pressure Gradient

	kPa
Atmospheric air	19.9
Alveolar air	14.6
Arterial blood	13.3
Venous blood	5.7
Cell cytoplasm	2.7
Mitochondria	0.3

Table 16.1. Partial oxygen pressure

16.3. Monitored Parameters Related to Oxygen Metabolism (Guidance Values)

Partial oxygen pressure in arterial blood [10.0 - 13.3 kPa]

Partial pressure in mixed/central venous blood [4.5 - 5.5 kPa]

Haemoglobin [130 - 160 g/L; 8.0 - 10.0 mmol/L]

Carbonyl haemoglobin [< 0.03] *

Methaemoglobin [< 0.02] *

Haemoglobin saturation with oxygen [0.95 - 0.98] *

P50 [3.3 - 3.9 kPa]

FiO_2 [0.21 atmosphere; 0.4 usually during pulmonary ventilation; 1.0 pure oxygen] *

Pulmonary shunts [0.02 - 0.05] *

Venous lactate-plasma [0.5 - 2.2 mmol/L]

Note: * [values are shown as a relative proportion]

16.4. Conditions for Adequate Oxygen Supply to Tissues and Possible Causes of Hypoxia

- Sufficient amount and partial oxygen pressure in the inhaled air
- Adequate ventilation – alveolar air exchange
- Optimal ventilation/perfusion ratio at the level of pulmonary parenchyma
- Efficient oxygen diffusion from alveolar air through the alveolar-capillary membrane to the blood
- (hypoxic hypoxia)
- Sufficient amount of functional (effective) haemoglobin in the blood
- Optimal oxygen affinity and bond to haemoglobin
- (anaemic/transport hypoxia)
- Adequate cardiac output per minute
- (circulatory/ischemic hypoxia)
- Effective oxygen release to tissues and cells
- (histotoxic hypoxia)

16.4.1. Sufficient amount and partial oxygen pressure in the inhaled air

The amount of oxygen available in the atmosphere depends on the barometric pressure.

The barometric pressure at the sea level is 760 mmHg ~ 760 Torr ~ 101 kPa (1 mmHg = 0.133 kPa).

According to Dalton's law, barometric pressure is the sum of partial pressures of gasses contained in the atmosphere.

$$P_{(atmospheric)} = pO_2 + pCO_2 + pN_2 + pH_2O + p_{other\ atmospheric\ gases}$$

The partial pressure of each atmospheric gas is calculated from the barometric pressure depending on their volume concentration:

Oxygen 20.93%,

Nitrogen 78.1%,

CO_2 0.03%,

Inert gases about 1%.

The percentage content of all atmospheric gases is the same at all altitudes.

Water vapour pressure [6.25 kPa] must be included in the partial pressure calculation.

A decrease in pO_2 in atmospheric air occurs at high altitudes, on alpine hikes, during high-altitude flights, or if oxygen is depleted by combustion in enclosed areas, or by oxygen displacement from the atmosphere by another gas such as CO_2 in some caves where accumulated, heavier CO_2 creates an oxygen-free zone at ground level.

16.4.2. Adequate ventilation – alveolar air exchange

Hypoventilation can be caused by a decreased tidal volume (shallow breathing) or decreased respiratory rate (hypopnoea). Since pulmonary ventilation is ensured by atmospheric gas inspiration by volume changes in the thorax (through the activity of respiratory muscles, including the diaphragm), painful respiratory movements (chest injuries,

pleuritis) lead to a reduced tidal volume. Respiratory centre inhibition may lead to an extreme decrease in the respiratory rate down to a few breaths per minute (e.g. morphine poisoning). Also, exhaustion of critically ill patients with signs of muscle weakness may be a cause of considerable hypoventilation and indications for artificial pulmonary ventilation.

The importance of dead space, the volume of air that remains in the lungs and airways after maximum expiration, increases during shallow breathing (low tidal volume). Dead space has a lower pO_2 and a higher pCO_2, which means it „dilutes" the inspired atmospheric air; the greater the ratio of relative dead space to tidal volume, the greater is such dilution.

The alveolar air pO_2 is about 14.6 kPa as compared with atmospheric air pO_2 of 19.9 kPa.

16.4.3. Optimal ventilation/perfusion ratio at the level of pulmonary parenchyma

Even in physiological situations, perfused but non-ventilated areas occur in the pulmonary parenchyma. Their proportion may increase considerably in pathological situations. Causes may include conditions leading to atelectasis. Blood flowing through these areas keeps its venous character and decreases haemoglobin saturation with oxygen in arterialized blood. This increases the value of pulmonary shunts, non-oxygenated blood after passage through the lungs (normally under 0.05; 5%). In character, pulmonary shunts correspond to the haemodynamic right-to-left cardiac shunt. Depending on the extent of the disorder, the pulmonary shunt area may reach tens of percent with highest values of up to 80%; 0.8 in ARDS with a shock lung.

16.4.4. Oxygen diffusion from alveolar air through the alveolar-capillary membrane to the blood

Oxygen diffusion is 20 times more difficult/slower than CO_2, which is why it is more sensitive to conditions hampering gas diffusion in the lungs. Structural or pathophysiological (alveolar/capillary) diffusion disorders alter oxygen transport considerably but have a minimal effect on the diffusion/excretion of CO_2.

Disorders of oxygen diffusion in the lungs may be caused by:

failure/destruction of alveoli

The normal surface area of alveoli is about the size of a tennis court. This area may decrease critically in emphysema, for example. Diffusion is considerably impaired by fibrotic processes such as interstitial pulmonary fibrosis.

pulmonary oedema

Oxygen diffuses from alveoli to pulmonary capillaries through a very thin space. The fluid accumulated in this space during pulmonary oedema extends the diffusion path and hampers the diffusion.

insufficient pulmonary perfusion

Reduced perfusion → embolism, pulmonary hypertension, cardiac failure

16.4.5. Sufficient amount of functional (effective) haemoglobin in the blood

A normal value of „effective haemoglobin" in fully grown red blood cells is essential for ensuring sufficient oxygen supply to tissues according to their current requirement. This haemoglobin represents the oxygen-carrying capacity of blood; 1 g of haemoglobin can bind 1.39 ml (0.062 mmol) of oxygen (O_2). A lowered effective haemoglobin concentration may cause anaemic hypoxia.

Apart from adult haemoglobin A, blood also contains variants of haemoglobin incapable of reversible oxygen binding, and that do not take part in oxygen transport within the body. They are called dyshaemoglobins, one of them being carbonyl haemoglobin (COHb – town dwellers under 5%, smokers up to 10%) with a reversible CO bond, which is 200 times stronger than the oxygen bond, and methaemoglobin (MetHb – under 3%), unable to bind oxygen because the Fe in its molecule is irreversibly oxidized to Fe^{3+}.

Sulphaemoglobin and cyanhaemoglobin also do not participate in the oxygen transport.

These forms of haemoglobin gain pathophysiological importance if their level (%) in the blood increases to such

extent that they may cause hypoxia (CO intoxication, effect of nitrites, iatrogenic action of drugs such as sulphones – organic sulphur compounds used in dermatology, etc.).

A special variant is foetal haemoglobin. With its higher oxygen affinity, it supplies the foetus with oxygen in specific intrauterine conditions, and is replaced with adult haemoglobin A following birth.

Forms of haemoglobin with genetically determined alterations to their primary structure (composition and the order of amino acids) that also alter the oxygen bond are called haemoglobin variants or haemoglobinopathies.

16.4.6. Optimal oxygen affinity and bond to haemoglobin

The saturation/dissociation curve shows haemoglobin's affinity for oxygen. Partial oxygen pressure is the independent variable (x-axis), and haemoglobin saturation with oxygen is the dependent variable (y-axis). Oxygen does not bind to haemoglobin uniformly – the bond is difficult and energy-consuming at first, because most electrostatic bonds between the alpha and beta subunits have to be broken. The easier bonding of additional oxygen molecules is represented by the shape (slope) of the haemoglobin dissociation curve.

The typical S-shaped form of the saturation/dissociation curve conceals important advantages to oxygen transport within the body.

The end part of the curve on the right with high partial oxygen pressure and saturation values reflects the situation in the lungs. The curve shows that even a relatively large drop in partial oxygen pressure will cause an insignificant decrease in haemoglobin saturation with oxygen (a 3 kPa drop in pO_2 leads to a 2 % drop in saturation).

On the contrary, the left part of the curve with low pO_2 values describes the situation in tissues. The same drop in pO_2 will cause a significant drop in haemoglobin saturation, and therefore a large amount of released oxygen available for diffusion into tissues and cells.

Haemoglobin's affinity for oxygen is not constant. It changes (increases and decreases) in response to changes in certain physical and chemical parameters (temperature, pH, pCO_2, 2,3-bisphosphoglycerate). This phenomenon is referred to as the Bohr effect.

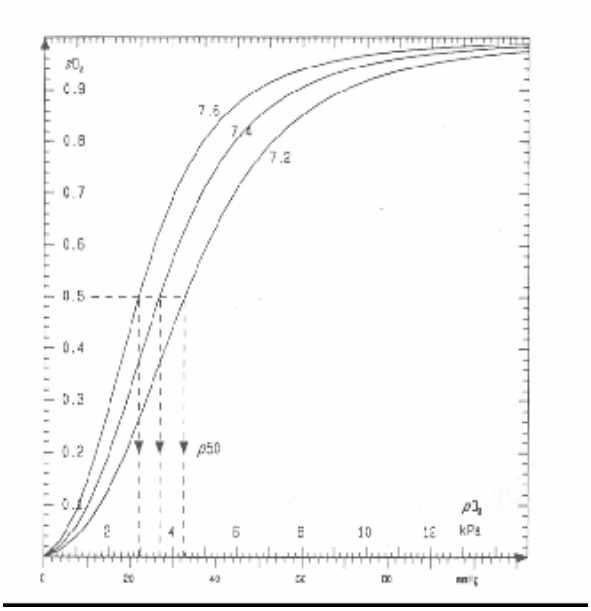


Figure 16.1. The effect of blood pH on the position of the haemoglobin dissociation/saturation curve. The position of the curve on the graph is characterized mathematically by the p50 value (3.0 << 3.6 >> 4.4 [kPa])

Haemoglobin's affinity for oxygen is increased by the following parameters:

Alkaline pH

Decreased temperature

Decreased pCO_2

Decreased 2,3-bisphosphoglycerate level

Haemoglobin's affinity for oxygen is decreased by the following parameters:

Acidic pH

Increased temperature

Increased pCO_2

Increased 2,3-bisphosphoglycerate level

Given the big functional reserve of the oxygen bond to haemoglobin in the lungs (see the haemoglobin saturation curve), the benefit of this phenomenon clearly points to oxygen metabolism in tissues (in terms of easier oxygen release).

The effect of the increased 2,3-bisphosphoglycerate level also manifests itself as a compensatory mechanism in developing anaemia, where it acts as a compensation for a potential hypoxic condition. By linking to the Hb molecule it decreases Hb's affinity for oxygen.

Patients with gradual/slow development of anaemia have an elevated 2,3-BPG concentration, which can partially explain why even patients with an extremely low Hb level are capable of ordinary activities.

There is also an increase in 2,3-BPG as part of the body's adaptation to high altitudes.

P50 value

P50 indicates the partial oxygen pressure at which haemoglobin, in a patient's current situation, is 50% saturated with oxygen.

An increase in haemoglobin's affinity for oxygen is graphically represented by a shift of the saturation/dissociation curve to the left; a decrease in affinity will shift the curve to the right.

The central position of the saturation/dissociation curve, or the range of its shift to the left or right, is expressed by the p50 value.

The p50 reference interval for pH 7.4 and T 37°C is 3.3 - 3.9 kPa. P50 differs from the standard value if pH deviates from 7.40, pCO_2 from 5.3 kPa, temperature from 37°C or 2,3-BPG concentration from 5.0 mmol/L. P50 for newborns is 2.4 - 3.2 kPa due to foetal Hb.

Increased P50 points to a shift of the Hb dissociation curve to the right, i.e. a lower Hb affinity for oxygen. Causes may include: hyperthermia, acidaemia, hypercapnia, elevated 2,3-BPG or a combination thereof.

Decreased P50 points to a shift of the dissociation curve to the left, and a higher oxygen affinity for Hb. Main causes include: hypothermia, acute alkalaemia, hypocapnia and low 2,3-BPG concentration. A pathophysiological impact of the elevated oxygen affinity for haemoglobin is less effective oxygen release from oxyhaemoglobin in peripheral tissues, and therefore a low tissue pO_2.

16.4.7. Adequate cardiac output per minute

Cardiac output per minute represents the volume of blood ejected by the cardiac function to the blood circulation (systolic volume x heart rate) per minute, and reacts to the current requirement for oxygen supply to tissues. A decrease in the minute output due to cardiac insufficiency or heart rhythm disorder leads to insufficient oxygen supply to tissues and hypoxia. In contrast, when there is insufficient oxygen supply to tissues, increased cardiac output will increase the oxygen supply to tissues, if the myocardium capacity allows it.

The cardiovascular system's adaptation reactions to hypovolaemia or shock with circulation centralization dramatically decrease the oxygen supply to hypoperfused tissues with subsequent signs of their ischaemia/hypoxia.

In severe sepsis, septic shock and other critical conditions, when the blood flow through the splanchnic region and the kidneys decreases, blood flow through the myocardium and brain is maintained.

16.4.8. Oxygen release to tissues and cells

Oxygen is released from the bond into tissues as a result of the large pressure gradient between the blood pO_2 and tissue cells, and the easy dissociation of the oxygen-haemoglobin bond at low pO_2 at the interface of the blood and tissue cell surface.

The pO_2 at the venous end of the capillary is about 5.0 kPa, so the normal arteriovenous pO_2 difference ranges between 5.0 and 8.3 kPa.

Arteriovenous difference

The a-v difference in the oxygen content shows the current amount of oxygen supplied by the blood to tissues.

	Arterial blood	Mixed venous blood
Haemoglobin concentration [mmol/L]	9.0	9.0
Haemoglobin saturation [0.00]	0.98	0.68
pO2 [kPa]	10.6	5.0
ctO2 [mmol/L]	9.0	6.7
A-v O2 difference [mmol/L]	2.3	
A-v O2 difference [mL/L]	51	

Table 16.2. shows an example of the a-v difference with different oxygen metabolism parameters.

ctO_2 – total oxygen content in the blood; 1 mol of haemoglobin binds 1 mol of oxygen.

An a-v difference of 51 mL O_2/L of blood is required to cover a rest oxygen consumption of about 250 mL/min (70 kg weight) and rest cardiac output per minute (5 L/min.).

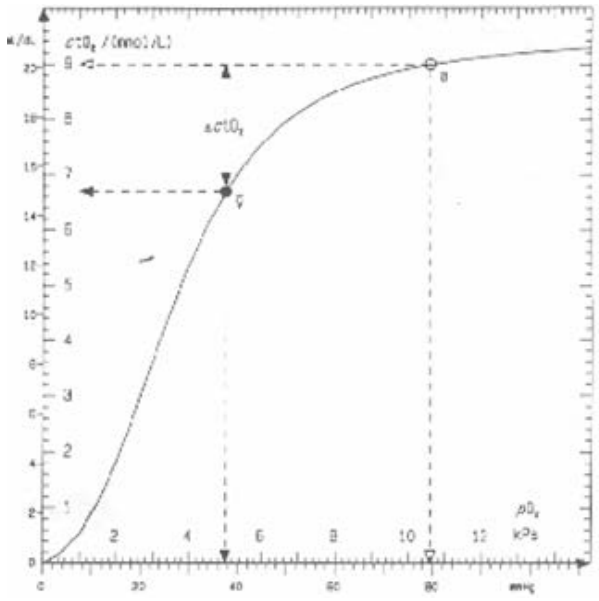


Figure 16.2. The values shown in the table are derived from the saturation/dissociation curve.

In addition to reaffirming the benefits of the S-shaped haemoglobin saturation/dissociation curve, the critical importance of a sufficient pressure gradient for the transport of oxygen must also be kept in mind. At a low p_aO_2 in the artery, a release of 2.3 mmol/L (51 mL/L) of oxygen has to be accompanied by a drop in p_vO_2 in mixed venous blood. Values under 4.0 kPa may lead to hypoxia. The critical value for p_vO_2 in mixed venous blood is < 2.5 kPa, which is when hypoxic brain damage occurs.

The total oxygen content ctO_2 is shown on the y-axis instead of oxygen saturation sO_2.

There is an assumption that 1 mol of haemoglobin binds 1 mol of oxygen. ctO_2 is 9.1 mmol/L at 100% Hb saturation (150g/L = 9.1 mmol/L). ctO_2 decreases in proportion to decreasing saturation.

16.5. Respiratory Insufficiency

Chronic respiratory insufficiency as a result of a chronic pathological process in the lungs is divided into partial, affecting the oxygen supply (hypoxia), and global, which is also associated with carbon dioxide retention (hypercapnia and respiratory acidosis).

Acute respiratory insufficiency – acute respiratory distress syndrome (ARDS) is a serious condition connected with high mortality. This condition is often connected with shock (shock lung), also as part of multiple organ dysfunction with different aetiologies.

16.6. Lactate

Lactate is the product of anaerobic glycolysis. Elevated lactate concentration in the plasma (> 2.2 mmol/L) may be caused by increased production or reduced elimination. In connection with oxygen metabolism, an increasing lactate concentration, lactate acidosis, goes together with tissue hypoxia (type A). Lactate acidosis may also occur as a consequence of lactate elimination disorder in renal and hepatic failure, due to nutritional and iatrogenic effects such as complications in biguanide treatment, realimentation in thiamine deficiency, and ethanol, salicylate or methanol intoxications, including congenital carbohydrate metabolism disorders (type B).

A rise in lactate concentration in critically ill patients is an alarming symptom, which often precedes and accompanies shock conditions. Continuous monitoring of the plasma lactate level and its trends is fully indicated to see whether therapeutic actions have been successful. A permanently elevated or increasing value is associated with a worsening prognosis and high mortality.

In general, 18 - 22% mortality is reported for shock, non-septic conditions with lactatemia 1.3 - 4.4 mmol/L, 73% mortality for hyperlactatemia 4.5 – 8.9 mmol/L, and 100% for values over 13 mmol/L. Values as low as 3.0 - 3.5 mmol/L suggest a poor prognosis for septic conditions. A rise in lactate is a signal that sepsis has transferred from a metabolically compensated into a decompensated stage, where tissues consume less oxygen even if the supply is sufficient.

16.7. Perinatal Asphyxia

Asphyxia is one of the commonest causes of death or severe damage to the foetus. Causes can be on the side of the mother (cardiac or pulmonary diseases, disturbed blood flow through the placenta, abnormal uterine contractions), or else asphyxia may be induced by umbilical cord compression or knotting, premature separation of the placenta, or pathology of the foetus (congenital heart defects, heart rhythm disorders, severe anaemia, sepsis, pneumonia, etc.). Respiratory distress syndrome (RDS) in newborn infants is a separate condition. It occurs in premature infants in particular, and is caused by insufficient pulmonary surfactant in the immature lungs, which prevents alveolar collapse in the expiratory phase and avoids functional atelectases. The degree of perinatal asphyxia is clinically evaluated using the Apgar score (the international system for rating a newborn's neonatal adaptation; the criteria are Appearance (skin colour), Pulse (pulse rate), Grimace (reflex irritability), Activity (muscle tone) and Respiration; the infant receives 0 - 2 points for each function, with 10 points as a maximum). A score of 0 - 3 at the 5^{th}, 10^{th}, 15^{th} and 20^{th} minute from birth is considered deeply pathological. Laboratory measurement of pH in the cord blood collected during delivery is used to evaluate the seriousness of foetal hypoxia; pH values < 7.0 are indicative of hypoxia.

16.8. High Altitude Effect

The composition of atmospheric air at sea level and on top of Mount Everest is the same; only the barometric pressure, and thus the partial pressure of oxygen, is different. Lower pO_2 causes hypoxia, which induces many adaptation reactions. Spending extended periods at high altitude often induces an increased production of erythropoietin (EPO) in the kidneys with subsequent stimulation of haemoglobin and red blood cell production and a resulting increase in the oxygen-carrying capacity of blood. There is also a rise in 2,3-bisphosphoglycerate, which facilitates oxygen release into tissues. These adaptation mechanisms are used to prepare mountaineers for high altitudes and often also to increase the oxygen-carrying capacity in different high-performance sports disciplines.

16.9. Diving

Divers face the opposite problem. As the immersion depth grows, the pressure and p_aO_2 of inhaled air increases

dramatically up to toxic levels. There is a risk of hyperoxia if air of the same composition as atmospheric air is inhaled. A gas mixture containing less oxygen (Trimix) is therefore used for deep diving. Surplus nitrogen and oxygen is replaced with an inert gas (helium).

16.10. Measured and Counted Oxygen Metabolism Parameters

- p_aO_2, Hb saturation, haemoglobin fractions, lactate, (pH, p_aCO_2, FiO_2, etc.)
- Pulmonary shunts, oxygen consumption, a-v difference

Anticoagulated (heparinized) arterial blood or arterialized capillary blood from the hyperaemized adult's earlobe or newborn's heel is most suitable to determine oxygen metabolism parameters.

Blood **paO_2** is measured using the Clark-type electrode on the amperometric principle. The electrode is a standard part of acid-base and blood gas parameter analyzers, together with the pH electrode and Severinghaus electrode to determine pCO_2. In addition to analyzers intended for clinical and biochemical laboratories, there are also instruments with very simple operation and maintenance for clinical use at the patient's bedside – Point Of Care Testing (POCT).

Haemoglobin oxygen saturation (SO_2) is usually derived in these instruments empirically from p_aO_2, pH and haemoglobin, from the normal shape and position of the Hb dissociation curve.

A correct result is conditioned by normal oxygen affinity for Hb, a normal 2,3-BPG concentration and the absence of dyshaemoglobins.

The drawbacks and limitations of the approximative method mentioned above are eliminated by direct measurement of all haemoglobin type contents in the blood based on their different absorption spectra (oxyhaemoglobin, reduced haemoglobin, carbonyl haemoglobin, methaemoglobin, sulphaemoglobin, foetal haemoglobin, oxygen saturation and total oxygen content ctO_2). Following ultrasonic red blood cell haemolysis, the haemolyzate is measured by spectrophotometry, and the haemoglobin levels present are computed from their maximum absorptions. The oxyhaemoglobin to total haemoglobin ratio is the usual result of the saturation calculation.

A CO-oximeter is a standard part of the advanced acid-base analyzers used today.

Pulse oximeters are used to monitor haemoglobin oxygen saturation.

The method is based on the principle of different absorption of red (or infrared) light by haemoglobin and oxyhaemoglobin. A sensor emitting radiation with wavelength of 660 nm (red light) and 940 nm (infrared light) is applied to an acral part of the body, usually a finger or earlobe. Light is passed through this part of the body and the transmitted light is measured. The light is absorbed by the tissue and venous, capillary and arterial blood, of which only the arterial component has a pulsing character. The absorption of light of the two wavelengths by pulsing and non-pulsing components is used to estimate haemoglobin oxygen saturation in arterial blood (SaO_2) in percent. Naturally, if the patient's blood contains other haemoglobin derivatives (COHb, MetHb), the method gives falsely high results of saturation. Neither is it suitable in the case of disturbances of peripheral circulation: the measured result will correspond to the value in the poorly perfused periphery, and not to arterial blood.

Pulse oxymetry is used for its simple and quick application when estimating a patient's blood oxygenation during anaesthesia in intensive care units and emergency wards, where patients are at risk of rapid development of respiratory insufficiency.

Pulmonary shunts are usually computed automatically from measured blood gas values in arterial and mixed/central venous blood, measured oxygen saturation, the FiO_2 value, temperature and barometric pressure.

The arteriovenous p_aO_2 difference and the position of the haemoglobin dissociation curve are used to derive the a-v difference in oxygen content, i.e. the content of oxygen available to tissues in that specific situation.

Information on current oxygen consumption can be obtained using instruments continuously analyzing oxygen content in inspired and expired air. Modern calorimeters work on the same principle to determine energy output/energy requirement using indirect calorimetry; 1.0 litre of consumed oxygen is equivalent to energy of 4.8 kcal (20.2 kJ).

In basal condition, the oxygen consumption of an adult is about 3.5 $mL.kg^{-1}.min^{-1}$. Cardiopulmonary system performance is expressed by the maximum oxygen consumption reached under an increasing dynamic load. This corresponds to about ten times the basal oxygen consumption in middle-aged men who do not take regular exercise.

16.11. Treatment for Hypoxia

Spontaneous breathing – oxygen therapy using oxygen-enriched air

Artificial pulmonary ventilation (ventilation modes)

At constant pressure – pressure-controlled ventilation

At constant volume – volume-controlled ventilation

PEEP – positive end-expiratory pressure

High-frequency ventilation

ECMO – extracorporeal membrane oxygenation

Hyperbaric oxygen therapy

(CO intoxication, caisson disease, Clostridium infections, risk of transplanted grafts/skin flap rejection)

Indications for oxygen therapy and controlled breathing

The following biochemical values are indications for oxygen therapy and controlled breathing:

paO_2 < 9.3 kPa and p_aCO_2 > 6.6 kPa in unventilated patient in acute condition; these are arterial blood values.

The aim of **oxygen therapy** is to increase haemoglobin oxygen saturation to 0.85 - 0.90 with subsequent normalization of blood gas values. It is advisable to select the lowest effective oxygen concentration in the inhaled mix (FiO_2) to prevent the toxic effects of oxygen on lung tissue. Oxygen concentrations under 60% are well tolerated over long periods of time. A permanent increase in FiO_2 over 60% will lead to the development of inflammatory alterations and infiltration of alveoli with subsequent fibrosis. If oxygen concentrations over 60 – 80% have to be supplied, controlled ventilation should be used.

Pure, 100% oxygen is supplied during emergency resuscitation; 100% oxygen inhalation is safe for at least 6 hours. Toxic effects start to develop after 48 hours.

Increased attention should be paid to oxygen therapy of chronic hypercapnia conditions; the reason is that the only stimulus for the respiratory centre is the low p_aO_2. If this last ventilation stimulus is removed, further p_aCO_2 increase may occur with depression or even arrest of ventilation.

Also, **controlled ventilation** leading to a p_aCO_2 decrease must be slow – the decrease rate should not be faster than 0.3 – 0.7 kPa/hr. Always keep in mind the persistence of the renal compensation of the disorder and the risk of transition to alkalaemia. Progress should be monitored by repeated blood gas analyses – before ventilation, 20 – 30 minutes after the start of controlled ventilation, and following each change in the ventilator set-up.

Pre-analytical phase

The essential pre-analytical requirement is to ensure an anaerobic method of blood collection and fast transport for the analysis. Anaerobic collection means the shortest possible contact of the collected blood with atmospheric air (capillary collection), and absolute prevention of air bubbles in the capillary tube or collection syringe. Failure to adhere to this principle may degrade the test completely and produce false results. The collected blood must be mixed perfectly with the anticoagulant (heparin) in order to prevent the formation of blood clots (collection syringe rotation, mixing the content of the capillary tube with a wire and magnet) that make the assay impossible and may clog up the analyzer and put it out of order.

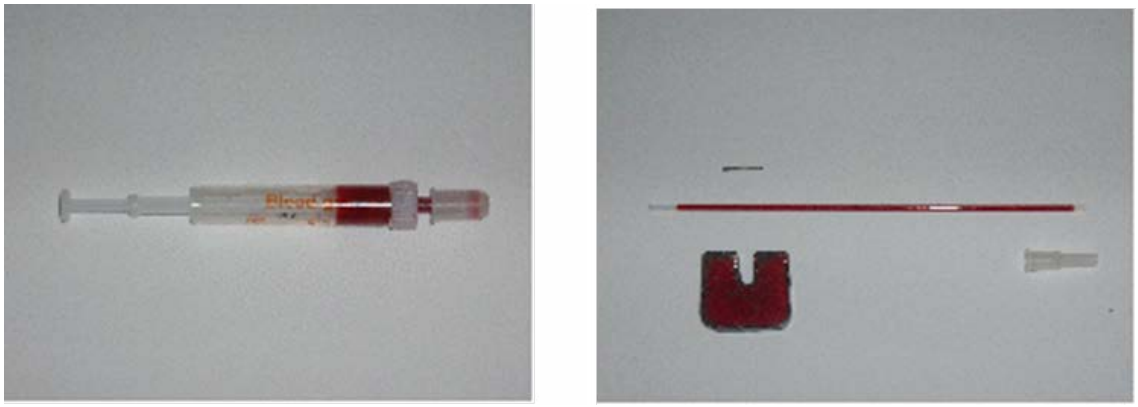


Figure 16.3. Anaerobic blood collected in a heparinized syringe and capillary tube; wire and magnet to mix anticoagulated blood; microstrainer to capture microclots during aspiration into the instrument.

CHAPTER 17

17. Importance of Osmolality Tests, Water Metabolism

Authors: prof. RNDr. Eva Táborská, CSc.; doc. MUDr. Milan Dastych, CSc.

Reviewers: prof. MUDr. Antonín Kazda, DrSc.; MUDr. Helena Brodská, Ph.D.

17.1. Osmolality and Water Metabolism

Water accounts for 60 percent of average human body weight. A uniform and constant distribution of water, and the substances dissolved in water, in different body compartments is one of the essential requirements for all vital functions of the organism.

The essential importance of water for the metabolism and proper function of cells, tissues, organs and the body as a whole is reflected in the perfect mechanism which regulates the volume, composition and distribution of water to each body compartment. Body water in cells provides an environment for all enzymatic and other physical and chemical reactions, the transport of substrates and products, and for maintaining the optimum volume of intracellular structures, cells and tissues. On the macro-scale of the body, water is essential for effective blood circulation with optimum tissue perfusion.

Water volume and water transfers within the body are regulated by systems equipped with sensitive osmoreceptors, volume receptors and baroreceptors. The sodium cation (Na^+), which is accountable for 90% of extracellular fluid osmolality, lies at the centre of regulation loops, the purpose of which is to maintain (preserve) the optimum content and distribution of water in the body.

The effects of mechanisms regulating osmolality, Na+ and water therefore interact and affect each other see also the chapter Importance of Na, K, Cl Assays.

17.1.1. Water Balance

In evaluation of water balance the intake means the water added to the body (food and beverages) and metabolic water (about 0.2 L per day). Water output, meanwhile, includes losses in urine, stool (usually 0.1 L/day, up to several litres during diarrhoea), sweat (0.1 L), by evaporation through the skin in thermoregulation (0.35 L) and losses in air exhaled from the lungs (0.35 L). Losses through the skin, in sweat and lungs, and metabolic gain, are referred to as insensible (unquantifiable when determining water balance), yet must be taken into account. Immeasurable losses are about 700 mL/day in normal circumstances, but they increase during fever and enhanced sweating, and, conversely, may decrease when artificial ventilation is applied.

The regulation of the volume, osmolality and composition of water in different body compartments serves primarily for ensuring optimum conditions in the intracellular space in reaction to changes in the ECF.

The unique function of the individual parts of nephrons ensures that about 99% of the daily volume of about 180 litres of glomerular filtrate (primary urine) is reabsorbed to produce definitive urine with osmolality from 50 to 1400 mmol/kg, depending on the extremes of variation in water intake.

Daily osmotic load excretion in the urine may fluctuate considerably even in physiological circumstances. Excreted osmotic loads may exceed 1500 mmol/24 hrs in healthy people with a high intake of salt and proteins, and who are highly physically active. In contrast, daily excretion of the osmotic load falls below 600 mmol if the protein intake is low and in resting non-catabolic patients.

Osmotic load excretion increases significantly in critical conditions. A major role in this excretion is played by high urea, ions and frequent glycosurias with a diminished tolerance for glucose, osmotherapy solutions, or low-molecular substances excreted during catabolism and intoxications. Losses over 1500 mmol/day occur in more than 1/3 of hyper-

catabolic patients, and even losses reaching 3000 mmol/day and more are not exceptional.

Excreting such a high osmotic load is very burdensome on the kidneys. If their renal concentrating ability is impaired following hypoxic accidents, intensive diuretic treatment or potassium depletion, for example, patients must be sufficiently hydrated to be able to excrete such high osmotic load and prevent the retention of osmotically active catabolites in the body.

17.1.2. Osmolality

Osmolality is a physical property of solutions based on the solute concentration expressed as mmol/kg (w/w). Osmolality is related to changes in many properties of the solution relative to solute-free water, such as decreasing the freezing point and water vapour depression. These properties are referred to as colligative properties (dependent on the number of molecules) and also form the basis of routine measurement of osmolality in biological samples.

The term *osmolarity* is an expression of osmolality in mmol/l (w/v = weight/volume). It is imprecise (unsuitable) for hyperlipidaemia and hyperproteinaemia, for urine measurements, and in the presence of some osmotically active substances such as alcohol or mannitol. In addition, it is related to the volume of the solution, which depends on temperature.

17.1.3. Osmolality and Body Water Regulation

17.1.3.1. Systemic Regulation of Osmolality

The maintenance of water and ion balance is based on the regulation of osmolality and extracellular fluid volume. Although this regulation is controlled by different mechanisms, both regulation systems have in common a relationship to plasma Na^+ ion concentration. Osmolality (affected mainly by Na^+ concentration) is regulated by changes in water balance, and intravasal volume is regulated by changes in Na^+ concentration.

The essential hormone for osmolality regulation is the antidiuretic hormone (ADH). ADH is a nonapeptide synthesized in the hypothalamus (nucleus supraopticus and paraventricularis). Secretory granules with ADH are stored in the posterior pituitary gland. ADH half-life is 15 - 20 minutes.

The hypothalamus reacts to an increase in blood osmolality, recorded by osmoreceptors in the hypothalamus, by an impulse to raise antidiuretic hormone (ADH) secretion and by inducing a sense of thirst.

The natural response to the sense of thirst is increased fluid intake, which leads to an increase in water content in the extracellular space, dilution of the elevated Na^+ concentration and a subsequent decrease in plasma osmolality. *Thirst is the decisive feature in fluid intake regulation in relation to plasma osmolality.* Thirst is stimulated by a 2% and greater increase in osmolality.

ADH is responsible for the hormonal regulation of osmolality by controlled water excretion through the kidneys. ADH is transported by blood into the kidneys where it activates aquaporins by acting on V2 receptors in the distal tubule and the connecting segment. This increases water absorption in the distal part of the kidneys. Water retained in the kidneys decreases plasma osmolality with the subsequent damping of ADH secretion. To keep osmolality at 285 ± 10 mmol/kg, osmoreceptors in the hypothalamus react very fast even to small changes in plasma osmolality. A 1-2% rise in osmolality will lead to a four times higher ADH concentration in the circulation. Conversely, a 1-2% drop in osmolality will stop ADH secretion. The response is fast.

There are also other factors that participate in ADH secretion. Aortal and carotid baroreceptors and arterial volume receptors stimulate ADH secretion when blood pressure drops by 10%. As pressure drops, the activity of sympathetic nerves in the heart and veins simultaneously increases, and the effect of ADH and the sense of thirst are supported by feedback. The ADH circulating in the blood acts on arterioles and induces their constriction. The decrease in blood flow through the kidneys induces the release of renin, which leads to a higher production of angiotensin II. Angiotensin II acts in the brain and induces the release of the ADH.

Renal regulation plays a major role in excess water management in the body, while thirst is a more important regulator preventing water deficiency and dehydration.

17.1.3.2. Osmolality Regulation at Cellular Level

The **effective osmolality** of fluids is decisive for water distribution and transfers between cells and extracellular spaces. The level of effective osmolality is influenced by substances that accumulate only in some spaces, and cannot spontaneously pass through the capillary wall or cell membrane. Typically these are sodium ions and glucose. If their concentration in the ECF increases, this fluid becomes hypertonic compared to ICF (ECF has a higher effective osmolality), whereas if their concentration decreases, effective osmotic ECF pressure will decrease. Changes in effective osmolality lead to the transfer of water to areas with higher effective osmolality. This process continues until osmotic balance is achieved. Urea, which freely passes through cell membranes, has no influence on effective osmolality or water transfers.

If the *cell is in a hypertonic environment*, typically in hypernatraemia, the cell starts to quickly lose water. The decreased intracellular volume activates processes that increase the volume of osmotically active substances, followed by the entry of water into the cell and adjustment to the cell volume. In particular, an influx of Na^+ ions occurs due to the mechanism of Na^+ Cl^- symport, Na^+ K^+ $2Cl^-$ symport and Na^+/H^+ antiport. These mechanisms, based on an increase in intracellular sodium ion concentration, are effective only for a short time because an increase in the Na^+/K^+ pump activity follows very quickly. This is the reason cells in regular contact with a hypertonic extracellular environment, brain cells in particular, have other secondary mechanisms at their disposal to maintain an optimum volume. They can start synthesizing special osmotically active organic compounds in a few hours. The nature of these molecules was unknown for a long time, and they were called idiogenic osmoles. Recent findings have shown they consist of multiple organic compounds (polyols, amino acids, amines, choline compounds, creatinine phosphate). As a consequence, patients with chronic hypernatremia or diabetic hyperglycaemia are able to tolerate high plasma osmolality levels without any damage to brain cells.

If the cell is *in a hypotonic environment*, the cell's volume will start to increase due to the transport of water. In reaction to this effort to maintain cell volume, transport processes become activated leading to a decrease in the volume of osmotically active substances in the cell, the balancing of osmolality and the maintenance of the cell volume. Different cells employ different regulation mechanisms, most often the enhanced release of K^+ through the stimulation of channels for K^+. The content of organic compounds (idiogenic osmoles) in brain cells decreases.

17.1.3.3. Intravasal Volume Regulation

Adequate intravasal volume is required to maintain blood pressure and to maintain the necessary perfusion of tissues and organs. The regulation of the content (amount) of Na^+ and water take part in the intravasal volume control.

The renin – angiotensin - aldosterone system has primary responsibility for decreases in intravasal volume.

Renin is excreted by juxtaglomerular cells in afferent renal arterioles close to renal glomeruli in reaction to reduced perfusion of the kidneys.

Renin converts angiotensinogen to angiotensin I, which transforms into angiotensin II at a later stage. Angiotensin II causes vasoconstriction, which quickly increases blood pressure as well as aldosterone secretion, which then increases the retention of Na^+ and the water accompanying it.

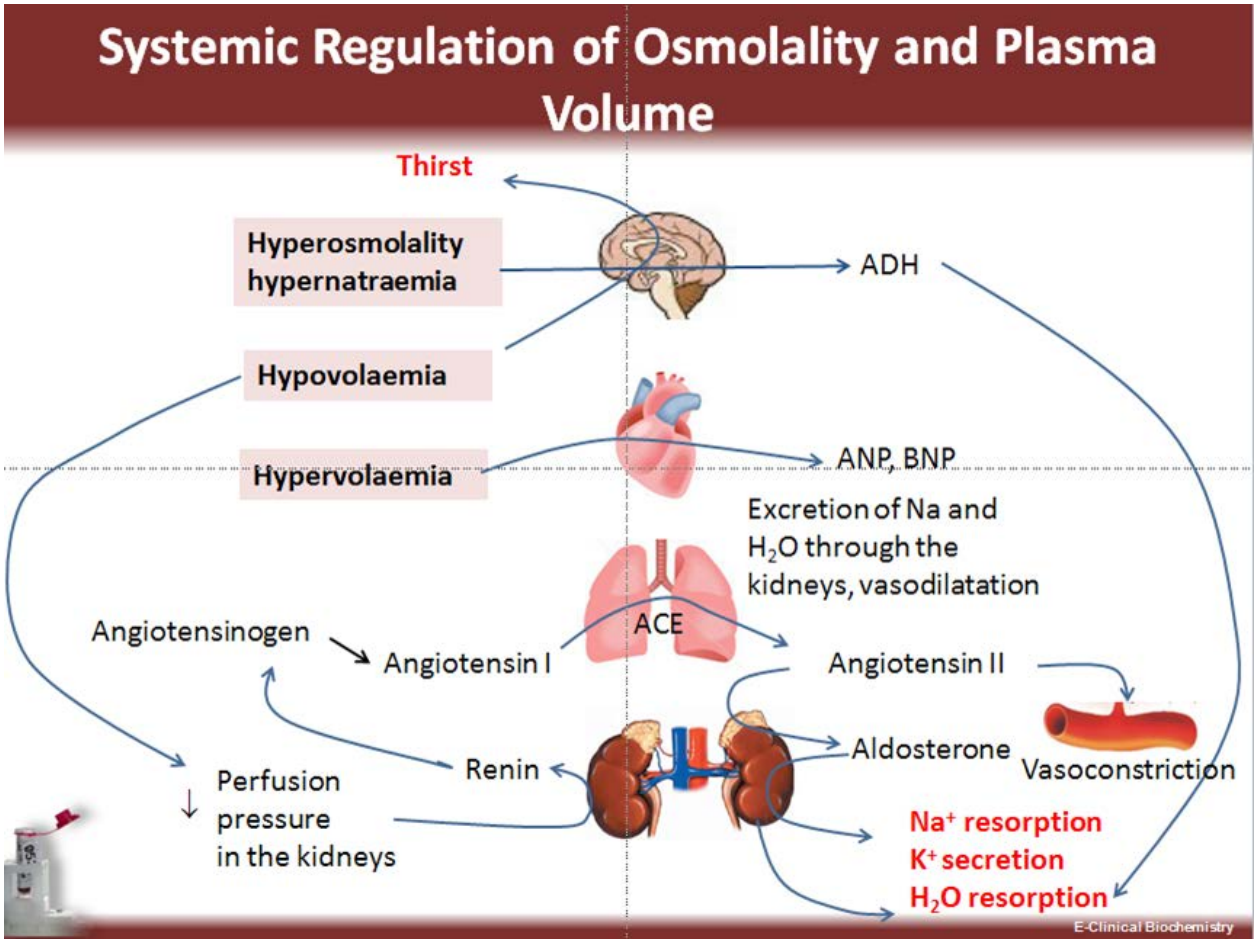


Figure 17.1. The effects of changes in osmolality and intravasal volume on Na^+ and water metabolism. (Clinical Chemistry: Techniques, Principles, Correlations. Sixth edition. M.L. Bishop, E.P. Fody, L.E. Schoeff. Wolters Kluwer/Lippincott Williams&Wilkins, 2010)

Intravasal volume changes are primarily detected by a series of tension receptors located at various points of the cardiopulmonary circulation, carotid sinus, aortic arch and glomerular arterioles. Signals from these receptors activate many responses aimed at adjusting intravasal volume through the adjustment of vascular resistance, cardiac output and renal retention or excretion of Na^+ and water.

- Additional factors affecting intravasal volume:

Natriuretic peptides: *atrial natriuretic peptide (ANP)* and *brain natriuretic peptide (BNP)* released from the atria and ventricles of the heart in response to the expansion of intravasal volume support Na^+ excretion in the kidneys; BNP and ANP co-act in regulating blood pressure and water balance.

Volume receptors also stimulate the release of the ADH independently of osmolality, which saves water through renal re-absorption.

Glomerular filtration increases and decreases with increasing or decreasing intravasal volume.

A 1 - 2% decrease in *tubular re-absorption of Na^+* may increase water losses by litres per day.

Urine osmolality has a broad range depending on water intake and the status of body hydration.

17.1.4. Water Deficit – Dehydration

Water deficit caused by insufficient intake or excessive losses leads to increased plasma osmolality. Both regulators, thirst and ADH, become activated. While ADH minimizes water losses, thirst is the major correcting factor. It is, however, inactive in unconscious people and may fail in elderly people, children, or in cases where drinking is impossible or it is otherwise impossible to secure access to water. The osmotic stimulation of thirst decreases considerably in people over 70.

Water deficiency, *dehydration*, leads to a more concentrated internal environment, reflected in the plasma by

haemoconcentration, i.e. an increase in the concentration of haemoglobin, proteins and urea, and an increase in the haematocrit level (an increase above the reference limits occurs only if baseline levels were within the reference limits; previously decreased levels can only be „adjusted“). Isotonic, hypertonic and hypotonic dehydration are distinguished based on the osmolality level.

17.1.4.1. Isotonic Dehydration

Isotonic dehydration is caused by a loss of a fluid that is isotonic with blood plasma (vomiting, diarrhoea, burns). This loss is not accompanied by a rise in ECF osmolality, and so the compensatory transfer of water from the ICF does not occur. There are signs of haemoconcentration, and skin turgor and diuresis are reduced. Lowered blood volume stimulates renin and aldosterone secretion. This increases sodium absorption from the distal tubule, which causes low sodium concentration in the urine. Given the relatively low ECF volume, hypovolaemia and the centralization of blood circulation and its failure may occur rapidly.

17.1.4.2. Hypertonic Dehydration

Hypertonic dehydration is caused by a lack/loss of fresh water. The most common cause is a reduced supply of water (unconsciousness, reduced sense of thirst, inability to swallow) or inadequately compensated losses (polyuric phase of renal failure, central or nephrogenic diabetes insipidus). Osmolality and Na^+ concentration in the ECF increase in addition to signs of haemoconcentration. Hyperosmolality of the ECF leads to the transfer of water from the ICF in an effort to correct the disturbed osmotic balance. Due to double the content of water in the ICF, the disorder may be corrected at first in the ECF, thereby delaying the clinical signs of hypovolaemia.

17.1.4.3. Hypotonic Dehydration

Hypotonic dehydration is less frequent. It occurs if fluids with relatively higher Na^+ content than that in the ECF are lost (pancreatic fistula, some types of diarrhoea). Relative hyperosmolality of the ICF leads to the transfer of water to this space and a deepening of existing hypovolaemia.

17.1.5. Excess of Water – Hyperhydration

Excessive intake or insufficient excretion of water leads to a decrease in osmolality.

Both regulators (thirst and ADH) are dampened. In the absence of ADH, water is not reabsorbed from the primary urine, and a large amount of diluted urine is produced. Normally functioning kidneys are so effective that even hypo-osmolality and hyponatraemia need not necessarily occur (beer drinkers).

In pathological situations, isotonic, hypotonic or hypertonic hyperhydration may occur.

17.1.5.1. Isotonic Hyperhydration

Isotonic hyperhydration occurs due to the accumulation of isotonic fluid leading to an increase in the ECF volume without a change in its osmolality. This disorder occurs in patients with cardiac failure, decompensated cirrhosis and nephrotic syndrome. These conditions are accompanied by secondary hyperaldosteronism with sodium and water retention. Signs of haemodilution (a drop in proteinaemia, haemoglobin and haematocrit) are present, osmolality and Na^+ concentration are in the reference range, and the Na^+ reserve is usually increased. Oedemas or ascites are present.

17.1.5.2. Hypotonic Hyperhydration

Hypotonic hyperhydration accompanies conditions where the ability to excrete excessive intake of water is limited, for example in renal failure and in cardiac patients. Capillary leakage of water leading to secondary hyperaldosteronism sometimes occurs in polytraumas, and so it is easy to hyperhydrate patients at the beginning of the therapy.

The laboratory signs of haemodilution, hyponatraemia and hypoosmolality are present.

Patients with the syndrome of inappropriate antidiuretic hormone secretion (*SIADH*) form a special group. SIADH may accompany some inflammatory and tumour diseases of the lungs and CNS, post-traumatic and post-operative conditions. These conditions feature increased ADH production not induced by increased osmolality or hypovolaemia.

A similar clinical and laboratory presentation is typical of the cerebral salt wasting syndrome (*CSWS*), caused by the pathological overproduction of natriuretic peptide. Hyponatraemia and hypoosmolality are present in both cases.

While signs of dehydration can be found in CSWS, in SIADH the intravascular volume is normal or only slightly elevated.

17.1.5.3. Hypertonic Hyperhydration

Hypertonic hyperhydration tends to be iatrogenic in origin, and develops by making up for pure water losses with saline (154 mmol Na^+/L), by administering a hypertonic solution, drinking seawater, etc.

The kidneys are forced to react to extreme situations in the water balance, which is made possible by their ability to produce urine in a broad range of osmolality (50 – 1500 mmol/kg).

17.1.6. Causes of hyperosmolality

- Hypertonic dehydration (thirst, central or nephrogenic diabetes insipidus)
- Hyperglycaemia → hyperosmolar diabetic coma
- Low-molecular substance poisoning (e.g. 1‰ ethyl alcohol = 23 mmol/kg, methanol, ethylene glycol)

During therapy for effective ECF osmolality disorders, the principle of carefully controlled, gradual adjustment should be adhered to (a maximum change in serum osmolality by 2 - 4 mmol/kg per hour), in particular with respect to the risk of damage to the CNS.

Hypoosmolality of the ECF leads to the penetration of water into the relatively hyperosmolar environment in the CNS to form cerebral oedema and an increase in intracranial pressure with neurological symptoms (headaches, disorientation, lethargy, vomiting) and a risk of brainstem herniation. To compensate this, the Na^+ and K^+ content in brain cells decreases within 24 hours, and the content of organic substances (polyols) decreases within 48 hours, with a drop in effective osmolality.

ECF hyperosmolality and too fast an adjustment to previous ECF hypoosmolality leads to brain dehydration with a dramatic reduction in brain cell volume, and demyelization and damage to brain capillaries. Clinical signs are also neurological.

17.1.7. Measurement of Osmolality

Blood plasma **osmolality** is determined using osmometers mostly based on freezing point depression measurement. Osmometers based on the cryoscopic principle are equipped with highly sensitive thermistors which measure with an accuracy of one thousandth of a degree centigrade. A -1.86°C drop in the freezing point is caused by 1 mol dissolved undissociated substance.

The range of blood serum osmolality in healthy people is 285 ± 10 mmol/kg. More than 90% of total plasma osmolality is attributable to the Na^+ cation and its associated Cl^- and $HCO3^-$ anions. The contribution of other cations and anions is low. As regards non-electrolytes, a major contribution is made by glucose and urea. Since urea readily penetrates the intracellular space and does not cause an osmotic gradient between the ECF and the ICF, it is not regarded as a substance with any effective influence on osmolality intended for therapeutic correction.

In clinical practice, blood plasma osmolality is also estimated by calculating from molar concentrations (mmol/L) of osmotically important components of the plasma. The concentration of Na^+ and the corresponding anions, urea, glucose and to a small extent proteins, are decisive for the value of plasma (serum) osmolality. There are many empirical formulas, such as:

$$\text{plasma osmolality (mmol/ kg } H_2O) \approx 2\,[Na^+] + [\text{glucose}] + [\text{urea}]$$
(all in mmol/l)

These calculations are made even if the value of osmolality has been found by measurement. The measured osmolality value is usually about 5 - 10 mmol/kg higher than the calculated value. Where the difference between the approximate calculation and the measured value (osmolal gap) is larger, it warns of the presence of an increased amount

of low-molecular substances that normally do not occur in the plasma, such as alcohol, acetone or ethylene glycol. If the difference between the measured and the estimated osmolality is greater than 10 mmol/kg, the presence of these substances in the plasma is highly probable. For example, 1‰ alcohol will increase osmolality by 23 mmol/kg. The presence of volatile artificial substances can be proved only by the cryoscopic measurement method, not by measurement on the principle of water vapour depression. One should keep in mind when evaluating the osmolal gap that the administration of osmotically active substances such as mannitol will also increase plasma osmolality.

Percentage - wise, a small but important contribution (~0.5% of the total osmotic pressure) is made by the molecules of proteins, albumin in particular. This property of proteins is referred to as **colloid osmotic (oncotic)** pressure, and this pressure is important for transfers of water between the plasma and the interstitium. High-molecular substances do not penetrate from the venous space to the interstitium in normal circumstances, thus they act against hydrostatic pressure and contribute to the maintenance of intravasal volume. A dramatic worsening of this function accompanies shock conditions with permeability disorder and albumin leakage into the interstitium.

Values < 250 mmol/kg and > 325 mmol/kg are regarded as critical serum osmolality values.

17.1.8. Importance of Osmolality Determination

17.1.8.1. Serum

- Differential diagnostics of water and ion balance system disorders
- Establishing the seriousness and level of hyperosmolal conditions
- Control and monitoring the decrease in osmolality in treatment for hyperosmotic conditions
- Finding the osmolal gap from the difference between measured and calculated osmolality

17.1.8.2. Urine

- Monitoring osmotic load excretion in hypercatabolic conditions
- Determining renal concentrating ability in the concentrating experiment (adiuretin test)

17.1.8.3. Serum/urine

- Determining the osmolal index (U_{Osm} / S_{Osm}) in the differential diagnosis of renal and prerenal disorders
- Calculation of osmolal clearance (Cl_{Osm}) and solute-free water clearance (Cl_{H2O})

CHAPTER 18

18. Serum Lipids and Lipoproteins, Relation to Atherogenesis

Author: doc. MUDr. Vladimír Soška, CSc.

Reviewer: MUDr. Tomáš Franěk

18.1. Lipids

Lipids include hydrophobic substances (Tg, FAs, esterified cholesterol) and substances simultaneously both hydrophobic and hydrophilic (phospholipids, free cholesterol). They can be divided into four groups: FAs, simple lipids, complex lipids and isoprenoid lipids.

18.1.1. Fatty Acids (FAs)

Fatty acids are substances with a simple carbon chain whose molecule may contain one or more double bonds. FAs in animal tissues only have an even number of carbons, usually 16 - 22 carbons. They are supplied from food and are also synthesized in the liver. In the blood they are transported while bound to albumin (free FAs) or in the form of complex lipids in lipoproteins. They are the main source of energy for cardiac and skeletal muscles. They are cleaved (beta-oxidation) in mitochondria, to which they are transported while bound to carnitine. Beta-oxidation may produce ketone bodies, an important source of energy. Free FAs in the blood have a very short biological half-life (about 2 minutes); they are taken up in the liver and re-esterified to Tg.

18.1.1.1. Saturated FAs

Saturated FAs do not have any double bond in the chain, and are primarily a ready source of energy. They can be synthesized in the liver from other energy substrates such as glucose, amino acids and alcohol, and can be stored in the form of Tg as an energy reserve. Physiologically important saturated FAs are lauric acid (C12:0), myristic acid (C14:0), palmitic acid (C16:0), stearic acid (C18:0) and arachidic acid (C20:0). Their primary source is animal fats (they are accompanied by cholesterol) and also some vegetable fats such as palm oil. They are relatively stable even at high temperatures and do not oxidize easily.

Relation to atherosclerosis: they are a rich source of energy and an increased intake of FAs supports the development of obesity. In addition, a high consumption has proatherogenic effects, leading to an increase of LDL cholesterol level in the blood (probably through the induction of elevated cholesterol synthesis in the liver). An exception here is stearic acid, which can be changed in the body into unsaturated oleic acid. It is recommended to reduce saturated FAs consumption to prevent atherosclerosis.

$$CH_3 - (CH_2)_n - COOH$$

Figure 18.1. Saturated FA structure

18.1.1.2. Unsaturated fatty acids

These FAs contain 1 or more double bonds in their chains, physiologically usually in the cis configuration. They are contained mainly in vegetable fats and fish oils.

18.1.1.3. Monounsaturated fatty acids

The main representatives are palmitoleic acid (C16:1, ω-7) and oleic acid (C18:1, ω-9). They can be synthesized in the liver and so are not essential; nonetheless, an exogenous supply of oleic acid is recommended, because oleic acid synthesis in the body is usually insufficient. Relation to atherosclerosis: if supplied sufficiently, oleic acid can lower the LDL-cholesterol level and is considered a part of the Mediterranean diet reducing the risk of atherosclerosis development.

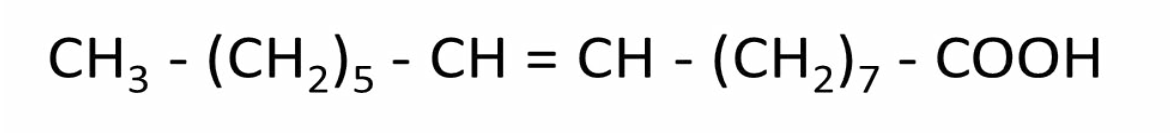


Figure 18.2. Example of monounsaturated FAs structure

18.1.1.4. Polyunsaturated fatty acids

They have multiple double bonds. The first bond is usually on the third (ω-3) or sixth (ω-6) carbon from the carboxylic tail. Among other things, they are used as a source for the synthesis of prostaglandins, leukotrienes and thromboxanes. They are less stable than saturated FAs and oxidize more easily.

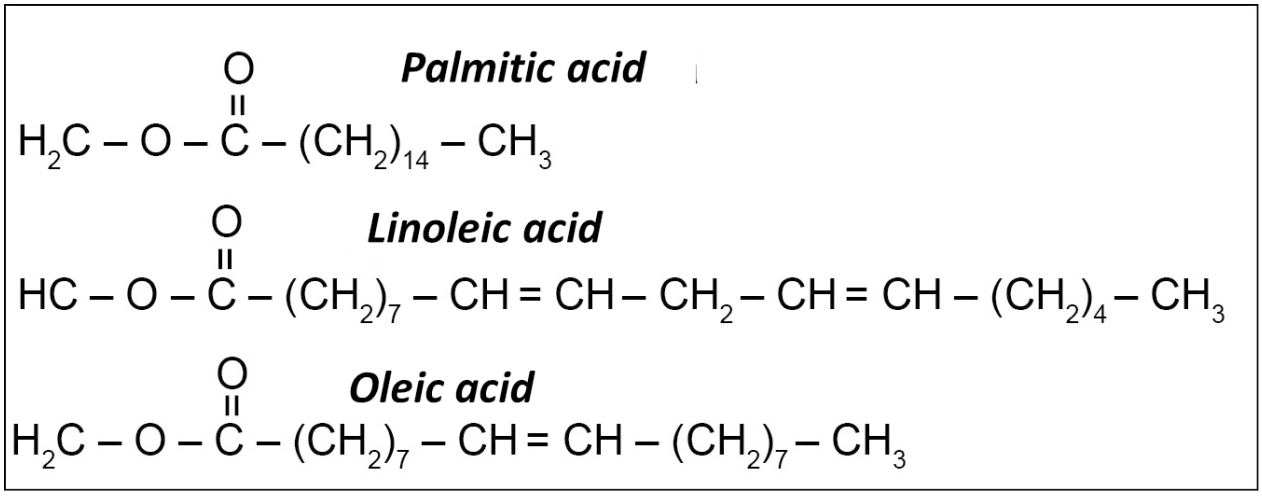


Figure 20.3. Some examples of the most important fatty acids

ω-6 fatty acids

The main representatives are linoleic acid 18:2 (9,12), γ-linolenic acid 18:3 (6,12,12) and arachidonic acid 20:4 (5, 8, 11, 14). They are a source for the synthesis of those eicosanoids which have pro-inflammatory and vasoconstrictive effects and increase thrombocyte aggregability. A high consumption may reduce the HDL-cholesterol level.

ω-3 fatty acids

Their main representatives are linolenic acid 18:3 (9,12,15), eicosapentaenoic acid 20:5 (5,8,11,14,18) and docosahexaenoic acid 22:6 (4,7,10,13,16,19). They are a source for the synthesis of those eicosanoids which have vasodilative and anti-inflammatory effects and decrease thrombocyte aggregability. If supplied in sufficient amounts, they lower the Tg level and act in preventing the development of atherosclerosis complications. They are contained primarily in fish oils, and to a lesser extent in vegetable oils.

18.1.1.5. Trans fatty acids

These are isomers of unsaturated FAs and their spatial arrangement is similar to saturated FAs (straight chain). They are produced from cis isomers FAs by the action of bacteria in the bowels of ruminants, from where they arrive at all meat and milk products. They can also be formed during the hydrogenation of unsaturated FAs (the production of “hard” margarines for baking and frying; soft margarines for direct consumption are produced using different

technologies and do not contain trans isomers of FAs). They are also produced if unsaturated FAs are exposed to high temperatures for an extended period, during frying, for example.

Therefore, normally consumed foodstuffs of animal origin, bread and pastry, and various “delicacies” with icing and chocolate substitutes are the main source of trans FAs from food. Trans FAs may have pro-atherogenic effects – an increased intake may accelerate atherosclerosis and raise the LDL-cholesterol level.

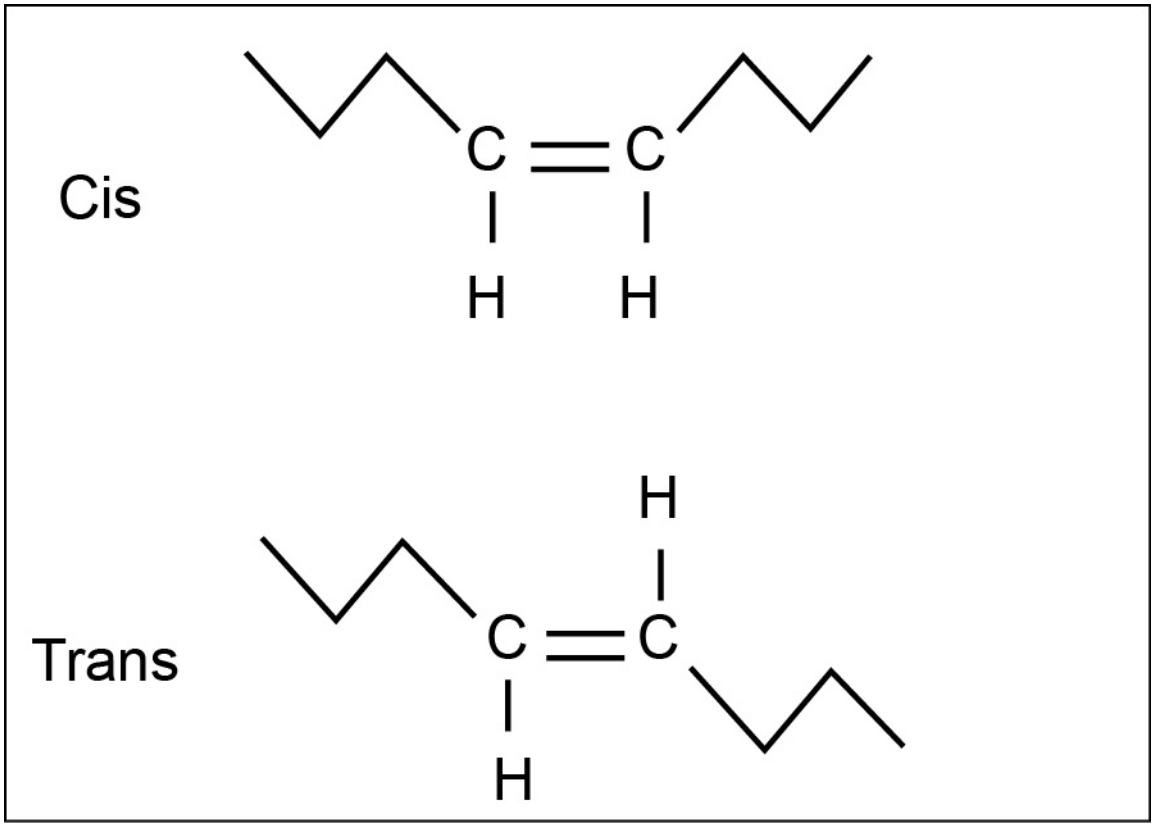


Figure 18.3. Trans and cis FAs structure

18.1.2. Simple Lipids

The group of simple lipids includes acylglycerols and waxes. Acylglycerols are produced by FAs bonding to trihydric alcohol glycerol, and waxes are produced by FAs bonding to primary or secondary aliphatic alcohol.

18.1.2.1. Triacylglycerols (Tg)

Triacylglycerols are produced by bonding 3 FAs to trihydric alcohol – glycerol (glycerol triesters). The first Tg carbon preferentially binds palmitic acid, the second carbon linoleic acid and the third carbon oleic acid (if these FAs are available). Tg are of exogenous (about 80 – 170 mmol/day are consumed in food) and endogenous origin (synthesis in the liver). Tg are cleaved in the fatty tissue by the action of hormone-sensitive lipase. This enzyme is inhibited by insulin, and so it is active only if insulinaemia decreases (during fasting). Increased Tg synthesis in the liver (an increased fasting concentration of Tg in the blood) is usually caused by an excessive intake of energy (carbohydrates, fats, alcohol) in food. Tg are the main reserve form of energy: a non-obese adult has about 15 kg of Tg. With a content of about 38 kJ of energy per gram of weight, this corresponds to about 570,000 kJ, which is a reserve of energy for almost 3 months of fasting. Tg are transported in the blood in all lipoproteins, but their highest content is in the VLDL and chylomicrons.

The physiological concentration of Tg in the blood is under 1.7 mmol/L, provided the patient has been fasting for 10 - 12 hours prior to blood collection. An elevated Tg concentration in the blood is a marker of increased risk of atherosclerosis and cardiovascular diseases. A very frequent cause of an elevated Tg level is excessive alcohol consumption. A decrease of Tg concentration in the blood can, in most cases, be simply achieved by limiting the energy (and alcohol) intake, irrespective of the food composition, and also by increased physical activity. Hyperinsulinaemia (e.g. in patients with type 2 diabetes mellitus or obese people) inhibits hormone-sensitive lipase in the fatty tissue, thus inhibits the splitting of Tg. This hormone becomes activated only if insulinaemia decreases – when restricting energy intake.

18.1.3. Complex Lipids

Complex lipids, which consist of polyhydric alcohol with the bound FAs and other components, and one of the basic structural substances of the body. Their molecules have hydrophobic and hydrophilic tails, which is decisive for their physiological functions – they are the basic building blocks of all cell membranes. Cell membranes are characterized by their double layer, with the outer layer consisting of choline phospholipids and glycolipids, and the inner layer consisting of aminophospholipids. Assays for measurement of phospholipids or glycolipids in the blood are not available, as their blood levels are not related to atherogenesis.

18.1.3.1. Phospholipids

They are a basic component of all cell membranes at the cellular and sub-cellular level, and also form the surface structure of blood lipoproteins. All cells are able to synthesize phospholipids. They are divided into glycerophospholipids and sphingophospholipids.

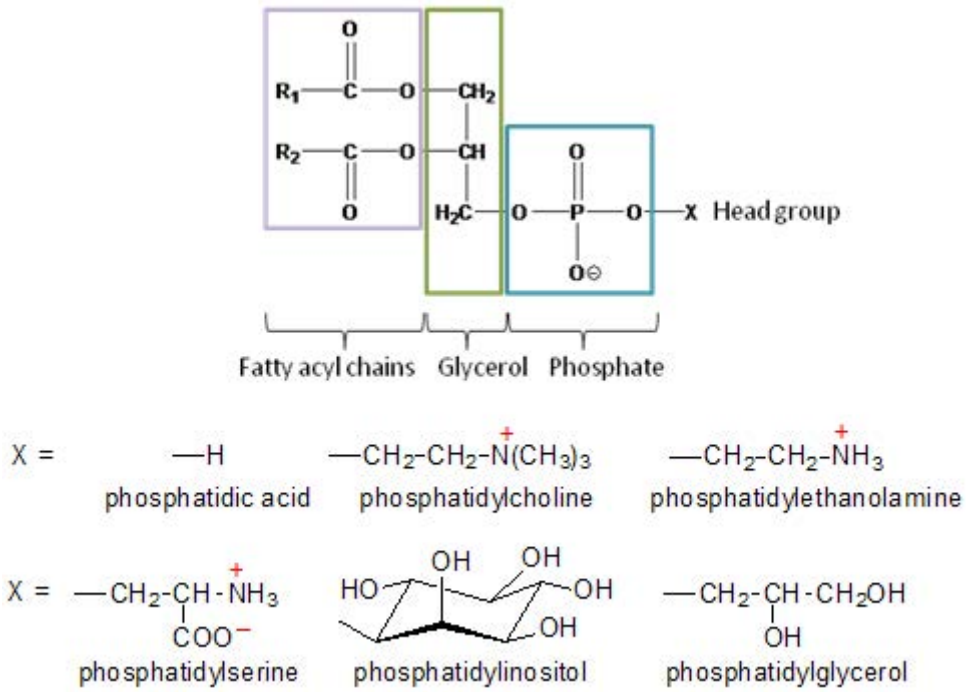


Figure 18.4. Structure of phospholipids

Glycerophospholipids

They are derived from Tg, where one fatty acid is replaced with phosphoric acid residue and another substance (choline, ethanolamine, serine). One representative phospholipid in the outer cell membrane layer is phosphatidylcholine (sometimes also called „lecithin“), while the inner layer of cell membranes is formed usually by phosphatidylserine.

Relation to atherogenesis: has not been conclusively proved.

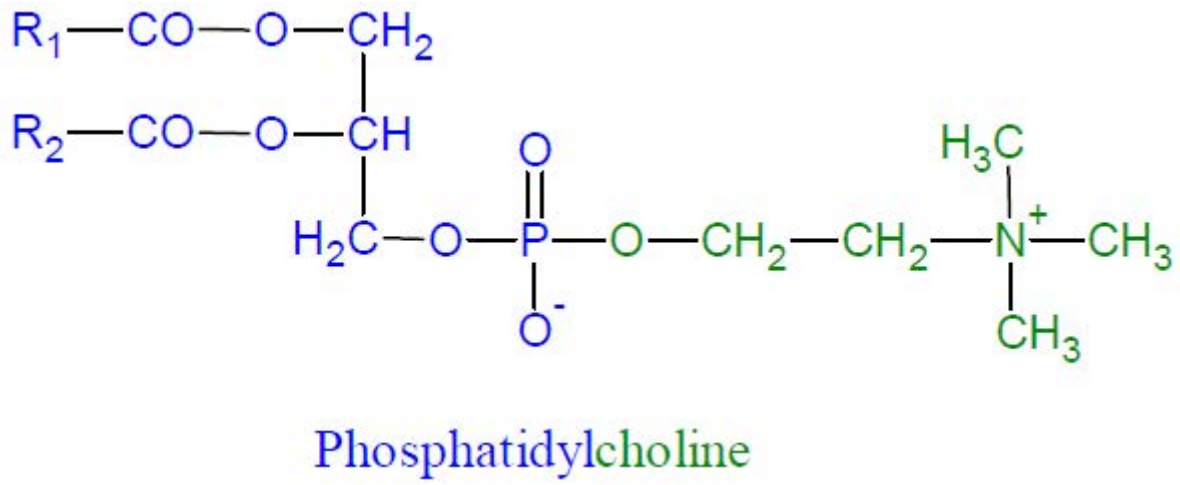


Figure 18.5. Structure of glycerophospholipids

Sphingophospholipids

A typical representative is sphingomyelin, which contains the amino alcohol sphingosine, 2 FAs, phosphoric acid residue and choline. Sphingosine is the main membrane phospholipid at the sub-cellular level and is also a dominant phospholipid in the nerve fibre myelin. It is also part of surface structures of lipoproteins.

18.1.3.2. Glycolipids

Glycolipids contain sphingosine to which is bound FAs and a saccharide component. They are components of cyto-plasmatic membranes, neuron sheaths and many other organic structures.

18.1.4. Isoprenoid Lipids

18.1.4.1. Cholesterol

Cholesterol is one of the basic biological compounds: it has a key role in the structure of membranes at the cellular and sub-cellular levels; it is a starting substance for the synthesis of steroid hormones, bile acids and vitamin D; and, it is required for the absorption and transport of Tg and fat-soluble vitamins. Practically all cells in the body are able to synthesize it; intensive synthesis takes place in hepatocytes, enterocytes, neurons and steroid hormone-producing endocrine glands. The enzyme 3-hydroxy-3-methyl-glutaryl-CoA (HMG-CoA) reductase is crucial for the regulation of cholesterol synthesis as it catalyzes the conversion of hydroxy-methyl-glutaryl-CoA to mevalonate. The activity of this enzyme is inhibited by statins – blood cholesterol-lowering drugs. The intensity of cholesterol synthesis in the cell is regulated by feedback at the HMG-CoA reductase level: a sufficient reserve of intracellular cholesterol inhibits this enzyme, while a lack of cholesterol activates it. This enzyme's activity is also stimulated by insulin, for example, and inhibited by glucagon. Another mechanism regulating intracellular reserves of cholesterol is the regulation of LDL--receptor synthesis: if supply of cholesterol is deficient, the synthesis of receptors increases (increased uptake of LDL lipoproteins from the blood), while high supply of cholesterol inhibits LDL-receptor synthesis. Cholesterol synthesis in the liver is often increased in obese people and following an increased intake of unsaturated FAs from food. Cholesterol in the blood is transported in lipoproteins, eliminated from the body in bile (both in an unchanged form and in the form of bile acids), and about 50% of this excreted cholesterol and about 95% of bile acids are re-absorbed in the intestine.

Free Cholesterol

Free cholesterol is partly hydrophilic and is a basic component of cell membranes at the cellular and sub-cellular levels. It is transported in the blood on the surface of lipoproteins.

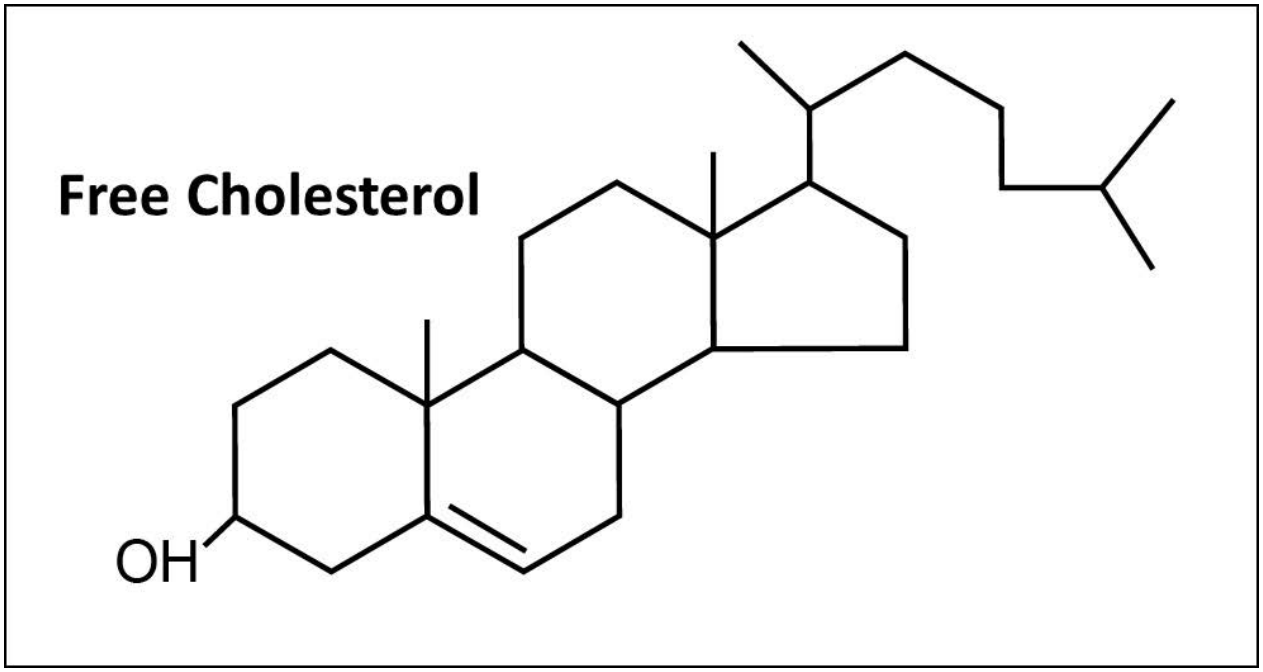


Figure 18.6. Structure of free cholesterol

Esterified Cholesterol

Esterified cholesterol has a bound fatty acid, and it is a reserve form of cholesterol in cells. Cholesterol esterification takes place intracellularly by the catalysis effect od acetyl-CoA cholesterol acetyltransferase, whose activity increases when cholesterol supply is sufficient and decreases when cells are lacking cholesterol reserves. In addition, cholesterol esterification takes place also in blood lipoproteins under catalysis of lecithin-cholesterol acyltransferase. Linoleic and linolenic acids are preferentially bound to cholesterol in lipoproteins, and oleic and palmitoleic acids preferentially within cells.

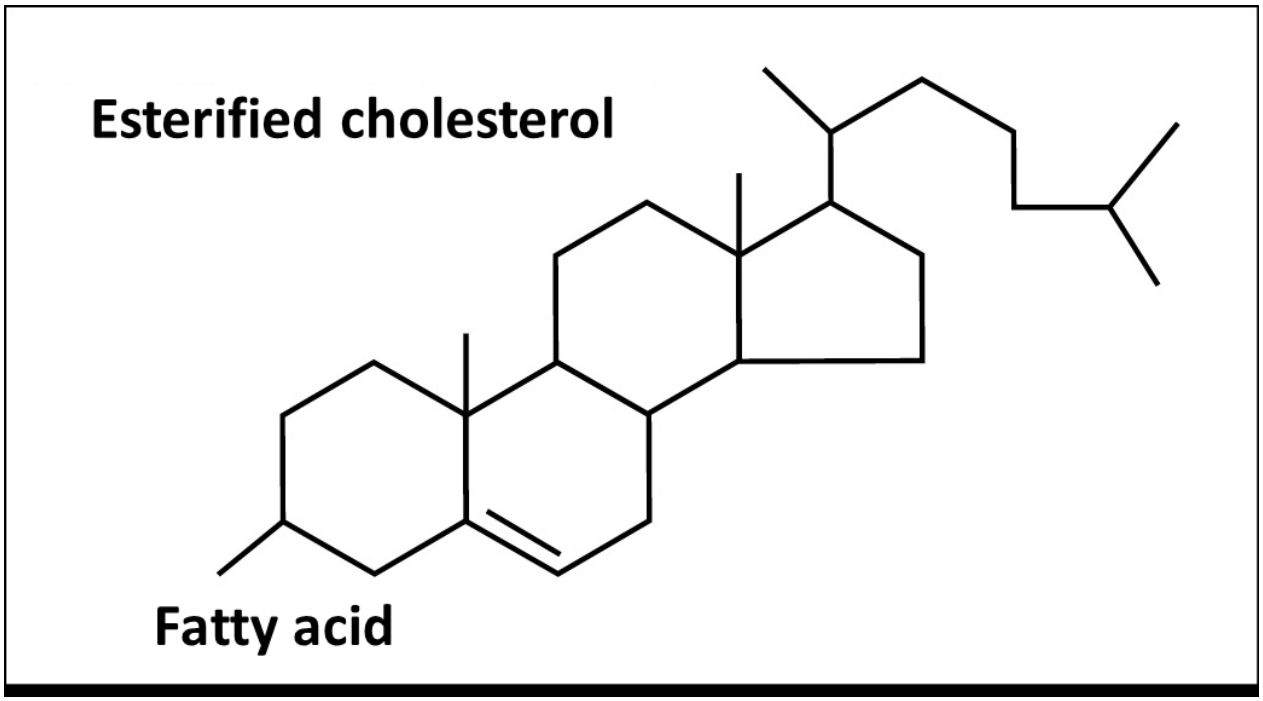


Figure 18.7. Structure of esterified cholesterol

Relation of Cholesterol to Atherogenesis

An elevated blood cholesterol level is a main risk factor for the development of atherosclerosis and cardiovascular diseases. Cholesterol is always a basic part of the atherosclerotic plaque (there is no atherosclerosis without choleste-

rol). As the cholesterol level grows, starting from a total cholesterol concentration as low as 3.9 mmol/L, so the risk of cardiovascular complication also rises. Average cholesterol levels in the adult Czech population range from 5.6 to 5.8 mmol/L. Total cholesterol level values under 5.0 mmol/L are considered optimal for the general population, and this value should also be indicated as the "physiological" upper limit for cholesterolaemia in laboratory results of cholesterol level measurement. Lower cholesterol levels are recommended for people with a high risk of cardiovascular complications under primary prevention (under 4.5 mmol/L). The optimum total cholesterol level in the blood for patients with cardiovascular diseases is under 4.0 mmol/L. Intervention studies have shown that a 1% decrease in the blood cholesterol level leads to about a 2% decrease in the incidence of ischaemic heart disease. A dramatic decrease in the cholesterol level may even lead to the regression of atherosclerosis.

18.2. Lipoproteins

Lipoproteins are spherical particles transporting lipids in the blood. Table 1 shows an overview of the basic lipoprotein classes. Chylomicrons and VLDL are synthesized *de novo*. IDL and LDL are products of VLDL catabolism. HDL production and metabolism are described in the relevant section of this chapter. An intensive metabolism of lipoproteins takes place in the blood, where besides their degradation, practically all their components (cholesterol, Tg, phospholipids, apolipoproteins) are exchanged and transferred between the lipoprotein types. Lipoprotein(a) is an atypical lipoprotein. It takes part in the fibrinolysis process and is described in the chapter Laboratory Markers of Atherosclerosis. Table 1 shows the relation of each lipoprotein class to atherosclerosis.

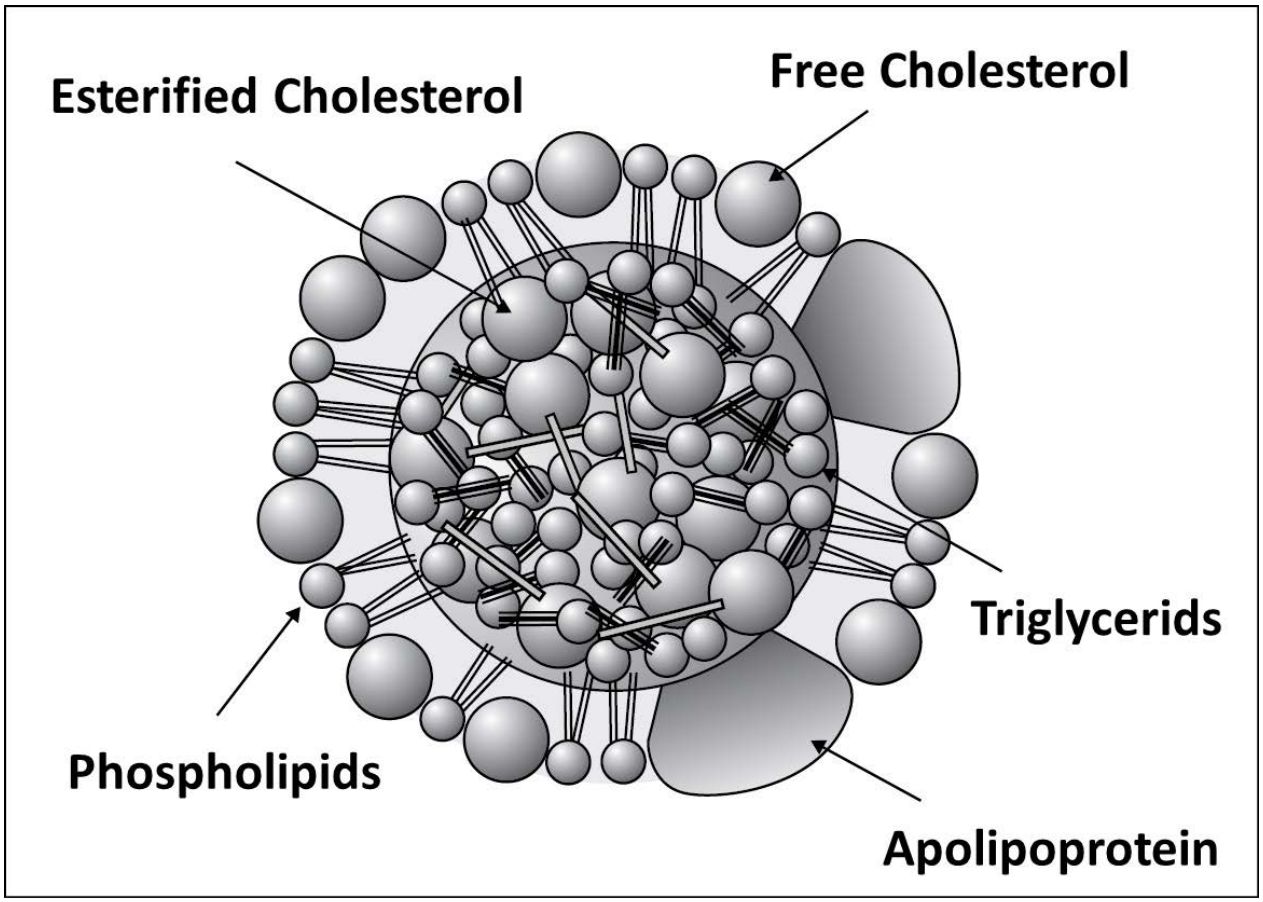


Figure 18.8. Shape and structure of lipoproteins

Lipoprotein Class	Acronym	Risk of Atherosclerosis
Chylomicra	CL	No effect
Chylomicron remnants	CL remnants	Increase (++)
Very low density lipoproteins	VLDLs	Increase (+)
Intermediate-density lipoproteins	IDLs	Large increase (++)
Low-density lipoproteins	LDLs	Large increase (++)
High-density lipoproteins	HDLs	Decrease (-)

Table 18.1. Relation of lipoproteins to atherosclerosis

18.2.1. Chylomicrons

Chylomicrons are synthesized in enterocytes (along with apolipoprotein B48). Their physiological function is to transport Tg from the intestine to peripheral tissues (fatty tissue in particular) and transport cholesterol from the intestine to the liver. They are partially degraded in blood capillaries of the peripheral tissues (fatty and muscle tissues in particular), and released Tg are hydrolyzed by lipoprotein lipase effect. This enzyme is localized on the blood capillary endothelium, mostly in skeletal muscles, myocardium, fatty tissue and the mammary gland during lactation. The particle gradually diminishes, and when about 70% of Tg have been split off, chylomicrons become „chylomicron remnants“ which are relatively rich in cholesterol. They are bound very rapidly, mainly in the liver, through the pathway of receptors for apolipoprotein E. The remnant particles are a source of exogenous cholesterol for hepatocytes. The highest concentration of chylomicrons in the blood is usually 3 - 6 hours after a meal, and they should not be present in the blood after about 9 hours of fasting.

Atherogenicity: chylomicrons do not have any atherogenic potential; their remnants, however, are strongly atherogenic (but they have a very short biological half-life in physiological circumstances).

18.2.2. VLDLs

VLDLs are synthesized in hepatocytes. The rate of apolipoprotein B100 synthesis is a limiting factor for their secretion into the blood. VLDLs transport endogenous Tg (synthesized in the liver) to peripheral tissues (fatty and muscle tissues in particular). While in the blood, VLDLs take over additional apolipoproteins (apo E, apo C) from HDL particles. In the blood capillaries of peripheral tissues (muscle and fatty tissues in particular) are VLDLs partially degraded during the hydrolysis of Tg by the action of lipoprotein lipase, that is localized on the endothelium of capillary vessel in the tissue with a high consumption of FAs. Lipoprotein lipase is activated by apolipoprotein C-II and also by insulin. The VLDL particle takes over esterified cholesterol from the HDL particles in exchange for Tg. The VLDL particle is thereby depleted of Tg and enriched with cholesterol esters, they decrease and becomes the IDL particle (VLDL remnant). The biological half-life of VLDL is about 2 - 4 hours.

Atherogenicity: increased VLDL concentration slightly increases the risk of cardiovascular diseases. Their presence in the blood (along with chylomicron remnants) is involved to elevated Tg level.

18.2.3. IDLs

IDLs are a product of VLDL catabolism and contain approximately the same molar content of both cholesterol and Tg. They may be metabolised in two ways: a) when flowing through the liver, they are taken up by hepatocytes using the receptor for apolipoprotein E (carried on the IDL surface) and degraded in the liver; b) when flowing through the liver, they are exposed to the effect of hepatic lipase, which hydrolyzes most of Tg remaining in them to form the LDL particle. In physiological circumstances, IDLs have a very short biological half-life and their concentration normally does not exceed 1/10 of the LDL concentration.

Atherogenicity: IDLs are highly atherogenic and may markedly accelerate the development of atherosclerosis.

18.2.4. LDLs

LDLs are a product of VLDL and IDL catabolism in the blood and carry apolipoprotein B100 on their surface, which is required for the LDL bond to the LDL-receptor on the surface of cells. Their function is to supply cholesterol to all peripheral tissues – practically all cells express the LDL-receptor on their surface. The entry of cholesterol into cells through the pathway of LDL-receptors is regulated by feedback. LDLs are the main source of cholesterol, that is primarily used for cell membrane synthesis. About 30 – 40% of LDL particles are usually catabolized within 24 hours, about 80% of these through the LDL-receptor pathway. LDLs have a long biological half-life and a part of them is subject to chemical modification such as glycosylation and/or oxidation. Such modified LDLs lose their ability to bind to LDL-receptors and are catabolized through the pathway of scavenger receptors, particularly by macrophages in sub-endothelial spaces. The entry of modified LDLs into cells through the pathway of scavenger receptors is not regulated by feedback; macrophages in sub-endothelial spaces therefore accumulate cholesterol uncontrollably and gradually change into foam cells that form the basis of the atherosclerotic plaque. There are several LDL subtypes, differing in size, density and composition (content of cholesterol and Tg): large LDL_1, medium-sized LDL_2 and small LDL_3. The small LDL_3 are produced in the blood of patients with high level of Tg a result of the degradation of originally very large (atypical) VLDL particles. These are synthesized in the liver due to excess energy intake (obesity) in type 2 diabetes mellitus and in patients with metabolic syndrome.

Atherogenicity: LDLs are highly atherogenic; LDL atherogenicity increases with diminishing size (the most atherogenic LDLs are the small LDL_3).

18.2.5. HDLs

These smallest lipoproteins have a high content of proteins (apolipoproteins and other proteins and enzymes). The main function of HDLs is the reverse transport of cholesterol from peripheral tissues and other lipoproteins back to the liver. Precursor of HDL is apolipoprotein A1, which is secreted from the liver into blood. Apolipoprotein A1 binds in the blood free cholesterol, phospholipids, and other apolipoproteins and enzymes to form discoid (nascent) HDL particles. Cholesterol in HDL particles is esterified by the action of LCAT (lecithin-cholesterol acyltransferase); esterified cholesterol is transported inside HDL particles to form larger, spherical HDL particle is then formed. This esterified cholesterol is exchanged at the same time for Tg between HDLs and Tg-rich lipoproteins (VLDLs, chylomicrons) by the action of CETP (the cholesterol-ester transfer protein). The HDL particles may have two potential fates:

- they are taken up by hepatocytes through the pathway of the HDL-receptor and hepatocytes excrete cholesterol from the HDL into the bile;
- Tg in the HDL are hydrolyzed by the action of hepatic lipase to form a smaller HDL particle, which returns to the process of reverse cholesterol transport.

Atherogenicity: HDLs protect against the development of atherosclerosis – they transfer excess cholesterol from the vessel wall (and from peripheral tissues and other lipoproteins) back into the liver. In addition, HDLs have many other „anti-atherogenic“ effects. For example, they protect LDL particles from oxidative modification (by the action of the enzyme paraoxonase on their surface), they stimulate nitric oxide synthesis in the endothelium and thus protect the vessel endothelium.

18.3. Apolipoproteins

Apolipoproteins are proteins which are part of the lipoprotein particle and have many functions in the lipoprotein metabolism: they are enzyme cofactors, ligands for receptors, they take part in the transport and exchange of substances between lipoprotein particles (cholesterol, Tg, phospholipids), and are the structural proteins of lipoproteins. Table

2 shows the characteristics of some main lipoproteins. Clinically significant lipoproteins which are determined and are most important for the diagnosis and treatment of patients with hyperlipoproteinaemia are apolipoprotein B100 and apolipoprotein A-I.

Apolipoprotein	Site of Synthesis	Contained in Lipoproteins:	Function
A I	Liver, intestine	HDL, CL	LCAT activator
A IV	Liver, intestine	HDL, CL	LCAT activator
B 48	Intestine	CL	Structural lipoprotein
B 100	Liver	VLDL, IDL, LDL	Ligands for LDL receptors
C I	Liver, adrenal gland	HDL, VLDL, CL	LCAT activator
C II	Liver, intestine	HDL, IDL, VLDL, CL	LPL activator
C III	Liver, intestine	HDL, IDL, VLDL, CL	LPL and hepatic lipase inhibitor
E	Liver, peripheral tissue	HDL, IDL, VLDL, CL	Ligands for LDL and apo E receptors

CL - chylomidrons; LCAT – lecithin-cholesterol acyltransferase; LPL - lipoprotein lipase

Table 18.2. Some important apolipoproteins and their functions

18.3.1. Apolipoprotein A-I (Apo A-I)

This polypeptide is bound reversibly to the surface of HDL and chylomicrons. When chylomicrons are metabolized in the blood stream, apo A-I is transported to HDL where it accumulates. It is a structural HDL apolipoprotein and LCAT (lecithin-cholesterol acyltransferase) enzyme cofactor, which is required for cholesterol esterification in HDL particles and thereby for reverse cholesterol transport. Apo A-I is also a ligand for the HDL receptors in hepatocytes. Apo A-I level is routine measured in biochemical laboratories. This assay is important to evaluate cardiovascular risks because it provides information about the amount of anti-atherogenic HDL particles in the blood (unless the blood also contains chylomicrons, i. e. unless the serum is chylous).

Atherogenicity: increased Apo A-I concentration reduces the risk of atherosclerosis.

18.3.2. Apolipoprotein B100 (Apo B)

This is the largest apolipoprotein (4536 amino acids). It is synthesized in hepatocytes (about 10 mg per day) and it is a structural component of VLDLs, IDLs and LDLs. It also serves as a ligand for the bond of LDL particle to LDL-receptor on the surface of all cells. Apo B synthesis is decisive for the rate of VLDL secretion from hepatocytes into the blood. Following VLDL secretion into the blood, the active part of the apo B100 chain is masked, so VLDL and IDL particles are not bound to the LDL-receptor. Only after the hydrolysis of Tg from VLDLs and IDLs (when the LDL particle is formed) the specific parts of the apo B100 chain are unmasked and the LDL particle can bind to LDL-receptors. Following internalization of the receptor - LDL complex, apo B100 is degraded in lysozymes. A mutation in the apo B100 gene (like mutations in the LDL-receptor gene) leads to a disease called „familial hypercholesterolaemia“. In both cases, LDL particles are not sufficiently taken up from the blood and a marked increase occurs in the LDL-cholesterol level in the blood. This causes a very early ischaemic heart disease. Concentration of Apo B100 is routine measured in biochemical laboratories.

Atherogenicity: apo B100 concentration is an indicator of the sum of the main atherogenic lipoproteins (LDLs, IDLs and VLDLs) in the blood. Therefore, an increased apo B100 level is a risk factor for cardiovascular diseases.

CHAPTER 19

19. Risk Factors for Atherosclerosis (Except Lipids)

Author: doc. MUDr. Vladimír Soška, CSc.

Reviewer: prof. MUDr. Richard Průša, CSc.

19.1. Markers of Inflammation

Atherosclerosis is characterized by chronic inflammatory process in the vessel wall and therefore it is not surprising that inflammatory markers are related to complications of atherosclerosis

19.1.1. C-Reactive Protein (CRP)

CRP is a clinically significant „acute phase protein", whose level increases particularly in systemic bacterial, infectious and tumour diseases. CRP levels in tens and often hundreds of mg/L are typical of systemic bacterial infections.

Clinical studies have shown, however, that a slightly increased CRP level in the range of 3 - 10 mg/L is associated with an increased risk of the manifestation of cardiovascular complications. Primarily, there is an increased risk of ischaemic heart disease, but the risk of ischaemic cerebrovascular accident and atherosclerosis in other locations also rises. Clinical studies performed on large sets have proved that people with slightly increased CRPs had higher cardiovascular morbidity and mortality rates. This relationship is valid not only for previously healthy people without cardiovascular complications (primary prevention), but also for patients with cardiovascular diseases (secondary prevention). CRP values under 3 mg/L are currently considered physiological, 3 - 6 mg/L indicate a borderline risk, and 7 - 10 mg/L indicate a elevated risk of cardiovascular diseases.

However, CRP is not a risk factor that induces atherosclerosis, it is just an indicator of risk (like other inflammatory markers). The CRP level is increased in people with major causal risk factors such as smoking, hypertension, obesity, metabolic syndrome, diabetes mellitus, etc. The CRP level decreases if a person manages to reduce the cholesterol level, lose weight, induce a regression of diabetes (overweight or obesity reduction), etc. In addition, a slight increase in CRP may not always be associated with an elevated risk of cardiovascular complications: the interpretation of a CRP result between 3 and 10 mg/L should be with cousion and individually, as a slight increase in CRP may be due to any inflammation in the body and may not have relationship to cardiovascular diseases (chronic diseases of the intestine, lungs, joints, etc.), or can be the result of a starting or subsiding acute inflammation.

19.1.2. Interleukin 6 (IL-6)

IL-6 is a cytokine which stimulates the synthesis of acute phase proteins in the liver, including CRP. IL-6 is also produced in atherosclerotic vessel wall lesions, particularly in macrophages and smooth muscle cells. Therefore, inflammatory (atherosclerotic) activity in the vessel wall is connected with an increased IL-6 level in the blood. Clinical prospective studies dealing with the risk of cardiovascular diseases have proved that an increasing IL-6 level also increases the risk of morbidity and mortality from cardiovascular diseases, particularly ischaemic heart disease. However, it is not yet clear yet whether there is a causal or only a mediating relationship between IL-6 and atherosclerosis. There are probably two mechanisms by which IL-6 may influence cardiovascular risk:

a) the process mediated by other cytokines and acute phase proteins (CRP, SAA, lipoprotein modifications);

b) the direct effect of IL-6 on endothelial cells.

Elevated levels of IL-6, however, should be interpreted (like CRP) very carefully, because it need not be related to cardiovascular complications, but to any other inflammatory diseases. In addition, the IL-6 level is often increased (like CRP) in patients with diabetes, dyslipidaemia and in smokers. Values ranging from 1.5 to 7.0 ng/L are considered

physiological.

19.1.3. Serum Amyloid A (SAA)

SAA is the common name for several proteins belonging to the family of acute phase reactants. It is also closely related to blood lipoproteins, particularly high-density lipoproteins (HDL). SAA is frequently associated with them – it is bound to their surface, most often to HDL_3 particles. In a chronic inflammatory condition, SAA increasingly binds to the HDL surface and displaces from binding to HDL enzyme paraoxonase-1. This enzyme provides HDL with antioxidant properties, which are required in preventing the development of atherosclerosis. The HDL particle loses its antioxidant properties becomes dysfunctional and may even acquire pro-atherogenic properties. Increased SAA levels are often found in people with obesity, insulinresistance, metabolic syndrome and diabetes mellitus. According to the results of clinical studies, increased SAA as a marker of cardiovascular risk is concurrent with an increase in another inflammatory marker of atherosclerosis, namely CRP.

19.1.4. Myeloperoxidase

Myeloperoxidase is an enzyme (haemoprotein) from the group of peroxidases produced by monocytes and neutrophils at the site of inflammation. When the inflammatory process in atherosclerotic plaques is active, the originally stable plaque is transformed by myeloperoxidase action into an unstable (vulnerable) plaque with a risk of plaque rupture and thrombus formation, causing artery lumen occlusion. Therefore, increased MPO activity in the blood is a sign of the risk of atherosclerotic plaque instability and also the risk of acute cardiovascular accident.

19.2. Markers of Haemostasis and Thrombosis

Acute complications of atherosclerosis (myocardial infarction, cerebrovascular accident) usually develop as a result of a thrombus adhering to a artery wall damaged by atherosclerosis. That is why the relationship between markers of haemostasis and thrombosis to atherosclerosis and its complications has been studied intensively. In addition to the factors listed below - lipoprotein (a), fibrinogen, factor VII, plasminogen activator inhibitor - other coagulation and prothrombotic factors may also take part in the atherogenesis process.

19.2.1. Lipoprotein(a) [Lp(a)]

Lp(a) is essentially the LDL particle, on which the surface of the addition of apolipoprotein B100 is bound one more protein - apolipoprotein(a). Apolipoprotein(a) is highly homologous with plasminogen; it competes with plasminogen for binding sites on plasmin, but does not activate it. Higher concentrations may thus block the fibrinolysis process. In addition, unlike the LDL particle, the Lp(a) particle is not catabolized through the LDL-receptor pathway. The Lp(a) level is usually genetically determined. It is often increased in patients with familial hypercholesterolaemia, but it is also increased in patients with chronic renal failure, nephrotic syndrome and post-menopausal women. Concentrations under 0.3 g/L (sometimes under 0.5 g/L) are usually considered physiological, but a part of the population has a very low or even non-measurable Lp(a) concentration.

Increased Lp(a) concentration is connected with an increased risk of arterial thrombosis. Epidemiological studies have shown that an increased Lp(a) concentration is an independent risk factor primarily for the development of early ischaemic heart disease. But it is very difficult to lower the Lp(a) level; most hypolipidemic agents (statins, fibrates, resins, ezetimibe) have practically no effect on the Lp(a) level. The only hypolipidemic agent which may lower the Lp(a) level is niacin (nicotinic acid). The measurement of Lp(a) concentration has not yet been standardized; the main problem is the existence of several tens of Lp(a) chain variants with different binding capacities for the antibodies used in diagnostic sets. With immunonephelometry method it is possible to measure even very low concentrations of Lp(a).

19.2.2. Fibrinogen

Fibrinogen is a coagulation factor: it is changed into fibrin by the action of thrombin, which is a key step in blood clotting. Commonly reported physiological concentrations are 2 - 4 g/L, but recent opinions suggest that the upper limit is only 3.5 g/L or thereabouts (levels around 4 g/L may be connected with a slight increased thrombophilic risk). Epidemiological studies have shown that elevated fibrinogen concentration is a risk factor for cardiovascular diseases, in particular for ischaemic heart disease and cerebrovascular accidents, but also for peripheral forms of atherosclerosis. Fibrinogen predicts the risk of these complications. The mechanism through which fibrinogen interferes with the

atherogenesis process is still under discussion and it is expected a few options:

a) increasing platelet aggregation, and thus an increased risk of a thrombus on the atherosclerotic plaque

b) increasing blood viscosity

c) infiltration through the arterial wall and cell proliferation stimulation

The blood fibrinogen level is reduced by some hypolipidemic agents, in particular fibrates and nicotinic acid. Fibrinogen is an indicator of cardiovascular diseases risk (such as inflammatory markers), but it is probably not a causal risk factor: a decrease in fibrinogen concentration in the blood did not lead to any decrease in the incidence of cardiovascular diseases in the interventional studies. Elevated fibrinogen concentrations are often associated with other risk factors: smoking, diabetes mellitus, obesity, metabolic syndrome, and it also increases during physical inactivity.

19.2.2.1. Factor VII

Factor VII is another coagulation factor. its physiological level in the blood is usually given as 70 to 120% of the standard activity. A raised factor VII concentration is associated with an increased risk of a thrombus in the artery wall that id damaged by atherosclerotic plaque. Results of epidemiological studies show that increased factor VII levels are associated with an increased risk of cardiovascular diseases. Raised factor VII levels (together with increased PAI-1) can be often found in people with metabolic syndrome.

19.2.2.2. Plasminogen Activator Inhibitor 1 (PAI-1)

Under the name of PAI-1 are included natural inhibitor of tissue plasminogen activator. An increased PAI-1 level leads to impaired fibrinolysis and that is why PAI-1 is considered a risk factor for atherosclerosis. Among other places, it is synthesized in the vessel endothelium, where the synthesis of PAI-1 may be increased either by hypertriglyceridaemia or an increase very low-density lipoproteins level. Elevated PAI-1 can be typically found (together with elevated factor VII) in people with metabolic syndrome. An increase in PAI-1 correlates with insulinaemia, and its synthesis in adipocytes (in intra-abdominal fat) may be induced by insulin. Decrease in PAI-1 levels in diabetic patients treated with metformin suggests that there is a relationship between the increase in PAI-1 and hyperinsulinemia.

19.2.2.3. Homocysteine

An elevated level of this non-essential amino acid in the blood is called hyperhomocysteinaemia (hHCy). Mild hHCy (16 - 30 µmol/L), intermediate hHCy (31 - 100 µmol/L) and severe hHCy (over 100 µmol/L) are distinguished. Severe hHCy is determined genetically and it is accompanied by thrombosis of the venous and arterial systems, thromboembolic disease and atherosclerosis complications. Fortunately, the incidence of severe hHCy in the population is very rare. In the context of atherosclerosis, is addressed primarily mild hHCy, which is relatively frequent in the population (prevalence of about 1:70). It has been proved that mild hHCy is associated with an increased risk of cardiovascular complications. A weak association between mild hHCy and CVD has been proved in healthy people; on the other hand, in patients with ischaemic heart disease, thromboembolic disease, diabetes mellitus and renal failure, mild hHCy is a strong predictor of future cardiovascular accidents.

The commonest cause of mild hHCy is polymorphism in the methylenetetrahydrofolate reductase gene, manifested particularly as a result of insufficient folic acid intake from food. In addition the homocysteine level increases also as a result of renal failure and smoking; it rises also with age and in post-menopausal women. Elevated homocysteine concentrations are frequent in patients with premature manifestation of atherosclerosis. The mechanism of the pro-atherogenic action of homocysteine is not yet clear. It was expected that homocysteine directly affects the vessel endothelium, dampens fibrinolysis, affects oxidative stress (increased production of free radicals) and may developed endothelial dysfunction with subsequent platelet activation and the formation of platelet-rich thrombi. The homocysteine level in mild hHCy can usually be easily decreased by the administration of B vitamins, in particular pyridoxine, B12 and folic acid. However, homocysteine is most probably not a causal risk factor: although there was a decrease in homocysteine concentration in practically all clinical interventional studies of folic acid and B-group vitamin administration, there was no any decrease of the incidence of cardiovascular events.

19.2.2.4. Lp-PLA2 (Lipoprotein-Associated Phospholipase A2)

Lp-PLA2 is synthesized primarily in macrophages (partially also in T-cells and hepatocytes), and it is bound to lipoproteins in the blood (mostly to LDLs, and to a small extent to HDLs). Lp-PLA2 in the blood catalyzes the transformation of phosphatidylcholine into free fatty acids and lysophosphatidylcholine. Theis process may accelerate inflammation and atherogenesis. Lysophosphatidylcholine is a chemoattractant of monocytes and T-cells, and may activate the development of vessel endothelium dysfunction, the induction of endothelial adhesive molecule expression, the proliferation of smooth muscle cells and macrophages in the vessel wall, and the expression of PDGF (platelet-derived growth factor). Oxidized free fatty acids are chemoattractants for monocytes and increase both adhesive molecule (VCAM-1) expression and cell membrane permeability. Lp-PLA2 is also present in atherosclerotic plaques, where it is produced by inflammatory cells.

It has been proved that an increased Lp-PLA2 concentration in the blood is associated with an increased risk of myocardial infarction and cerebrovascular accidents. It is due to fact that Lp-PLA2 is expressed primarily in unstable atherosclerotic plaques with a thin fibrous cover, so increased Lp-PLA2 concentration in the blood points to the risk of plaque rupture with consequent acute myocardial infarction or a cerebrovascular accident. The association between an increased blood Lp-PLA2 level and increased risk of cardiovascular diseases has been proved in many epidemiological studies. Results of these studies indicate that increased Lp-PLA2 concentration is a predictor of cardiovascular risk independent of the presence of other risk factors. Nevertheless, the Lp-PLA2 level also increases due to smoking, diabetes mellitus, obesity and metabolic syndrome. Concentrations over 200 - 235 ng/mL are considered elevated.

19.2.2.5. Microalbuminuria

Albumin excretion to the urine in the ranges 30 to 300 mg/24 hrs in is called “microalbuminuria”. This assay is performed especially in patients with diabetes mellitus, as it is an early indicator of diabetic nephropathy. Clinical studies have shown, however, that microalbuminuria is also associated with the risk of cardiovascular complications, not only in patients with diabetes mellitus but also in non diabetic population. Microalbuminuria is therefore a predictor of cardiovascular morbidity and mortality in diabetic and hypertonic patients as well as in the general population. It is supposed that microalbuminuria is not only the indicator of glomerular damage, but also generally a marker of the damaged vessel permeability that is present in endothelial dysfunction. Vessel endothelium dysfunction is one of the key steps in the development of atherosclerotic focus.

19.2.2.6. Disorder of Glucose Homeostasis

Practically all disorders of glucose homeostasis (insulinoresistance with hyperinsulinaemia, impaired fasting glycaemia, impaired glucose tolerance, type 2 DM) are risk factors for the premature atherosclerosis and the development of cardiovascular complications. Cardiovascular diseases are also the main cause of mortality of patients with diabetes mellitus. The incidence of ischaemic heart disease is 2-4 times higher in diabetics; the risk of stroke is 3-5 times higher; and the risk of lower-extremity ischaemia is even up to 30 times higher compared with non-diabetics. Mechanisms accelerating atherogenesis in diabetis subjects have not yet been fully clarified. Diabetes mellitus is usually accompanied by dyslipidaemia hypertension and obesity (obesity is the main cause of type 2 diabetes mellitus). Besides the risk factors mentioned above, many other pathophysiological processes take part in atherosclerosis acceleration in diabetic subjects: endothelial dysfunction, platelet hyperreactivity, impaired fibrinolytic balance, impaired blood flow, and chronic inflammation. Platelets in patients with diabetes mellitus increasingly adhere to the vessel endothelium and their aggregability is increased; enhanced von Willebrand factor production in the damaged endothelium also contributes to their activation. The concentration of other coagulation factors (fibrinogen, factor VII, VIII, XI, XII, kallikrein) is also increased with concurrent inhibition of the fibrinolysis system (increase in PAI-1). The permeability of cell membranes is impaired. Type 2 diabetes mellitus is commonly accompanied by dyslipidaemia, characterized by the presence of small LDL_3 particles, a decrease in HDL-cholesterol and increased triglycerides. The atherosclerotic process becomes accelerated in the period before diabetes manifests itself, when hyperinsulinaemia and insulinoresistance are present, and they initiate the development of vessel endothelium dysfunction as the first step in the atherosclerotic process.

CHAPTER 20

20. Free Radicals, Relation to Diseases and Protection against Them

Authors: Prof. MUDr. Tomáš Zima, DrSc.,MBA, prof. MUDr. Marta Kalousová, PhD.

Reviewer: MUDr. Daniel Rajdl, PhD.

20.1. What Is Oxidative Stress?

Oxidative stress is a mechanism of damage to the body which co-participates in the pathogenesis of many diseases and their complications, such as atherosclerosis, diabetes mellitus, renal disease, ischaemia-reperfusion injury, neurodegenerative diseases, carcinogenesis and inflammatory diseases.

Oxidative stress is defined as an imbalance between the increased production of oxidants and insufficient antioxidant defence mechanisms, which results in damage to tissues.

20.2. Oxidants – Free Radicals and Reactive Forms of Oxygen and Nitrogen

Oxidants are free radicals and other reactive forms of oxygen and nitrogen which are closely related to radical reactions. Free radicals are atoms, molecules or ions containing one or more unpaired electrons in the bonding orbital. They are unstable, highly reactive and tend to make chain reactions. Examples of important oxidants include the hydroxyl radical $OH^{\cdot}$, superoxide $O_2^{\cdot -}$, singlet oxygen $1O_2$, hydrogen peroxide H_2O_2, nitric oxide $NO^{\cdot}$, hypochlorous acid HClO, and the peroxyl and alkoxyl radicals $ROO^{\cdot}$ and $RO^{\cdot}$. Figure 1 shows the reactions leading to oxidant generation.

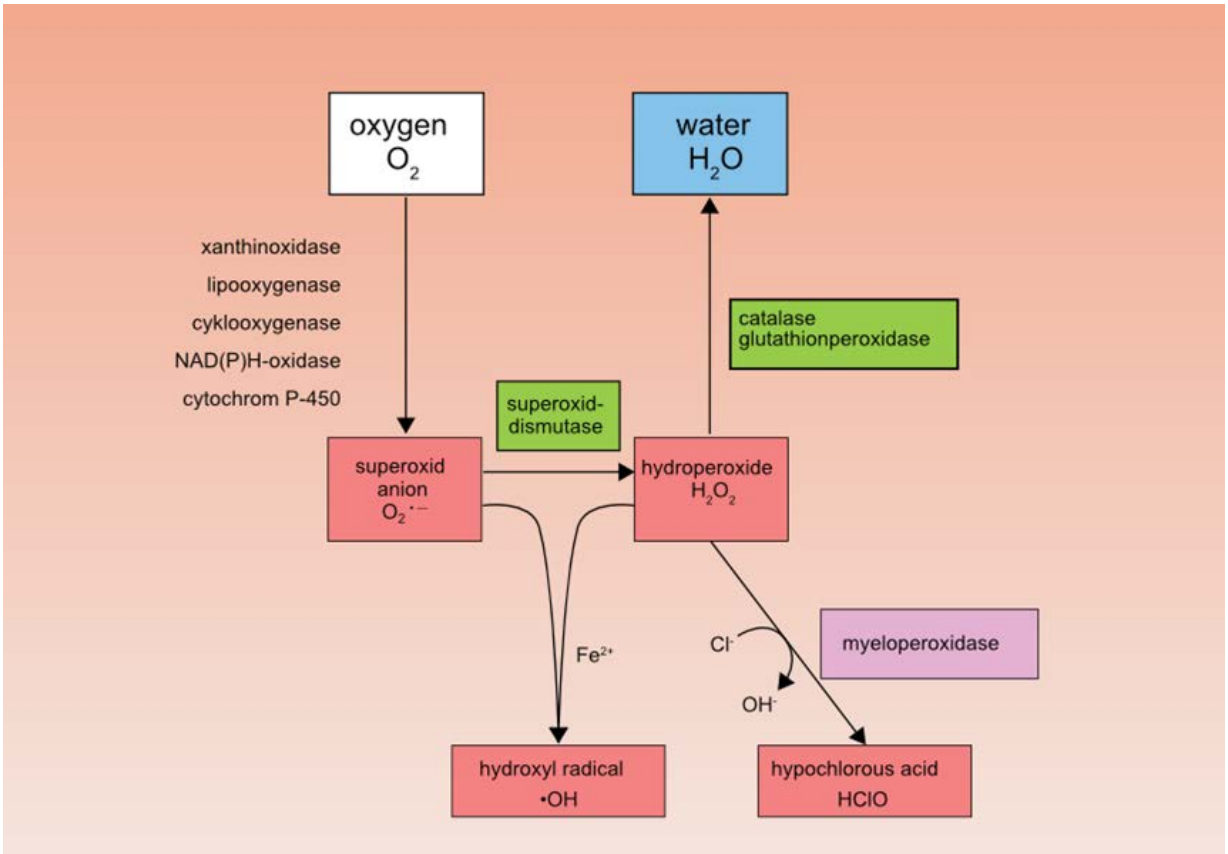


Figure 20.1. Oxidant generation and oxidant inter-reactions

The most important source of free radicals in the body is the mitochondrial respiratory chain. Oxygen commonly accepts 4 electrons and is transformed into water. This reaction, however, takes place in successive steps: oxygen → superoxide → hydrogen peroxide → hydroxyl radical → water. Another important producer of oxidants is phagocytes (neutrophils and monocytes-macrophages), specifically their NADPH oxidase and myeloperoxidase. In the respiratory burst (Figure 21.2), following activation by inflammatory stimuli, NADPH oxidase reduces molecular oxygen to superoxide, which is subsequently converted to hydrogen peroxide. These substances can be another source of oxidants – nitric oxide, peroxynitrite and hydroxyl radical. In the presence of chloride ions, hydrogen peroxide is also metabolized by myeloperoxidase to hypochlorous acid, which may subsequently react with endogenous amines to form chloramines. Free radicals are also generated during xenobiotic detoxification and in many other chemical reactions, such as reactions catalyzed by cytochrome P450, xanthine oxidase, lipoperoxidase or cyclooxygenase. Their generation is accelerated by transition metals - iron, copper – Fenton's reaction:

$$H_2O_2 + Fe^{2+} \rightarrow HO^{\cdot} + OH^{-} + Fe^{3+}$$

Nitric oxide ($NO^{\cdot}$) is generated from L-arginine by the action of NO synthase (NOS). Constitutive NOS is found primarily in the endothelium and neurons, and inducible NOS primarily in macrophages.

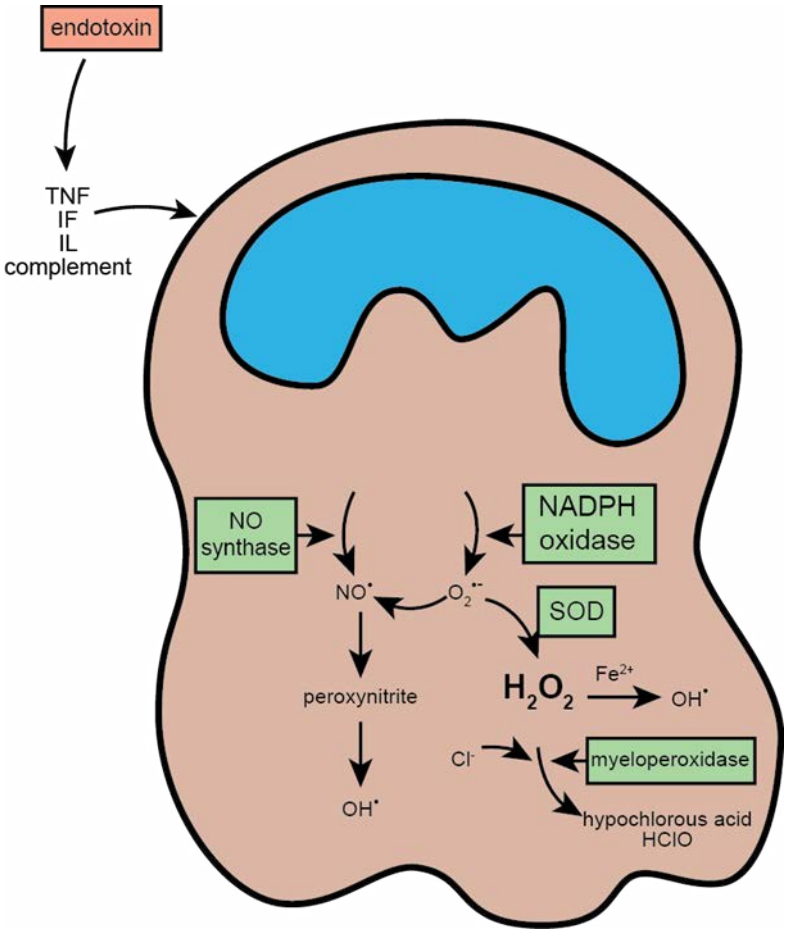


Figure 20.2. Respiratory burst of neutrophils

20.3. Antioxidants – Substances Acting against Oxidants

Antioxidants are substances working against the generation and effect of free radicals – as scavengers (interceptors) of free radicals, as free radical generation inhibitors through the bond to transition metals, for example, or else they eliminate hydroperoxides and repair damage. Antioxidants in the human body can be divided into two groups:

Enzymes and enzyme cofactors = trace elements (superoxide dismutase – copper, zinc, manganese; glutathione peroxidase – selenium; catalase – iron; cytochrome P450, lactoperoxidase;

Substrates – lipophilic (tocopherols and carotenoids) and hydrophilic (ascorbic acid, glutathione, thiols, caerulo-

plasmin, transferrin, ferritin, albumin, bilirubin, uric acid and others).

Antioxidant properties can also be found in many synthetic compounds such as the iron and copper chelates deferroxamine and penicillamine, the xanthine oxidase inhibitors allopurinol and oxypurinol, probucol, lazaroids, angiotensin-converting-enzyme inhibitors and statins.

20.4. Compounds Generated Due to Oxidative Stress – Radical Reaction Products and Their Importance in Tissue Damage

Oxidative stress **modifies biologically important compounds**. It changes their structure by fragmentation, aggregation, cross-linking; and, it changes their properties – hydrophobicity, proneness to proteolysis and function, as well as their immunological properties. These changes depend on the chemical nature of the agent and the intensity of attack. The damage affects lipids, proteins, carbohydrates and nucleic acids. Fatty acids with multiple double bonds – polyunsaturated fatty acids – are most prone to radical reactions in lipids, and the process is referred to as **lipid peroxidation**. Their structure is gradually rearranged; conjugated dienes, peroxyl and alkoxyl radicals and hydroperoxides are generated, and the modified molecule may also cleave into shorter products to form malondialdehyde or 4-hydroxynonenal. Membrane lipids as well as lipoproteins may become damaged. The amino acids tyrosine, methionine, cysteine and tryptophane are subject to damage in **proteins** by oxidation, nitration, chlorination and dimer formation. Proteins may then aggregate and cross-link to generate advanced oxidation protein products (AOPP). **Advanced glycation end products** (AGEs) are generated with the help of carbohydrates and also due to carbonyl stress. Proteins can also be modified through the action of lipid peroxidation products (malondialdehyde lysine, hydroxynonenal and acrolein-protein adduct) to form advanced lipid peroxidation end products (ALEs). **DNA damage** affects deoxyribose and bases, which results in DNA chain breaks and chain cross-linking. *Figure 3* shows damage to biological structures and its effects.

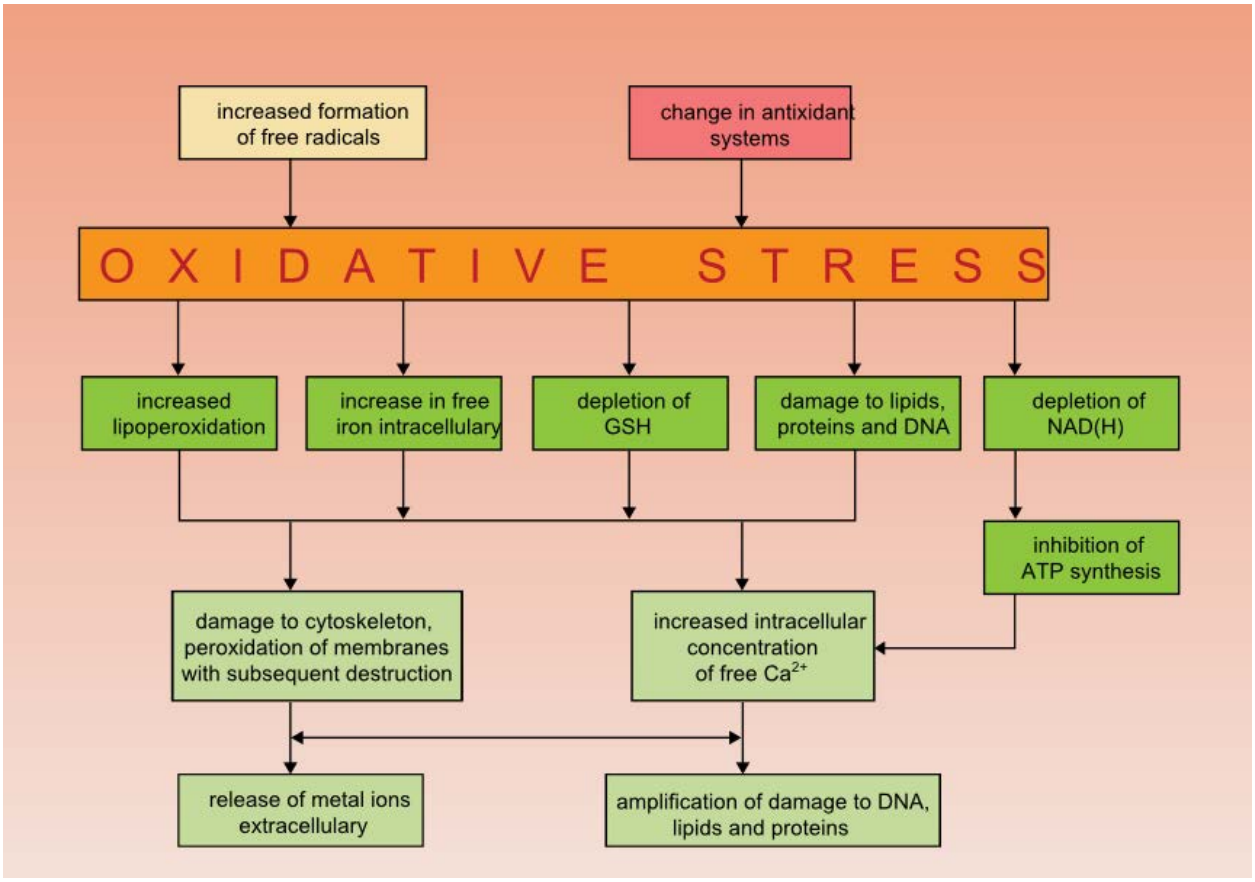


Figure 20.3. Effects of oxidative stress

Oxidative stress is closely related to **carbonyl stress**, which is characterized as an increase in reactive carbonyl compounds caused by their reduced production and/or reduced elimination and excretion. These are compounds containing a carbonyl group such as glyoxal and methylglyoxal, or products of lipid peroxidation, malondialdehyde and 4-hydroxynonenal. Reactive carbonyl compounds may also be generated from carbohydrates, lipids and amino acids in oxidative and non-oxidative reactions, and may give rise to ALEs and AGEs. AGEs, a heterogeneous group of compounds such as carboxymethyllysine, pentosidine, glyoxal-lysine dimer and methylglyoxal-lysine dimer, have been

characterized the best; they are generated spontaneously by non-enzymatic glycation, but also due to oxidative and carbonyl stress. Apart from structural damage and change in the biological properties of important substances, the end products of advanced damage by oxidative and carbonyl stress (AGEs, ALEs, AOPP) have many other effects mediated through specific receptors, of which **RAGE** (receptor for AGEs, with AGE-R1, R2 and R3 being essential) was the first to be described.

RAGE is a multiligand receptor, and AGEs were described as the first ligands. β-amyloid, S100-proteins/kalgranulins or HMGB1 (high mobility box group protein 1) bind to RAGE. **Ligand-RAGE interactions** trigger a signalling cascade involving p21ras, MAP-kinase, ERKs and JNKs; in addition, the nuclear factor κB becomes activated, which is also connected with oxidative stress. This is followed by the stimulation of gene transcription of cytokine and growth factors (TNF, IL-1, PDGF, IGF-1, interferon γ) and of adhesion molecules (ICAM-1, VCAM-1), the stimulation of cell proliferation, an increase in vascular permeability, the induction of macrophage migration, etc. The AGE-RAGE interactions, subsequent oxidative stress and AGEs also directly damage DNA. In addition, the RAGE activation affects the expression and activity of glyoxalase I, which is an enzyme detoxifying the AGE precursors methylglyoxal and glyoxal. RAGE plays a role in the pathogenesis of many diseases and their complications, such as diabetes mellitus, atherosclerosis, Alzheimer's disease, rheumatoid arthritis, inflammatory diseases and tumour metastasizing.

20.5. Physiological and Pathological Role of the Reactive Forms of Oxygen and Nitrogen, Importance in Pathogenesis

When discussing oxidative stress, its definition must include not only the imbalance between oxidants and antioxidants in favour of oxidants, but also take into account the subsequent damage to the body, since oxidants, i.e. free radicals and other reactive forms of oxygen and nitrogen, **have many important biological functions** in the inflammatory response, xenobiotic metabolism, vascular tonus, or as second messengers in signal transduction (*Table 1*).

Physiological role	Specific function
Inflammatory reaction	Respiratory burst of phagocytes
Signal transduction	Intracellular signalling – second messenger role
Monooxygenase reactions	Steroid hydroxylation
	Xenobiotic metabolism
Vascular tonus regulation	NO – relaxation
	superoxide – constriction
Mediators	NO – CNS neuromodulator

Table 20.1. Physiological roles of free radicals

In contrast to this, the **pathological effect of radicals and radical reaction products** involves extensive damage to biological structures, damage signalling through receptors with the subsequent triggering of many cascade reactions depending on the cell type – pro-inflammatory responses, cell proliferation or apoptosis, and the effect on other structures by the chemical compounds thus produced. Therefore, oxidative stress plays a role and co-acts in the pathogenesis of many diseases and their complications, from inflammatory diseases (sepsis, chronic inflammatory diseases), through cardiovascular diseases (hypertension, atherosclerosis, ischaemia-reperfusion), diabetes mellitus and its complications, renal impairment and its effects, neurodegenerative diseases (Alzheimer's disease, multiple sclerosis) through to tumour diseases and their treatment, or pregnancy, and especially to pathological conditions in pregnancy (pre-eclampsia).

Pathological role	Specific function

Damage to biological structures	Lipid peroxidation
	Glycoxidation
	Protein modification
	DNA modification
Role in the pathogenesis of many diseases and their complications	Atherosclerosis
	Diabetes mellitus
	Renal diseases
	Tumour diseases
	Inflammatory diseases
	Neurodegenerative diseases

Table 20.2. Pathological roles of free radicals

20.6. Laboratory Diagnostics

The determination of free radicals in body fluids is problematic due to their short biological half-life. It is therefore more appropriate to measure levels of compounds generated due to oxidative stress or antioxidation protection. Assays can be made on various biological materials such as serum, plasma, urine, whole blood, blood elements or exhaled air. Laboratory diagnosis must always be comprehensive and the results of regular tests may often point to the probable presence of oxidative stress. Laboratory results must be evaluated on an individual basis in relation to the patient's clinical condition. Laboratory assays for oxidative stress and antioxidant parameters are used mainly to study the cause and course of a disease; they are not used in normal clinical practice.

Direct measurements include assays for generating radicals and the reactive forms of oxygen and nitrogen using methods such as pulse radiolysis, electron spin resonance and chemiluminescence. These methods are used especially in basic research and are not commonly available.

Indirect measurements determine substances generated in radical reactions, antioxidants and autoantibodies. It is also possible to determine degradation products in NO generation (nitrites, nitrates) or use an immunochemical assay (ELISA) for enzymes generating radicals – myeloperoxidases or xanthine oxidase.

It is possible to determine many **substances generated in radical reactions (classic oxidative stress markers)**; for example, lipid peroxidation is characterized by malondialdehyde, oxidized LDLs or 8-isoprostane as an arachidonic acid metabolite. Protein impairment can be mapped by measuring modified amino acids or as AOPP; AGEs or sRAGE (soluble receptor for AGEs) are typical for the effects of carbohydrates. 8-hydroxy-2´-deoxyguanosine is a marker of DNA damage. ELISA, high-performance liquid chromatography (HPLC) and gas chromatography with mass spectrometry (GC/MS) are the most commonly applied methods.

Assays for selected **antioxidants** are more available: vitamins A and E (HPLC), trace elements, zinc and selenium in particular (atomic absorption spectrometry); it is also possible to determine the activity of antioxidant enzymes or total antioxidant capacity using spectrophotometry.

Another possibility, again more used for research, is the **assay for autoantibodies** generated in reaction to the modification of biological structures. The oxidative stress process involves the generation of compounds that are immunogenic and the body produces antibodies against them. Another system (immune system) thereby becomes involved and may be impaired by oxidative stress; an example would be antibodies against oxidized LDL determined by ELISA.

Molecular biology techniques are able to determine a genetic predisposition to tissue damage (e.g. polymorphism of genes for antioxidant enzymes or enzymes degrading reactive carbonyl compounds, or genes for receptors mediating tissue damage such as RAGE – receptor for AGEs).

Basic laboratory assays can point to various pathologies and signal the probable presence of oxidative stress. The

function of kidneys is examined routinely, and it is known that oxidative stress is present in patients with impaired renal function. Oxidative stress also has an important connection with inflammatory reaction (indicators such as C-reactive protein, orosomucoid) and the atherogenesis process (modified LDL particles, myeloperoxidase). An elevated glucose or lipid level means more substrates for the production of reactive carbonyl compounds. On the other hand, albumin and bilirubin (higher levels without hepatic involvement, e.g. Gilbert's syndrome) and also glucose have antioxidant effects. Erythrocytes contain another important antioxidant, glutathione, and so a drop in total glutathione is expected in anaemia.

Table 21.3 summarizes selected assays for detecting oxidative stress. The methods listed below describe the reactions and processes that take place in the human body during oxidative stress. Individual analytes should be considered in context of their mutual relations before evaluating the degree of damage to the body by oxidative stress.

Group of Assays	Parameters
Direct measurement – determination of generated radicals	Pulse radiolysis, electron spin resonance, chemiluminescence
Oxidative stress damage markers	Lipid peroxidation – malondialdehyde, conjugated dienes, oxLDL, ALEs, 8-isoprostane
	Amino acid and protein oxidation – modified amino acids (3-chlorotyrosine, 3-nitrotyrosine, dichlorotyrosine), AOPP
	Glycoxidation – AGEs, sRAGE
	Oxidative DNA damage – 8-hydroxy 2´deoxyguanosine
	Degradation products in NO generation (nitrites, nitrates)
	Radical-generating enzymes – xanthine oxidase, myeloperoxidase
Antioxidants	Vitamins A and E
	Trace elements – Se, Zn
	Albumin, bilirubin
	Glucose, uric acid
	Antioxidant enzymes – superoxide dismutase, glutathione peroxidase, catalase
	Total antioxidant capacity
Autoantibodies	Antibodies against oxidized LDL
Molecular biology techniques	Polymorphism of genes for antioxidant enzymes and enzymes degrading reactive carbonyl compounds, or genes for receptors mediating tissue damage such as RAGE – receptor for AGEs
Basic laboratory assays	Renal function – estimated glomerular filtration, serum creatinine
	C-reactive protein, albumin, glucose level, HbA1c, microalbuminuria, lipids, bilirubin, blood count

Table 20.3. Selected assays for detecting oxidative stress

20.7. Possible Therapies

There are no standard recommendations for oxidative stress therapy at present. However, many studies have

shown the positive effect of administering some antioxidants, and it is also known that some drugs which are frequently administered for different indications have antioxidant properties (angiotensin-converting-enzyme inhibitors, angiotensin II AT_1 receptor blockers or statins). The physiological balance between oxidants and antioxidants and potential adverse effects should always be taken into account when considering therapy. For example, some antioxidants may also have pro-oxidizing effects. The administration of antioxidants is not currently part of the recommendations for therapy or prevention of diseases, except for some rare diseases. On the contrary, the unsuitable administration of antioxidants (e.g. beta-carotene in smokers) may even lead to damage to the body. Antioxidants are administered as a supportive or complementary treatment, not as a causal therapy.

CHAPTER 21

21. Biochemical Tests for Liver Diseases

Authors: Prof. MUDr. Libor Vítek, PhD., Doc. MUDr. Milan Jirsa CSc.

Reviewer: MUDr. Milan Dastych

A standard set of biochemistry assays sometimes referred to as "liver tests" or "liver function tests" has been put together over the years. These assays are not reflecting liver function tests in the real sense of the word; rather, they are tests for markers substituting specific liver function tests. Biochemical tests can be divided into several groups based on different points of view (see the table below):

Tests indicative of hepatocyte integrity	ALT, AST
Tests indicative of disorders at the level of bile duct system and the canalicular pole of hepatocytes	ALP, GT (GGT)
Tests measuring protein synthesis by the liver	Albumin, prealbumin, cholinesterase, prothrombin complex/coagulation factors
Tests measuring liver capacity to transport organic anions and eliminate endogenous and exogenous substances from circulation	Bilirubin, bile acids
Tests measuring the liver's ability and capacity to metabolize xenobiotics or endogenous substances	Ammonia, CDT, lidocaine, aminopyrin
Laboratory tests for the diagnosis of specific hepatic diseases	Viral hepatitis, haemochromatosis, Wilson's disease, porphyric diseases, autoimmune hepatic diseases, etc.

Table 21.1. Summary of biochemical tests of liver function

21.1. Tests indicative of impairment of hepatocyte integrity

- Aspartate aminotransferase (AST) and alanine aminotransferase (ALT)

Serum AST and ALT activities are elevated in most hepatic diseases. The highest values can be seen in fulminant viral hepatitis, drug-induced liver injury and toxicity, and circulatory shock. The magnitude of the increase reflects the extent of liver injury, but does not correlate with the prognosis. A decrease in values may reflect both declining disease and massive necrosis.

The AST/ALT ratio (de Ritis index) is particularly useful in alcoholic liver injury, where it is often higher than 2. The activities of both enzymes is affected by many factors; there is a correlation between body weight and AST/ALT activity – the both enzymes have up to 40% higher activities in obese people, most often as a result of the development of non-alcoholic steatohepatitis (NASH). This is also related to the fact that the asymptotic elevation of aminotransferase activity correlates with total mortality (similar association has been reported for elevation of gamma-glutamyl transferase (GT) activity, see below). At the same time, the asymptomatic elevation of aminotransferase activities is highly prevalent; 10-20% of the total population is afflicted due to alcoholic injury and/or non-alcoholic fatty liver.

21.2. Tests indicative of disorders at the level of bile duct system and the canalicular pole of hepatocytes

- Alkaline phosphatase (ALP)

Among liver diseases, the highest values can be seen in cholestatic diseases. This is caused by the localization of ALP (and also GT) at the canalicular pole of hepatocytes and in the apical membrane of the bile duct epithelium. ALP activity increases in extrahepatic and intrahepatic cholestasis; some patients with malignant tumours have an increased ALP activity as a result of ectopic ALP production in the malignant tissue (Regan isoenzyme). Other extrahepatic causes of higher ALP activity include bone diseases, some rheumatological diseases, vasculitis, hyperparathyroidism, hyperthyroidism, or acromegaly.

- Gamma-glutamyl transferase

GT is a non-specific marker of cholestasis, but it also increases following the administration of some drugs and alcohol abuse as a result of microsomal isoform induction. The half-life of the GT elevation after alcohol consumption has been discontinued is long (26 days). In this case, a combination of assays for GT activity and the carbohydrate-deficient transferrin (CDT) level are very useful in confirming alcohol abuse. CDT is a transferrin lacking the usual carbohydrate component in the side chain. Transferrin glycosylation is post-translational and is adversely affected by alcohol abuse.

The GT value correlates with ALP in liver diseases. It is considered a more sensitive test for hepatobiliary disease than ALP. The diagnostic value of GT is limited by low organ specificity – GT activity also rises in many extrahepatic diseases such as cardiac infarction, renal insufficiency, obstructive bronchopulmonary disease, diabetes and pancreatic diseases. In addition, marked overweight may increase GT by up to 50%, so it is no surprise that higher GT activity is associated with higher cardiovascular morbidity and mortality.

21.3. Tests measuring protein synthesis by the liver

The liver synthesizes practically all plasma proteins except immunoglobulins. Serum level measurement allows for a rough estimate of protein synthetic capacity of the liver.

- Albumin

Albumin is quantitatively the most important serum protein synthesized in the liver. The half-life of albumin is 19 - 21 days, and so albumin is not a suitable parameter for evaluating protein synthesis in acute hepatic diseases. Serum albumin level is not only influenced by albumin synthesis, but also by the nutritional status, distribution space and catabolism of albumin.

Serum albumin is typically reduced in patients with alcoholic liver cirrhosis and ascites. Other possible causes of reduction include nephrotic syndrome, protein losing enteropathy, burns, catabolic conditions and corticosteroid treatment. The serum level correlates very well with the prognosis, which is why it is one of the criteria for the Child - Pugh score. The serum level in patients with ascites may be decreased considerably due to albumin leakage into the peritoneal cavity. A rise in albumin level is typical of haemoconcentration in dehydration and diuretic therapy.

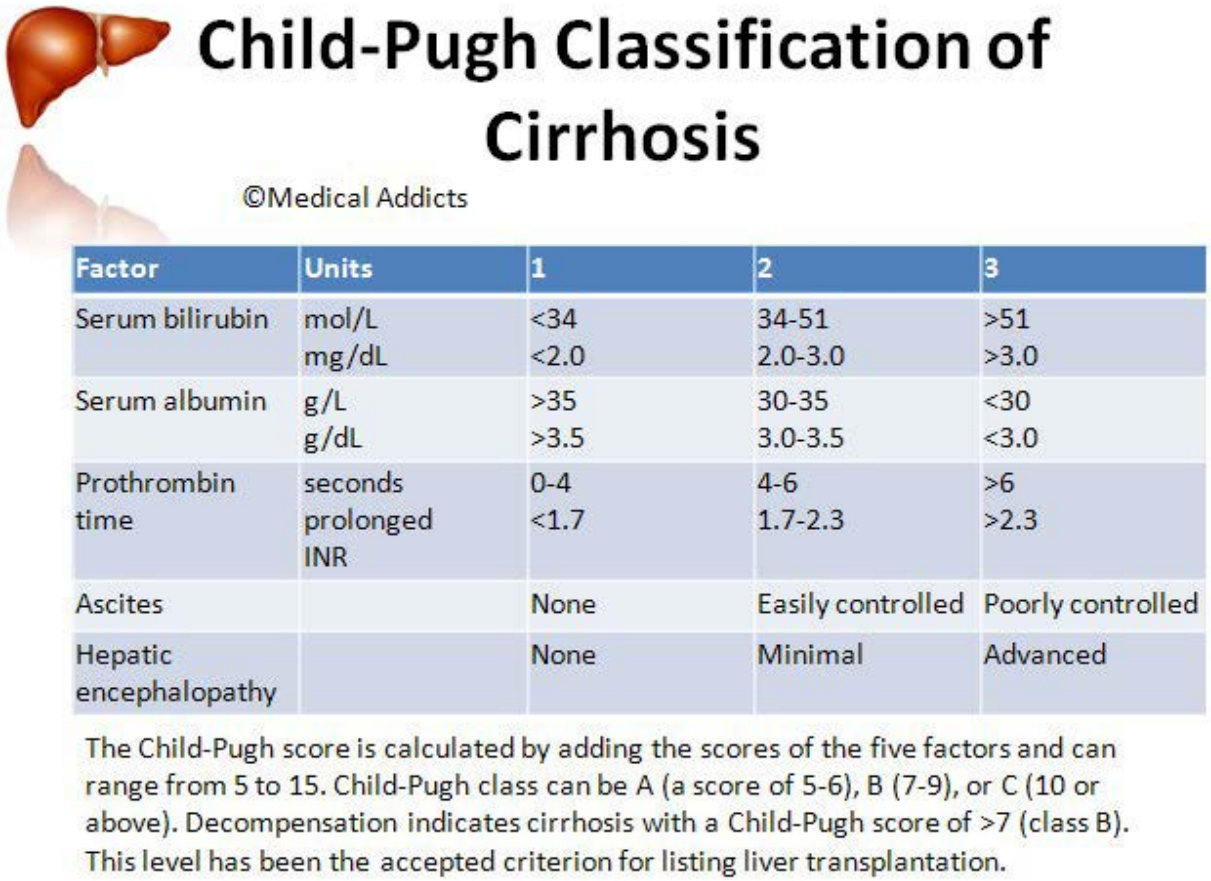


Figure 21.1. The Child-Pugh score (sometimes the Child-Turcotte-Pugh score) is used to assess the prognosis of chronic liver disease, mainly cirrhosis. Although it was originally used to predict mortality during surgery, it is now used to determine the prognosis, as well as the required strength of treatment and the necessity of liver transplantation.

- Prealbumin

The half-life of prealbumin is only 1.9 days, so the prealbumin level is used to evaluate proteosynthesis in the liver parenchyma in acute liver lesions.

- Changes in γ-globulins

Polyclonal gammapathy is a standard finding in patients with hepatic cirrhosis. An increase in each -globulin group is typical of some hepatic diseases; elevated IgM is typical of primary biliary cirrhosis, IgG of autoimmune and cryptogenic liver cirrhosis, and IgA of alcoholic cirrhosis.

- Coagulation factors

Many coagulation factors such as factor I (fibrinogen), II (prothrombin), V, VII, IX and X are synthesized in the liver. The synthesis of coagulation factors II, VII, IX and X depends on vitamin K. Most of these factors are abundant in the serum, and decrease only if the proteosynthetic function of the liver is considerably impaired. Coagulation disorders commonly occur in hepatic diseases, most often due to disorders of the proteosynthetic functions of the liver parenchyma or cholestatic liver lesions, when the absorption of fat-soluble vitamins, including vitamin K, is impaired.

The prothrombin time test (Quick test) is used most often to evaluate the synthesis of coagulation factors produced by the liver. The test reflects external coagulation system activity by measuring prothrombin-to-thrombin conversion in the presence of Ca^{2+} ions and tissue thromboplastin. Given the short half-life of prothrombin, the test reflects changes very fast and is highly responsive. It is therefore an important test for acute liver conditions or the progression from chronic hepatic disease to hepatic failure. The measurement result is commonly referred to as the INR (International Normalized Ratio), the ratio of tested to control prothrombin time. Values within the range of 0.75 - 1.25 are considered normal; an INR over 1.5 is accompanied by an increased risk of bleeding.

The prothrombin time is one of components for rating liver function capacity according to the Child-Pugh score.

21.4. Analytes measuring the transport and excretory capacity of the liver

a), b) Bilirubin and urobilinogen – [refer to Chapter 24]

- Bile acids

The determination of serum bile acids is a sensitive test of cholestatic liver lesions. Patients with cholestatic liver lesions are often treated with ursodeoxycholic acid; this treatment has to be discontinued at least one week before scheduled blood collection for the bile acid assay. The highest levels can be seen in primary biliary cirrhosis, primary sclerosing cholangitis and congenital defects in the bile acid biosynthesis pathway or bile pigment transport by the hepatocyte. Very important is the assay for serum bile acids in women with suspected intrahepatic cholestasis of pregnancy, where a relatively slight increase in the bile acid level (over 40 µmol/L, with normal levels under 7 µmol/L) poses a serious risk of cardiodepression for the foetus.

21.5. Tests measuring the liver's ability and capacity to metabolize endogenous and xenogenous substances

- Ammonia

An increase in the ammonia level can be seen in hepatic diseases, particularly at the phase of hepatic insufficiency. High levels are also present in urea cycle enzyme deficiency, Reye's syndrome, acute and chronic hepatic encephalopathy, and in bleeding into the digestive tract accompanied by increased production of ammonia by the gut flora.

Ammonia determination has many pre-analytic limitations. The correlation is better if arterial blood is taken instead of venous blood. The plasma has to be separated from red blood cells within 1 hour, ideally within 15 minutes in patients with cirrhosis.

- Carbohydrate-deficient transferrin – see above
- Dynamic tests

These tests measure the liver's capacity to metabolize foreign substances. Non-toxic substances metabolized only by the liver are used. Most of these substances are metabolized by microsomal enzymes, and the tests therefore provide a good illustration of the microsomal functions of the liver. The best-known are the antipyrine and aminopyrine breath tests, either with direct measurement of the metabolized substance, or the method using a stable isotope ^{13}C-labelled substrate to measure exhaled labelled CO_2. Despite their relatively high diagnostic value, these tests are not widely used in practice, in particular due to the difficulty of determining stable isotopes in the exhaled air. Chromo-excretion tests are also available, but they pose a potential risk of serious anaphylactic reactions.

21.6. Laboratory assays for the diagnosis of specific liver diseases

- Laboratory assays for suspected haemochromatosis
- Transferrin saturation: physiological range 19 - 49%, a more than 60% increase is indicative of haemochromatosis in men and 50% in women.
- Plasma/serum iron: physiological range 6 - 29 µmol/L, values > 30 µmol/L are suggestive of haemochromatosis.
- Serum ferritin: physiological range 30 - 400 µg/L, values > 1.000 mg/L are usually suggestive of haemochromatosis.
- Iron content in liver dry matter: physiological range 0 - 2.000 µg/g dry matter, > 2000 µg/g dry matter is indicative of haemochromatosis.
- Gene testing: Mutations of the haemochromatosis gene on the short arm of chromosome 6; homozygous C282Y gene mutations or heterozygous C282Y mutations in combination with H63D in 80% of afflicted patients.
- Laboratory assays for suspected Wilson's disease
- Serum caeruloplasmin: physiological range 0.2 - 0.6 g/L, levels under 0.2 g/L are suggestive of Wilson's disease – this reduction is present in 90% of afflicted patients.
- Serum copper level: physiological range 12 - 24 µmol/l, levels > 24 µmol/L are suggestive of Wilson's disease, but levels may also be normal.
- Free copper in serum: physiological range 0 - 1.6 µmol/L, an increase in the fraction over 2.0 µmol/L is suggestive of Wilson's disease.

- Urinary copper excretion: physiological range 0.03 - 1.26 µmol/L, an increase over 1.5 µmol/L is indicative of Wilson's disease; the penicillamine challenge test is also available.
- Copper content in liver dry matter: physiological range under 50 µg/g dry liver tissue; an increase over 250 µg/g liver tissue is found in patients with Wilson's disease, and high levels can be found also in cholestasis of any aetiologies.
- Gene testing: the commonest H1069Q gene mutation is present in 40 - 50% of afflicted patients.
- Laboratory assays for suspected porphyria

Porphyrias belong to a group of haem biosynthesis disorders characterized by considerable variability in the clinical presentation. Clinically, porphyrias can be divided into acute and chronic types. The commonest chronic porphyria in the Czech Republic is porphyria cutanea tarda (PCT), and the commonest acute porphyria is acute intermittent porphyria (AIP).

- Porphobilinogen test on fresh morning urine

Importance: acute porphyric attack marker, more sensitive than δ-aminolevulic acid (ALA)

Reference range: < 0.25 mg/dL

- D-aminolevulic acid test on fresh morning urine

Importance: like porphobilinogen (PBG), ALA is also a marker for acute porphyric attack. The urinary ALA assay is important for lead poisoning diagnosis (lead reversibly inhibits the ALA-dehydratase enzyme) and porphyrias due to congenital ALA-dehydratase deficiency (incorrectly plumboporphyria).

Reference range: < 0.45 mg/dL

Both analytes are tested in 10 mL of fresh urine. PBG and ALA are both sensitive to light, heat, storage time and pH fluctuations. The PBG test has fewer interferences than ALA.

- Total porphyrin test on collected urine and stool

Importance: total porphyrin output in the urine is a marker for acute and chronic porphyria, as it is for secondary coproporphyrinuria (typically ranging from 200 to 500 µg/24 hours).

Reference range: < 200 µg/24 hours

The test for total porphyrin in faeces (sample of about 5 g stool) is similar.

Reference range: < 200 ng/g dry mass

There is also a set of special methods available in the CR at the Hepatology Laboratory, Institute of Medical Biochemistry and Laboratory Diagnostics of the General University Hospital and of the First Faculty of Medicine, Charles University in Prague (affiliation of the author – LV). These methods include the tests for plasma porphyrin emission maximum, urine and faeces porphyrin fractionation, a test for the activities of some haem biosynthetic pathway enzymes, or molecular diagnostics (provided in the CR by Prof. Martásek's site at the Paediatric Clinic of the First Faculty of Medicine, Charles University in Prague).

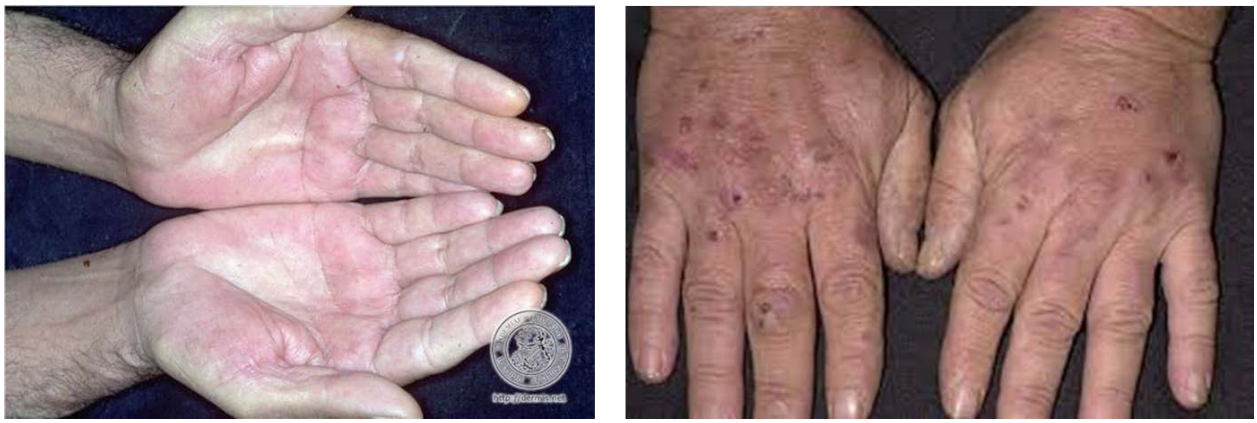


Figure 21.2. Porphyria cutanea tarda (commonly referred to as PCT). Typically, patients who are ultimately diagnosed with PCT first seek treatment following the development photosensitivities in the form of blisters and erosions on commonly exposed areas of the skin. This is usually observed in the face, hands, forearms, and lower legs. It heals slowly and with

scarring. Other skin manifestations like hyperpigmentation (as if they are getting a tan) and hypertrichosis (mainly on top of the cheeks) also occur.

- Laboratory assays for suspected hepatocellular carcinoma
- α_1-fetoprotein (AFP)

This is a glycoprotein synthesized in the yolk sac, foetal liver and malignant epithelial hepatic carcinomas. An increase is associated with hepatocellular carcinoma, but also embryonal carcinomas, germinal tumours and cleft neural plate.

- Serological laboratory tests for liver diseases

Serological hepatological assays play a major role in the diagnosis of viral hepatitis. Nowadays, they represent one of the main causes of acute and chronic hepatic diseases. The diagnosis currently relies on a combination of serological and molecular genetics methods.

Hepatitis A

The diagnosis relies on the presence of antibodies against the hepatitis A virus (anti-HAV)

- In the IgM class = acute hepatitis A
- In the IgG class (anamnestic antibodies) = history of hepatitis A, HAV vaccination

Hepatitis B

Acute hepatitis B

- Acute hepatitis B may be diagnosed based on serological findings only if basic patient history is known, because serological findings can be identical with findings for an acute hepatitis B virus infection at certain stages of chronic infection.
- Positive hepatitis B surface antigen (HBsAg) and antibodies against hepatitis B core antigen (anti-HBc) of the IgM class
- Secretory antigen (HBeAg) – persists for more than 10 weeks, probable transition to chronicity
- Occurrence of anti-HBs antibodies means recovery from acute infection and, with titre over 100 IU/mL, also permanent immunity. HBsAg may disappear in acute infection several weeks before anti-HBs antibodies appear (the period called diagnostic gap or window) – the only serological sign of HBV infection is the presence of anti-HBc antibodies.

Chronic hepatitis B

- Positive HBsAg, specific and sensitive test
- Anti-HBs antibodies persist permanently in most people with a history of hepatitis B, and are also a sign of successful HBV vaccination.
- HBe antigen – can be proved only if HBsAg is present at the same time. The presence of HBeAg reflects active replication. It is negative in pre-core HBV mutations. In such case, replication should be confirmed by the HBV DNA test.
- Transition to chronicity is characterized by the presence of HBsAg > 6 months, HBeAg (±) and HBV DNA (±) in the serum.
- The occurrence of anti-HBe antibodies and the disappearance of HBeAg is referred to as seroconversion in the HBe system and means the end of viral replication.
- HBV DNA test using PCR

Hepatitis C

Demonstration of infection:

- Serological tests demonstrating anti-HCV antibodies
- Molecular genetics tests – demonstrate HCV RNA in the blood, enable the quantification of viraemia, and determine the genotype of virus

Serological tests to demonstrate rarer **hepatitis D** and **E** virus infection are also available.

- Immunological tests in diagnosing liverdiseases

Assays for autoantibodies are performed for all diseases where a certain amount of autoimmunity is expected, particularly for suspected autoimmune hepatitis, primary biliary cirrhosis and primary sclerosing cholangitis. Autoantibodies are also present in a high number of patients with chronic HCV infection, however. The antibodies are type--nonspecific and may occur temporarily in healthy people as well. Autoimmune hepatitis is characterized primarily by elevated aminotransferase activity in laboratory results, with marked hypergammaglobulinaemia being typical.

Autoimmune hepatitis is divided into types I, II and III depending on the prevailing type of autoantibodies.

Antibody	AIH type I	AIH type II	AIH type III
ANA	+		±
ASMA	+		±
AAA	+		
pANCA	+		
LKM1		+	
LKM2		±	
LKM3		±	
AMA	±		
LCA		±	
SLA			+
dsDNA	+		

Table 21.2. The basic types of antibodies

Positive antimitochondrial antibodies (AMA, type M2) are typical of primary biliary cirrhosis and occur in up to 95% of all patients with this disease. Some patients also have anti-nuclear antibodies (ANA), which correlate with the prognosis of the disease. Perinuclear antibodies ANCA (pANCA) together with anti-nuclear antibodies (ANA) and anti--smooth muscle antibodies (ASMA) occur in patients with primary sclerosing cholangitis.

CHAPTER 22

22. Laboratory Diagnosis of Jaundice

Authors: Doc. MUDr. Milan Jirsa CSc., Prof. MUDr. Libor Vítek, PhD.

Reviewer: MUDr. MIlan Dastych

22.1. Classification of Hyperbilirubinaemias

Hyperbilirubinaemia may be caused by either a predominantly unconjugated or predominantly conjugated type of bilirubin, depending on the aetiology and pathogenesis of the disease. Predominantly unconjugated hyperbilirubinaemias - pre-microsomal and microsomal types (see *Figure 1*) – pre-hepatic jaundice according to former classification, where the proportion of conjugated bilirubin does not exceed 20% of total bilirubin, are caused by an excessive influx of unconjugated bilirubin to the liver, insufficient unconjugated bilirubin uptake, or reduced conjugation capacity. The impairment of conjugation can be congenital or acquired.

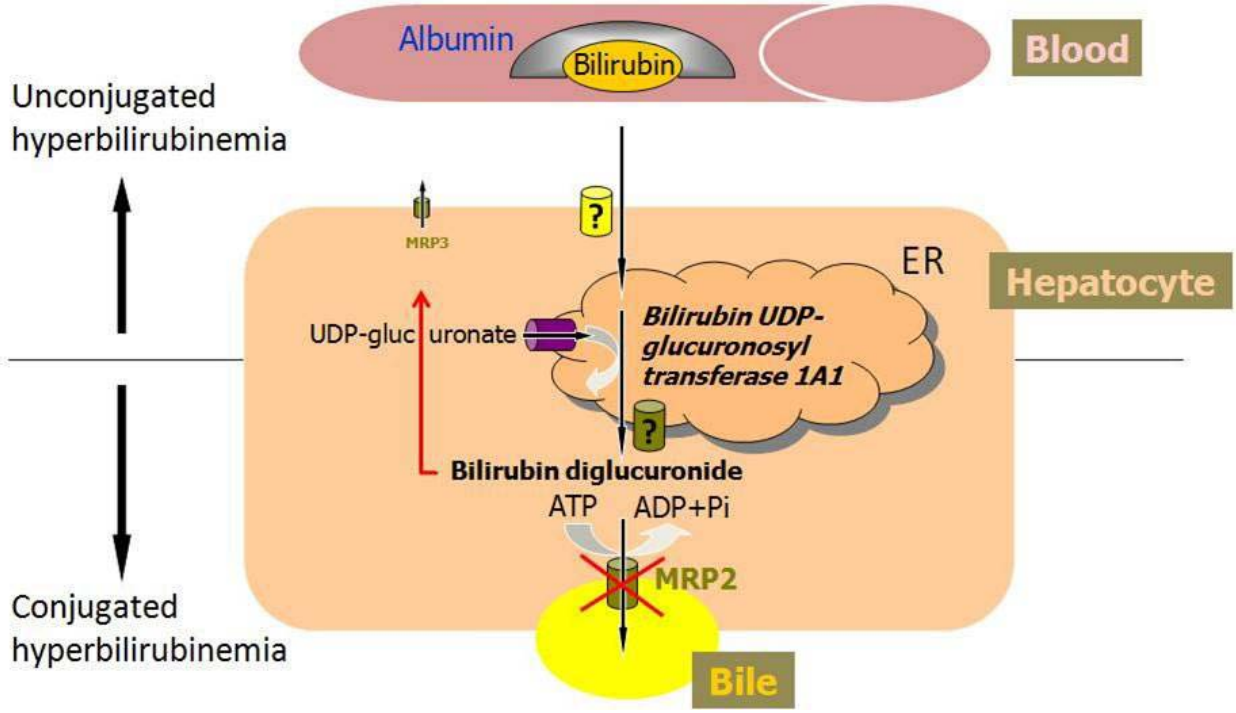


Figure 22.1. Diagram of bilirubin conjugation and secretion in the liver and classification of jaundice. From the biochemical point of view, jaundice can be classified into predominantly unconjugated (direct bilirubin does not exceed 20% of total bilirubin) and predominantly conjugated types (direct bilirubin exceeds 20% of total bilirubin). Mixed jaundice with direct bilirubin from 20 - 50% of total bilirubin is sometimes classified as a separate type. ER – smooth endoplasmic reticulum

The cause of predominantly conjugated hyperbilirubinaemias (post-microsomal type, *Figure 1*, includes hepatic and post-hepatic jaundice according to the former classification), where the proportion of direct bilirubin exceeds 50% of total bilirubin, is the inability to secrete conjugated bilirubin from the liver to the bile, or an obstacle in the bile drainage. Besides congenital disorders of bilirubin metabolism, most predominantly conjugated and mixed hyperbilirubinaemias, where the proportion of bilirubin is 20 – 50% of total bilirubin, are subject to a combination of multiple mechanisms resulting in a variable ratio of conjugated to unconjugated bilirubin.

22.2. Predominantly Unconjugated Hyperbilirubinaemias

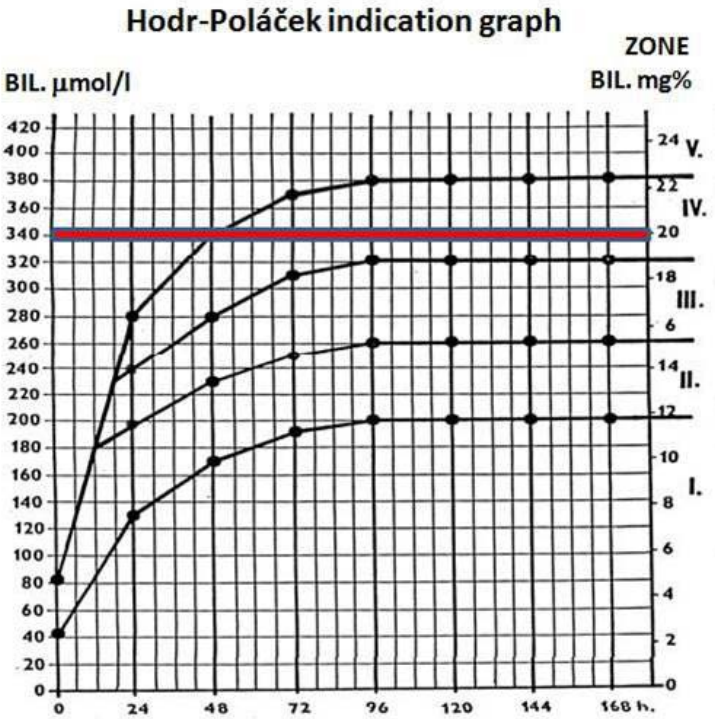
22.2.1. Hyperbilirubinaemia due to unconjugated bilirubin overproduction (pre-microsomal pre-hepatic jaundice)

The common causes of predominantly unconjugated hyperbilirubinaemia are conditions associated with haemolysis. The liver capacity in adults for the uptake, conjugation and secretion of bilirubin is substantial. Therefore, haemolytic icterus occurs at an adult age only due to massive and prolonged haemolysis, and the concentration of unconjugated blood bilirubin usually does not exceed 70 - 100 µmol/L. An exception is acute haemolytic crisis which accompanies some types of haemolytic anaemias, where bilirubin levels can be significantly higher. Bilirubin overproduction based on haemolysis can be found in sickle cell anaemia, hereditary spherocytosis, toxic or allergic reactions to xenobiotics, conditions connected with ineffective erythropoiesis (thalassaemia) and other haematological diseases.

The situation is different in newborn infants. The lower conjugation capacity of the liver and the absence of gut flora catabolizing bilirubin to non-toxic urobilinoids play a role here, in addition to haemolysis. Physiological neonatal jaundice, icterus neonatorum, occupies the boundary between haemolytic icterus and hyperbilirubinaemia based on insufficient bilirubin conjugation. This jaundice afflicts almost one half of all healthy, full-term newborns in the first 5 days of life. It is associated with unconjugated bilirubin levels of around 70 to 80 µmol/L, only rarely with higher levels of up to 220 µmol/L. These values revert to normal within 8 to 10 days, although normal neonatal levels are higher than levels in the adult population. Pathological neonatal hyperbilirubinaemia is characterized by its development in the first 24 hours after the birth, where hyperbilirubinaemia is over 220 µmol/L in full-term or over 256 µmol/L in premature infants, by bilirubinaemia rising by about 85 µmol/L a day, by the proportion of conjugated bilirubin being over 10%, and by the persistence of icterus for one (full-term infants) or two weeks (premature infants). Pathological neonatal hyperbilirubinaemia is an indication for phototherapy, or exsanguination transfusion in extreme cases. The criteria for starting a therapeutic intervention are shown in the Hodr-Poláček graph (*Figure 2*); similar recommendations given by the American Academy of Pediatrics and the UK National Institute for Health and Clinical Excellence are available. Pathological neonatal jaundice is caused mainly by haemolytic conditions induced by Rh incompatibility between the mother and the foetus, and incompatibility in the AB0 system. In addition to which, haemolytic neonatal jaundice may be a sign of a congenital disorder such as spherocytosis or sickle cell anaemia, or result from a toxic or allergic reaction.

The principle of phototherapy consists in the isomerization of bilirubin present in the skin and subcutis by exposure to 425 - 475 nm blue light. The radiation leads to the generation of cis-trans isomers on double bonds in the methylene bridges, which results in breaking the intramolecular hydrogen bonds between propionyl side chains and pyrrole nitrogens, and to the cyclization between rings A and B (Figure 4). Photoisomers have a polar character; they are water-soluble and are excreted in the bile in an unconjugated state. Due to their polar character, photoisomers do not pass the blood-brain barrier and are non-toxic for the newborn's central nervous system. In laboratory assays it is important that, unlike cyclobilirubins, configurational isomers are diazo reactive in common tests for bilirubin. Therefore, serum bilirubin assays may be biased due to phototherapy.

Breast-fed newborns generally have higher serum bilirubin concentrations. This is caused by the presence of UGT1A1 enzyme inhibitor in the breast milk.



> 37 g. w.			< 37 g. w.	
Rh	AB0	other	Rh	other
ET	ET (PhT)	FT (PhT)	ET	ET
ET (PhT)	PhT	PhT	ET	ET
PhT	B 2x	B 2x	ET (PhT)	FT
B 2x	B 2x		PhT	B 2x
B 2x			B 2x	B 1x

Figure 22.2. Hodr-Poláček graph. The nomogram on the left divides postnatal bilirubin level development into five zones. The table on the right shows the therapeutic intervention indications or the frequency of bilirubin assays depending on the zone, age and cause of jaundice. Key: ET – exchange transfusion, PhT – phototherapy, B1x, B2x – bilirubin assays once or twice a day

The effect of this enzyme can be eliminated by heating milk to 56°C. Serum bilirubin levels may reach 250 to 450 µmol/L in the first 10 to 19 days of life. This temporary non-haemolytic unconjugated hyperbilirubinaemia is usually not connected with neurological damage; nevertheless, every hyperbilirubinaemia over 340 µmol/L is considered potentially very dangerous.

22.2.2. Hyperbilirubinaemia due to insufficient unconjugated bilirubin uptake (pre-microsomal hepatic jaundice)

Hyperbilirubinaemia of the predominantly unconjugated type also occurs in toxic and viral hepatocellular injury when hepatocellular icterus develops. This is caused by many hepatocellular noxae such as alcohol, chloroform, tetrachloromethane, acetaminophen, etc. Some types of intoxication involve a combination of the conjugation disorder with cholestasis. Mixed unconjugated and conjugated hyperbilirubinaemia develops in these cases (see below). Depending on the type and extent of liver parenchyma injury, the corresponding response can be seen in liver tests and acute phase reactants.

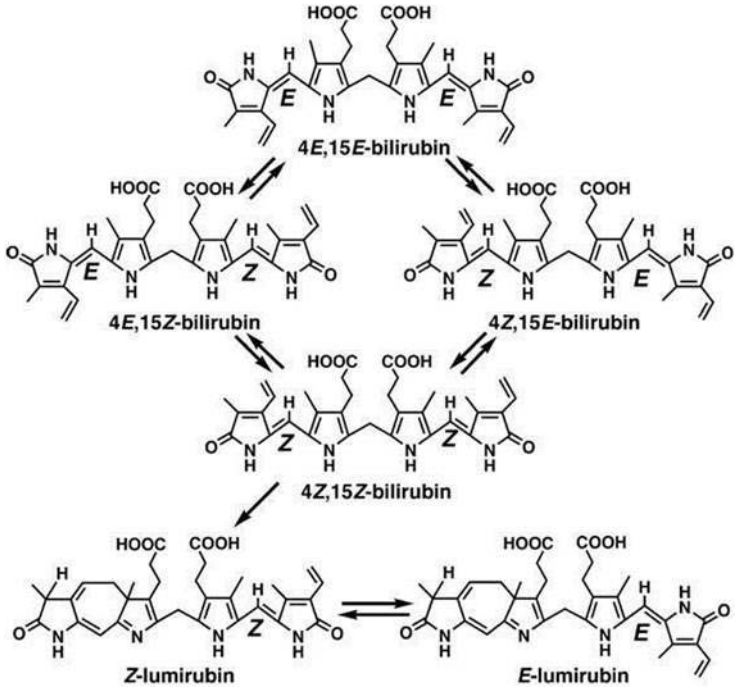


Figure 22.3. Chemical structures of bilirubin IXα, its photoisomers and their photochemical interconversion. Adopted from McDonagh AF et al. Pediatrics 2009;123:67-79

22.2.3. Hyperbilirubinaemia due to unconjugated bilirubin conjugation disorders (microsomal hepatic jaundice)

Gilbert's syndrome (icterus intermittens juvenilis) is characterized by chronic, fluctuating icterus and a slight increase in unconjugated serum bilirubin. Serum bilirubin usually does not exceed 85 µmol/L (exceptionally over 100 µmol/L). The syndrome is caused by reduced activity of hepatic UGT1A1, either due to the reduced expression of the enzyme as a result of gene UGTA1 TATA box promoter prolongation (motif allowing the recognition of the start of transcription) from sequence A(TA)6TAA) to A(TA)7TAA, or directly by mutations in UGT1A1 gene coding regions, resulting in decreased activity of the normally expressed enzyme (Figure 4). Gilbert´s syndrome is an autosomal recessive disorder with incomplete penetration and afflicts about 10% of Europeans. Concurrent UGT1A1 defect with slight haemolysis has also been described. Given the indistinct clinical presentation, the syndrome is often diagnosed by accident based on a biochemical assay where all liver test results except unconjugated bilirubin are repeatedly normal for at least two years.

Crigler–Najjar syndrome is a familial non-haemolytic icterus, an autosomal recessive inherited disorder associated with the total (type I) or partial (type II) defect of enzyme UGT1A1. The type I syndrome is completely lacking conjugated bilirubin excretion, and non-conjugated blood bilirubin levels reach 300 to 800 µmol/L. The defect manifests itself from birth, and afflicted newborns without intensive phototherapy die early due to the involvement of the central nervous system. The administration of conjugation inductors is not therapeutically effective in this type of disorder.

Allele	Exon No.	Consequence	Reference
A[TA]$_7$TAA	promoter	↓ transcription	Bosma 1995
-3279T>G	promoter	↓ transcription	Sugatani 2002
Gly71Arg	1	↓ specif. activity	Koiwai 1995
Phe83Leu	1	↓ specif. activity ??	Sutomo 2002
Pro229Gln	1	↓ specif. activity	Koiwai 1995
Arg367Gly	4	↓ specif. activity	Koiwai 1995
Tyr486Asp	5	↓ specif. activity	Koiwai 1995, Maruo 1998

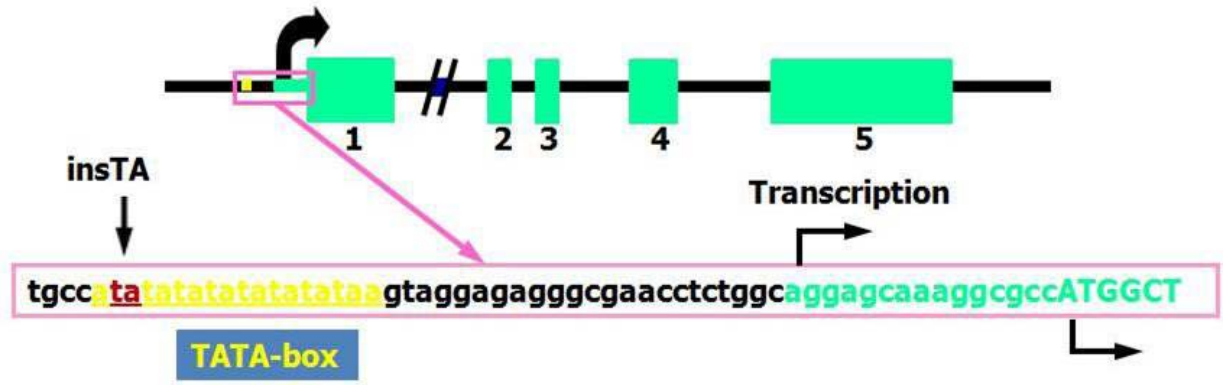


Figure 22.4. Mutations in UGT1A1 gene associated with Gilbert's syndrome. In Europe, by far the most common is the TA insertion in the TATA-box of the UGT1A1 promoter, shown in the chart below.

The defect of conjugation in type II is incomplete, and as a result the progress of the disease is milder, with unconjugated bilirubin concentrations in the blood of 100 - 350 µmol/L. Bilirubin-monoglucosiduronate is found in the bile of patients suffering from this disorder. Reduced conjugation can be induced and the progress of the disease favourably affected by the administration of phenobarbital. Both forms are subject to mutations in the UGT1A1 gene.

22.3. Predominantly Conjugated Hyperbilirubinaemias

22.3.1. Mixed hyperbilirubinaemia due to hepatic parenchyma involvement (post-microsomal hepatic jaundice)

Mixed-type jaundice often accompanies acute and chronic liver diseases such as viral hepatitis, bacterial infections, toxic liver damage, autoimmune hepatitis, metabolic liver diseases and cirrhosis of any aetiology. Jaundice is caused by the concurrent inability of damaged hepatocytes to take up unconjugated and conjugated bilirubin from the blood, conjugate unconjugated bilirubin or excrete it into the bile. It is not rare for jaundice to be accompanied by intrahepatic bile duct involvement. However, jaundice cannot be equated with liver and bile duct diseases. A serious liver disease - hepatitis or advanced cirrhosis - may develop surreptitiously, without manifested jaundice, and the disease is not discovered until detailed laboratory testing is made. In contrast, a rise in bilirubin in patients with hepatic cirrhosis signals the progression of a liver dysfunction. That is why bilirubin, alongside other analytes, was included in scoring systems (Child-Pugh and MELD) (*Figure 5*) used to evaluate liver function and predict the survival rates of patients with advanced liver diseases.

Focal liver diseases usually do not induce jaundice. If present, it is mostly a late trait signalling an advanced disease. The mechanism of jaundice is complex, and it also often involves bile duct obstruction.

MELD = 3.78 x $\log_e$ serum bilirubin (mg/dL) +
11.20 x $\log_e$ INR +
9.57 x $\log_e$ serum creatinine (mg/dL) +
6.43 (constant for liver disease etiology)

NOTES:
If the patient has been dialyzed twice within the last 7 days, then the value for serum creatinine used should be 4.0

Any value less than one is given a value of 1 (i.e. if bilirubin is 0.8, a value of 1.0 is used) to prevent the occurrence of scores below 0 (the natural logarithm of 1 is 0, and any value below 1 would yield a negative result)

Clinical and Lab Criteria	Points*		
	1	2	3
Encephalopathy	None	Mild to moderate (grade 1 or 2)	Severe (grade 3 or 4)
Ascites	None	Mild to moderate (diuretic responsive)	Severe (diuretic refractory)
Bilirubin (mg/dL)	< 2	2-3	>3
Albumin (g/dL)	> 3.5	2.8-3.5	<2.8
Prothrombin time			
Seconds prolonged	<4	4-6	>6
International normalized ratio	<1.7	1.7-2.3	>2.3

Child-Turcotte-Pugh Class obtained by adding score for each parameter (total points)
Class A = 5 to 6 points (least severe liver disease)
Class B = 7 to 9 points (moderately severe liver disease)
Class C = 10 to 15 points (most severe liver disease)

Figure 22.5. Child-Pugh (down) and MELD (up) scoring systems.

22.3.2. Hyperbilirubinaemia due to conjugated bilirubin secretion disorders (post-microsomal hepatic jaundice)

Dubin-Johnson syndrome is a benign, autosomally and recessively inherited, predominantly conjugated hyperbilirubinaemia. Apart from jaundice, its symptoms are indistinct, which is why laboratory testing is primarily used for diagnosis. The serum exhibits distinctive hyperbilirubinaemia (usually 35 to 85 µmol/L, exceptionally 350 to 400 µmol/L), 50% of which is conjugated bilirubin. Increased excretion of coproporphyrin I in the urine is also typical, but total porphyrin output is not increased. Other laboratory findings are within the normal range. Patients' livers are macroscopically dark and the presence of black pigment, probably related to melanin, is demonstrated histologically in hepatocyte lysosomes.

Dubin-Johnson syndrome is caused by a defect in conjugated bilirubin secretion from the hepatocyte into the bile due to mutations in the ABCC2 gene. The secretion of glutathione bromosulphophthalein conjugates and other anion pigments used previously for liver function tests are affected in a similar way. Today, they are substituted by cholescintigraphy with radiopharmaceuticals as excreted anions. The secretion of bile acids is unaffected.

Another rare chronic benign, predominantly conjugated, hyperbilirubinaemia is Rotor syndrome. In this syndrome, too, the other common haematology and biochemistry tests are in the normal range. However, the liver lacks any pigmentation and up to five times the normal amount of coproporphyrin is excreted in the urine, of which more than 65% is coproporphyrin I. Cholephilic anion pigments are excreted from the blood very slowly (*Figure 6*). Again, the disease is familial, with autosomal recessive trait of inheritance. The disease is subject to the concurrent presence of mutations

in the genes for OATP1B1 and OATP1B3, and consists in the absence of conjugated bilirubin and other organic anion uptake by the liver; the secretion of conjugated bile salts is maintained, however.

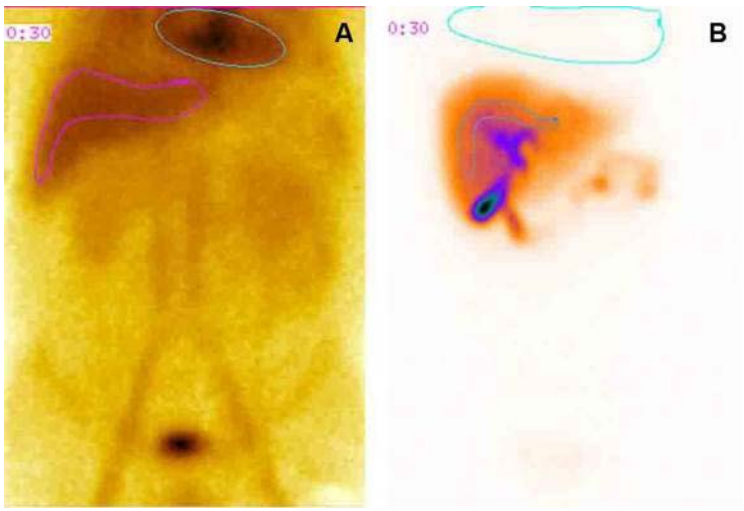


Figure 22.6. Cholescintigram: Rotor syndrome (A). Panel B shows a normal finding without signs of hepatic chromosecretion disorder. Adopted from Hřebíček M, et al. Liver Int 2007;27:485-91.

22.3.3. Hyperbilirubinaemia due to intrahepatic and extrahepatic bile drainage disorders (post-hepatic jaundice)

Obstacles in the bile drainage may be caused by mechanical obstructions or inflammation of hepatic or extrahepatic bile ducts. Bile accumulation is often connected with hyperbilirubinaemia and obstructive icterus.

Intrahepatic cholestasis is caused through different mechanisms by many pharmaceuticals, therapeutic hormones, natural substances and other noxae: oestrogens and C-17 alkylated steroids (ethinyl oestradiol), tauro and glycolithocholate, chlorpromazine, cyclosporin, colchicine, phalloidin and others. Oestrogens and progesterone take part in the pathogenesis of intrahepatic cholestasis of pregnancy. These substances affect the expression, localization and activity of each canalicular transporter. The cholestatic effect of many substances is subject to their genetic predisposition.

Conjugated bilirubin, conjugated bile acids, enzymes and other hepatocyte proteins are transferred to the blood in intrahepatic cholestasis or biliary obstruction. Continuous, although limited, transfer of these substances to the blood, mediated by carriers localized in the basolateral membrane of hepatocyte, takes place in physiological conditions as well. During cholestasis, other carriers that also take the role of canalicular transporters (e.g. MRP3 relative to canalicular MRP2 with decreasing expression) are expressed in the basolateral membrane, and the cell loses its original transporting polarity. The basolateral carriers then intensively take away toxic metabolites from the hepatocyte and transfer them to the blood.

The transfer of membrane proteins and enzymes such as alkaline phosphatase (ALP), gamma-glutamyl transferase (GGT) and 5´-nucleotidase to the blood in intrahepatic or extrahepatic cholestasis is explained by damage of bile duct epithelium due to the detergent effect of bile acids and subsequent regurgitation of bile into the blood circulation through chemically and mechanically (by compression) damaged cell junctions. The mode of cytosolic and mitochondrial protein and enzyme transfer from the hepatocyte to the blood is even less clear. Increased permeability of cell membranes caused by the primary pathological process probably plays a part.

Familial benign recurrent intrahepatic cholestasis is a disease diagnosed most often in adolescence. Nevertheless, it afflicts all age categories and is characterized by attacks of cholestasis with jaundices and intermittent asymptomatic periods. At the preicteric phase, the bile acid concentration in the blood rises and alkaline phosphatase activity also increases. Increased conjugated bilirubin is tested in the serum at the icteric phase, aminotransferase activity may be slightly increased and the prothrombin time is prolonged due to vitamin K malabsorption. Unlike most other cholestatic liver diseases, GGT activity is normal as a result of reduced bile salt secretion. Laboratory findings between attacks are within the normal range. In most patients the disease is caused by mutations in the FIC1 gene (familial intrahepatic cholestasis type 1), whose role in the mechanism of bile salt and bile pigment secretion is not yet known.

Extrahepatic cholestasis (obstructive icterus) in adults is most commonly caused by a stone stuck in the bile duct or a tumour. Other causes include strictures (strictures due to primary or secondary sclerosis cholangitis, caused by

inflammation around the bile duct and in the ampulla of Vater region, post-operative or iatrogenic strictures) and compression of the bile duct (by tumour, node, etc.). Biliary atresia should be considered at neonatal age and infancy.

22.4. Laboratory assays for differential diagnosis of jaundices

Basic biochemistry testing, including direct and total serum bilirubin, urinary bilirubin and urobilinogen, and the activity of aminotransferases (ALT and AST) and cholestatic enzymes (ALP and GGT) in the serum, is usually the first step in the differential diagnosis of jaundice following the patient history and physical examination. In the case of predominantly unconjugated isolated hyperbilirubinaemia, where ALT, AST, ALP and GGT levels are normal, haemolysis should be confirmed or ruled out. Blood count and reticulocyte count are used for this purpose. The differential diagnosis of haemolytic anaemias is then the domain of haematology. The diagnosis of inherited bilirubin conjugation disorders can be confirmed by molecular genetic tests. A result testifying to obstruction of the bile duct is followed by ultrasonic examination and other imaging (CT, MRCP) or endoscopic (ERCP) techniques focused on diagnosing and/or treating bile duct disease. Ultrasonography of the bile duct is also used for hepatic parenchyma diseases. If the ultrasound finding is normal, tests focused on liver disease aetiology follow. The basic differential diagnosis chart for jaundice based on biochemistry findings is shown in *Figure 7*.

Figure 22.7. Laboratory findings in each type of jaundice.

	Type of jaundice			
	pre-hepatic and microsomal	hepatocellular	Dubin-Johnson and Rotor syndrome	post-hepatic
S - total bilirubin	↑	↑	↑	↑
S - direct bilirubin	norm.	↑	↑	↑
U - bilirubin	+	+	+	+
U - urobilinogen	-	+	-	-
S - ALT	norm.	↑	norm.	norm.
S - AST	norm.	↑	norm.	norm.
S - ALP	norm.	norm.	norm.	↑
S - GGT	norm.	norm.	norm.	↑

CHAPTER 23

23. Bone Metabolism

23.1. Bone Metabolism – Preparation

Author: *MUDr. Richard Pikner, Ph.D.*

23.1.1. Introduction

This chapter provides an overview of the bone structure, bone remodelling process, bone remodelling regulation factors and options for bone metabolism laboratory tests. The key information offers a summary of the basic knowledge from this chapter.

23.1.2. Function and Structure of Bones

The human body is made up of 220 bones, which constitute about 15% of the body weight. Bone is not only the basis of the musculoskeletal system, but it is also a highly metabolically active tissue, permanently degrading (osteoresorption) and recreating (osteoformation). This process is called bone remodelling.

Bone remodelling provides a replacement for and repairing of damaged bone parts (microfractures and macrofractures) as well as adjustment of the bone shape to long-term mechanical load, to which athletes are exposed to, for example.

The bone has the following 3 basic functions in the human body:

Structural – biomechanical support of the limbs and protecting cavities in the body, points of attachment for muscles and tendons

Haematopoietic – the bone provides a spatial matrix for haematopoietic cells.

Metabolic – as a deposit for calcium, magnesium, sodium and phosphates, the bone contributes to regulating calcium and phosphate homeostasis and phosphate buffer systems

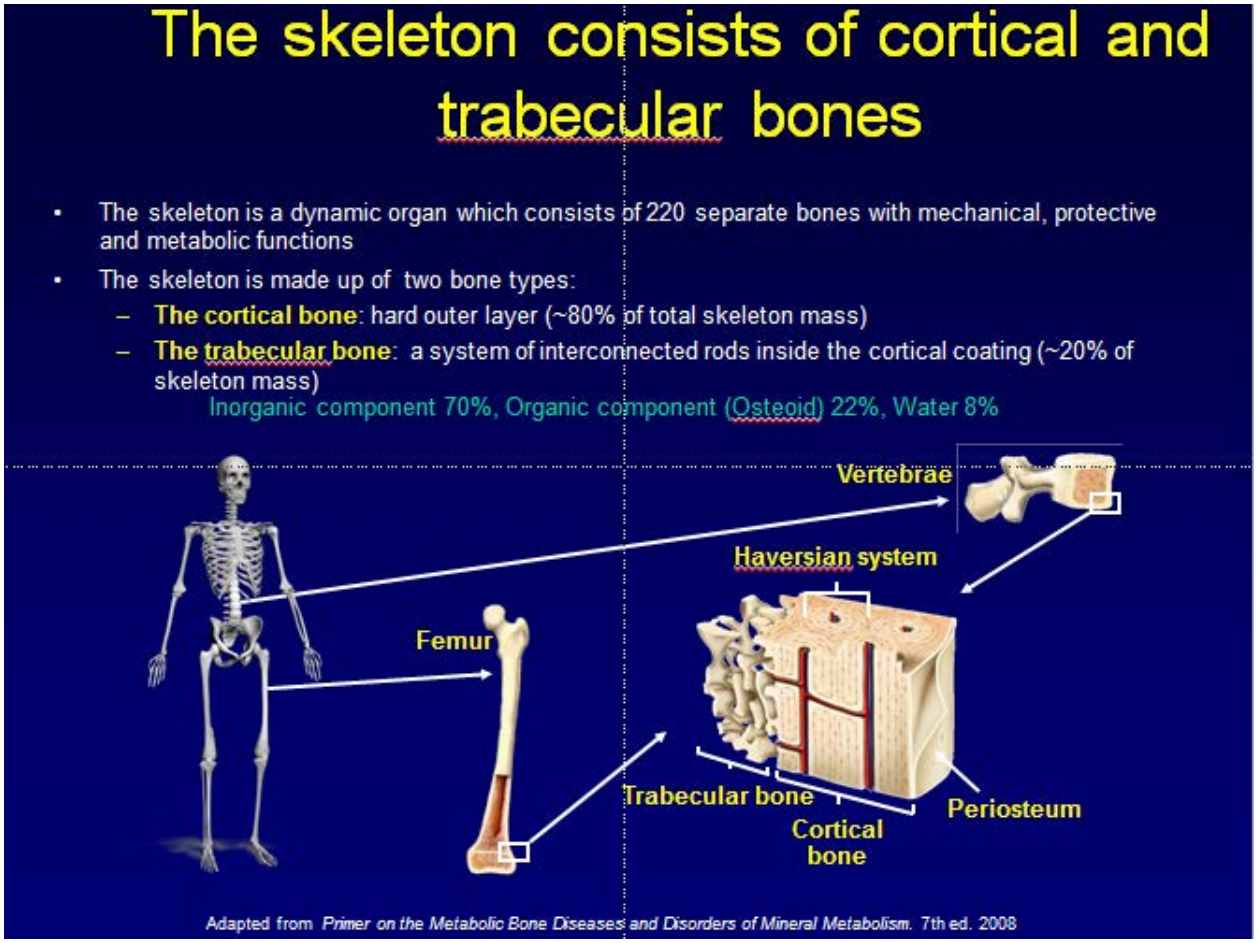


Figure 23.1. The skeleton consists of cortical and trabecular bones

There are two basic types of bone:

1. cortical bone
2. trabecular bone

The cortical (compact) bone makes up about 80% of overall bone mass, consists of dense layers of mineralised collagen (mineralisation is up to 90%), whose main function it is to provide biomechanical support and structure. Microscopically, the cortical bone consists of regular lamellae arranged concentrically around blood vessels, referred to as the Haversian lamellae. The Haversian canal with a vascular bundle is found in the middle. The lamellae with the canal form the basic structural unit, the osteon. Osteocytes, mutually interconnected bone cells, lie between osteons. The cortical bone is the main component of long bones and forms a protective coating for the other bones. The cortical bone's remodelling activity is 10 times less than that of the trabecular bone. About 25% of trabecular and only 3% of the cortical bone is remodelled annually in the human body.

The trabecular (cancellous) bone, conversely, makes up less than 20% of the bone mass, yet assumes a prevailing part in the overall metabolic activity of bones due to its much greater surface. Mineralisation is from 5 to 20%. Microscopically it is formed from a spongy structure consisting of individual interconnected trabeculae. The orientation and thickness of trabeculae depends on the long-term mechanical load of the given site. The trabecular structure markedly enlarges the overall surface, thus allowing a relatively fast metabolic exchange.

Different bones have different contents of bone types. Lumbar vertebrae contain 70% of trabecular bone, the heel 75%, proximal femur 50 - 75%, and the distal radius 25% of the trabecular bone. As corticoids primarily suppress the trabecular bone's remodelling activity, the vertebrae suffer from pathological compression. Primary hyperparathyroidism causes considerable demineralisation of the cortical bone, which can be documented in an X-ray image of the distal radius.

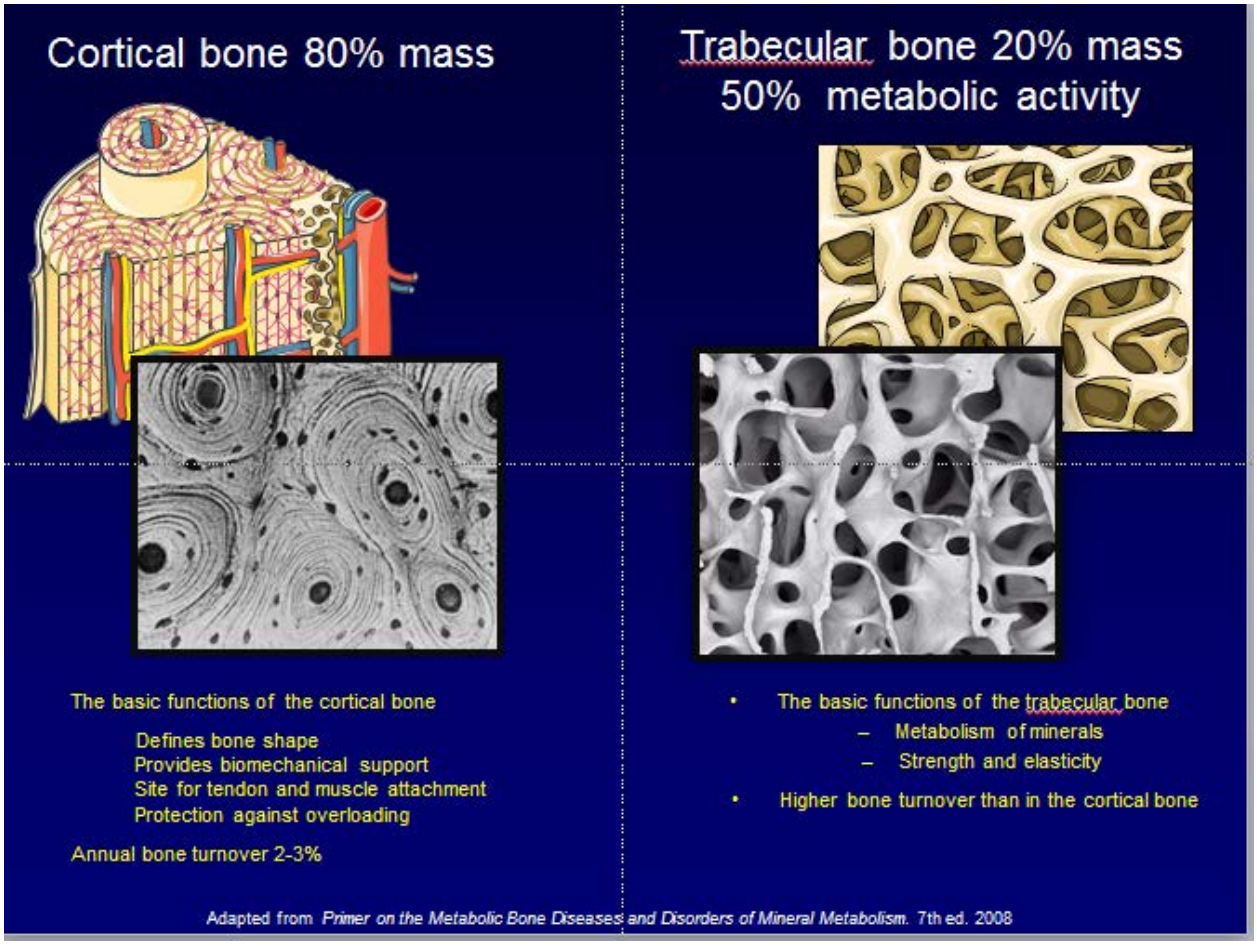


Figure 23.2. The stucture of cortical and trabecular bone

23.1.2.1. Composition of Bone

Bone is made up of water (8%), and inorganic (70%) and organic (22%) components. The basic part of the inorganic component is hydroxyapatite $Ca_{10}(PO_4)_6(OH)_2$ and mineral salts such as calcium phosphate, calcium carbonate, magnesium phosphate and magnesium carbonate, zinc, fluoride and strontium. 90% of the organic component (osteoid) is made up of collagen I, however it also contains collagen V, VI, VIII and XII and other non-collagen molecules, 20% of which is made up of osteocalcin and the rest osteopontin, vitronectin, sialoprotein, heparan and chondroitin sulphates, etc.

- Collagen I

Collagen I contains two identical 1-chains, and the third, different 2-chain. It is produced by osteoblasts in the form of individual pro-chains (procollagens); the C and N-terminal ends are cleaved off extracellularly and the typical triple helix stabilised by many cross-linking connections of pyridinoline and deoxypyridinole is formed. These links form by combining hydroxylysine or lysine residues from chains, and provide stability to the triple helix. The cleaved-off N and C-terminal ends are released to the serum and can be detected here as PICP or PINP (new bone formation markers). Collagen I is degraded from the bone using proteolytic enzymes: cathepsin K (osteoclasts) or matrix metalloproteinase 9 (MMP9 from tumour cells). The product of degradation is collagen fibres, released into the circulation. They can be detected in the plasma or serum as CTX, NTX or ICTP (bone resorption markers).

- The bone tissue contains 3 types of cells:

3. Osteoblasts
4. Osteocytes
5. Osteoclasts

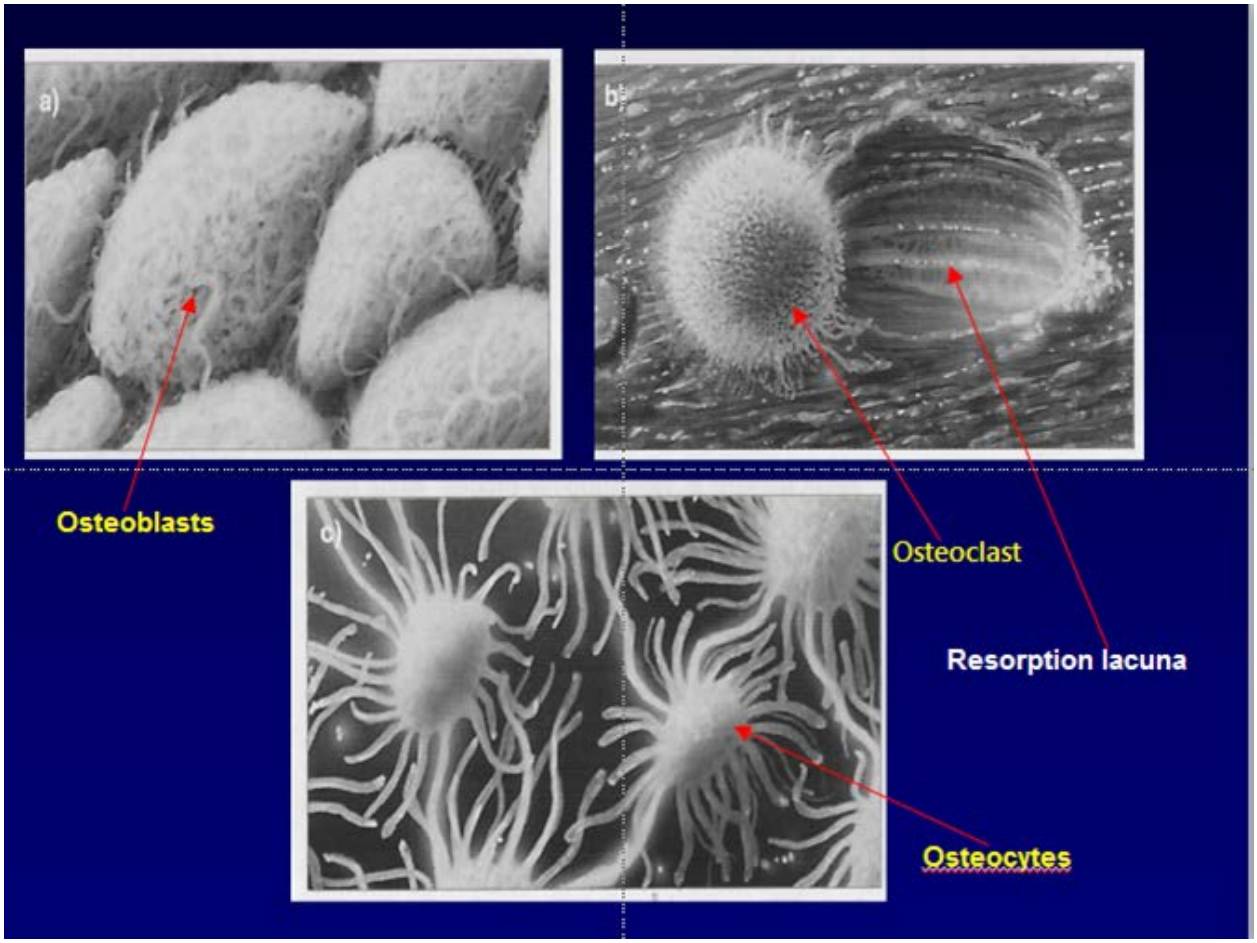


Figure 23.3. Osteoblasts, osteoclasts and osteocytes view in microscope

6. Osteoblast is the basic cell for new bone formation. Osteoblasts form in precursor mesenchymal cells (common stem cell for the myoblast, myocyte, adipocyte, chondrocytes and osteoblasts) in bone marrow. Differentiated osteoblasts produce a new bone matrix (osteoid) and initiate its mineralisation. A part of osteoblasts caught up in the newly formed bone stays here and creates osteocytes. The osteoblast is the basic cell for bone remodelling regulation and produces local regulation cytokines RANKL and osteoprotegerin (OPG). Its surface bears receptors for most systemic hormones (PTH, oestrogens, androgens, cortisol, etc.). Hormonal stimulation and local regulation factors result in regulation of the bone remodelling level.
7. Osteocytes are formed from osteoblasts located between cortical lamellae and inside trabeculae. Osteocytes have minimal synthetic activity, however they use cytoplasmic projections to create a cross-linked spatial network. They act as mechanosensors in the bone. Osteocytes are activated by a permanent load change or microdamage to their mutual links. They are a source of local signalling molecules which activate or inhibit the bone remodelling process in the given area. A good example is long-term inactivity. Mechanical load in bedridden patients is reduced, which results in a very fast osteoresorption in the unloaded bones, while regular physical strain leads to local remodelling resulting in increased new bone formation (gain) and altered orientation of the trabeculae in the direction of the mechanical load.
8. Osteoclasts come from the hematopoietic myeloid stem cell (common stem cell for erythrocytes, megakaryocytes, monocytes and granulocytes). Osteoclast comes from the monocytic line and is a specialised macrophage. The osteoclast is a multinucleate cell formed by fusing pre-osteoclasts from monocytic cells. Following contact with the bone surface, osteoclasts create a specially structured part of the plasma membrane called a ruffled border. This ruffled part of the membrane creates a space where H-K-ATPase secretes hydrogen protons. Together with other enzymes (cathepsin K), this acidic environment with a pH of about 4 creates the extracellular lysosome which degrades the adjoining bone. The result is the resorption lacuna - a space left after the completely resorbed bone. Osteoclasts bear receptors for local signalling molecules on their body. The only systemic hormonal receptor they have on their surface is the receptor for calcitonin. Therefore, calcitonin directly inhibits osteoclast activity.

23.1.2.2. Bone Remodelling

Bone remodelling is a continuous process of activation of bone resorption and new bone formation involving all of the three bone cell types. This results in a new bone protein matrix (osteoid), which is subsequently passively mineralised. In bone remodelling, osteocytes play the role of mechanosensors and also trigger the bone remodelling cycle. Any mechanical damage to osteocytes activates the remodelling cycle at the site of injury (osteoresorption) followed

by new bone formation activation. The balance between osteoresorption and new bone formation is controlled by osteoblasts.

One site in the bone where the remodelling cycle takes place is referred to as the bone metabolic unit (BMU). Millions of BMUs at different stages of activity are concurrently present in the human body.

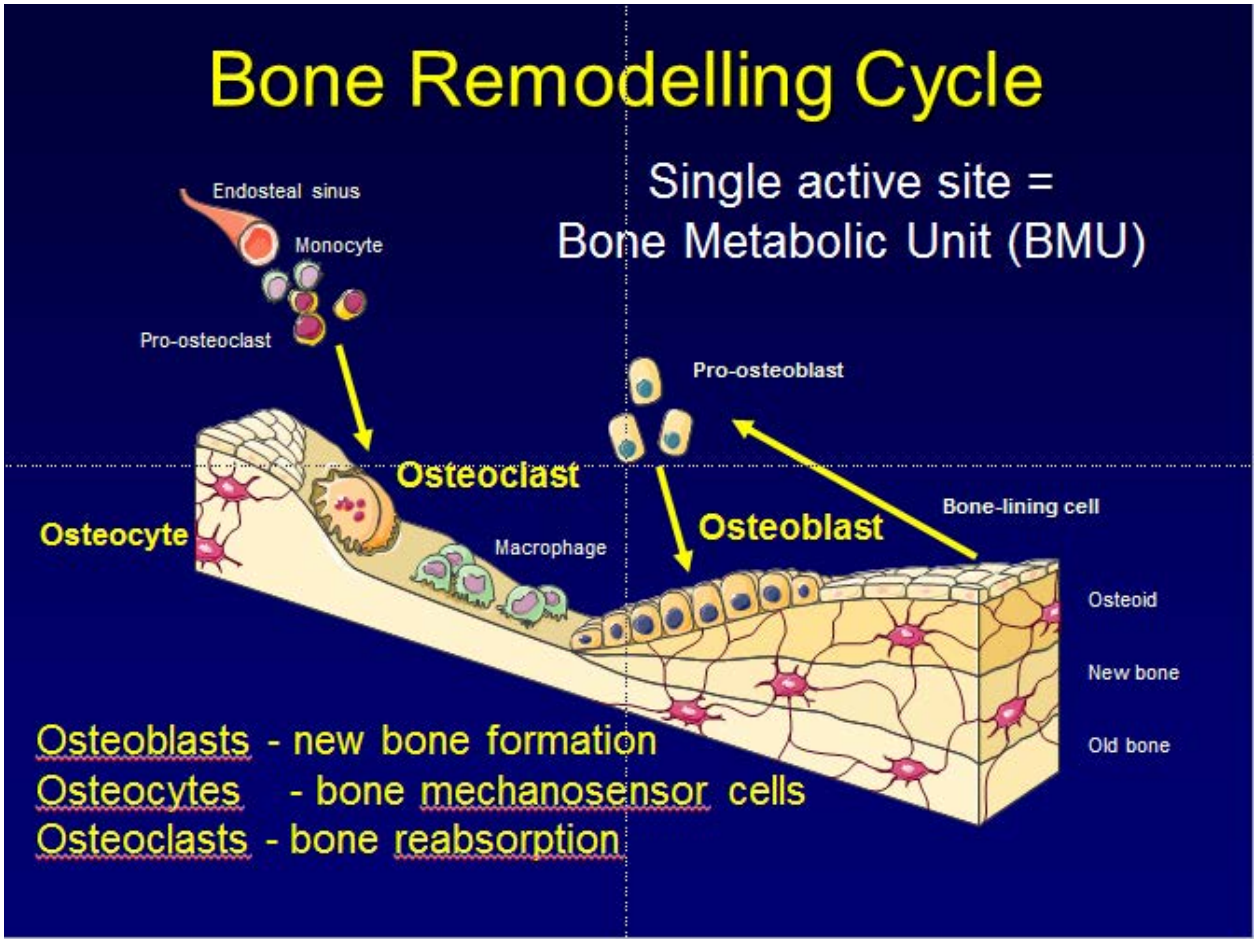


Figure 23.4. Bone remodelling cycle

- The remodelling process lasts about 180 days and may be divided into several phases:

Activation phase – bone cells are in the idle state. Osteocytes become activated through the action of different stimuli (load, damage, hormonal signals). Osteocytes reduce osteoblast activity by means of sclerostin and, conversely, begin to stimulate the formation of osteoclasts by M-CSF and RANKL cytokines. The formation of active osteoclasts is supported and induced by the production of RANKL in osteoblasts, concurrently decreasing the production of its natural inhibitor, OPG. This effect is the result of sclerostin activity and other hormonal influences.

The osteoresorption phase lasts about 10 – 14 days; M-CSF and RANKL activate the differentiation of monocytic cells into pre-osteoclasts and osteoclasts, and the bone in resorbed. OPG (osteoprotegerin) is the regulator of the degree of osteoresorption; it is a natural competitive inhibitor of RANKL. Bone resorption is further stimulated by TNF-1, IL-1, interferons-gamma and prostaglandins. Proteolysis of the bone matrix as well as the release and activation of TGF--ß1 take place during bone resorption. TGF-ß1 activates osteoblasts and the transition to the next phase.

New bone formation phase - Osteoblasts producing the bone protein matrix (osteoid) are responsible for new bone formation followed by the passive process of primary osteoid (50% of minerals) mineralisation. This process lasts for around 150 days.

Resting phase – after the primary mineralisation phase, the newly formed bone further mineralises (the percentage of crystals increases) provided there is a sufficient supply of minerals and vitamins. This process is called secondary mineralisation, and the time required for reaching 100% absorption capacity of the bone protein matrix is not known. The bone remains in the resting phase until the remodelling cycle in the give place has been activated.

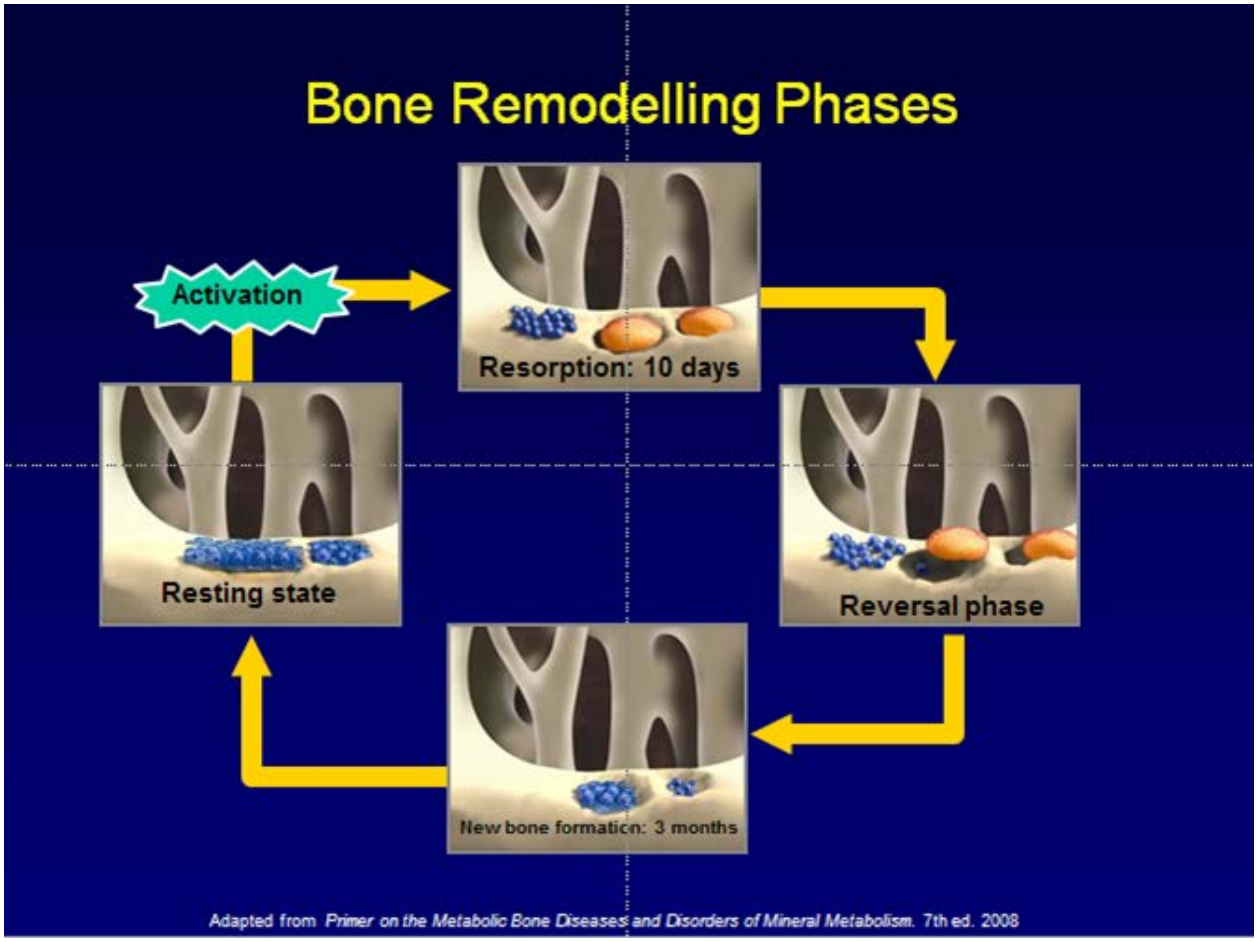


Figure 23.5. Bone remodelling phases

23.1.2.3. Bone Remodelling Regulation

- The remodelling process is regulated by:

Systemic hormones: parathormone, calcitriol, calcitonin, FGF23, oestrogens, IGF1, leptin, thyroid hormones, cortisol

Systemic cytokines: IL-1, IL-6, TNF, OPG, RANKL, prostaglandins, etc.

Local cytokines: signalling molecules identical to systemic and specific molecules such as sclerostin, BMPs (bone morphogenetic proteins)

Intercellular communication between osteoclasts and osteoblasts

Minerals (calcium, magnesium, zinc) and *vitamins* (K, C, B6, A): sufficient amounts are essential for correct osteoid mineralisation

Osteocytes trigger the remodelling cycle by RANKL and sclerostin production. RANKL activates osteoclasts and sclerostin inhibits osteoblast activity. However, the main regulator is osteoblasts mediating the action of osteocytes and other hormones.

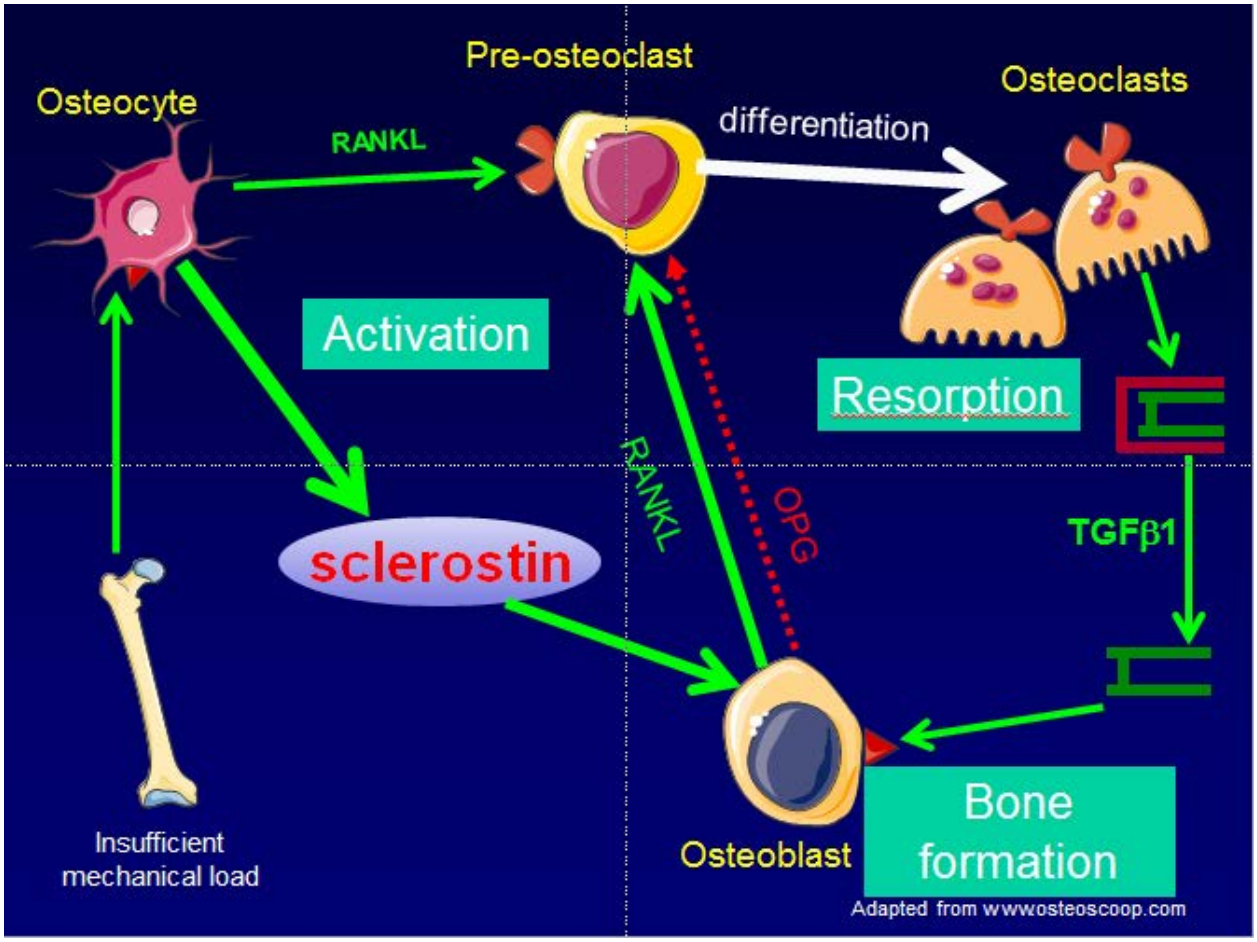


Figure 23.6. The interaction of cells and sclerostin in the bone formation

Osteoblasts and osteoclasts are key players in bone remodelling. The result of remodelling depends on their co-ordinated interaction. If the regulation favours osteoclast activity, bone tissue is lost. This happens in osteoporosis. Conversely, if the regulation favours new bone formation, bone tissue increases. This is physiological in the period of growth. The peak bone mass is subsequently the result of nutritional, genetic and hormonal factors together with mechanical load during the growth period with subsequent secondary mineralisation. The human body reaches peak bone mass between 25 and 30 years of age.

Osteoblasts are superior to osteoclasts in regulation. Their interaction is coordinated through osteoblasts. Osteoblasts and osteoclasts are located close to each other in the bone. Their coordination is based on:

Intercellular communication – this close intercellular communication is provided by epinephrines.

The *RANKL-OPG-RANK* system is the key regulatory mechanism for the remodelling unit's activity. The system consists of the RANK receptor on osteoclasts. When the receptor is stimulated, pre-osteoclasts differentiate into osteoclasts and their metabolic activity is maintained. RANKL stimulates its activity, while OPG is a competitive inhibitor. Both RANKL and OPG are produced by osteoblasts, and their proportion is decisive for the activation or inhibition of osteoclasts. Systemic hormones transfer their signal to the bone by inducing the production of RANKL and OPG in different ways. Oestrogens increase the production of OPG and inhibit the production of RANKL, meaning they have an anabolic effect on the bone. Corticoids increase the production of RANKL, decrease the production of OPG, and increase bone resorption. M-CSF triggering and maintaining the production of osteoclast precursors is also an important factor.

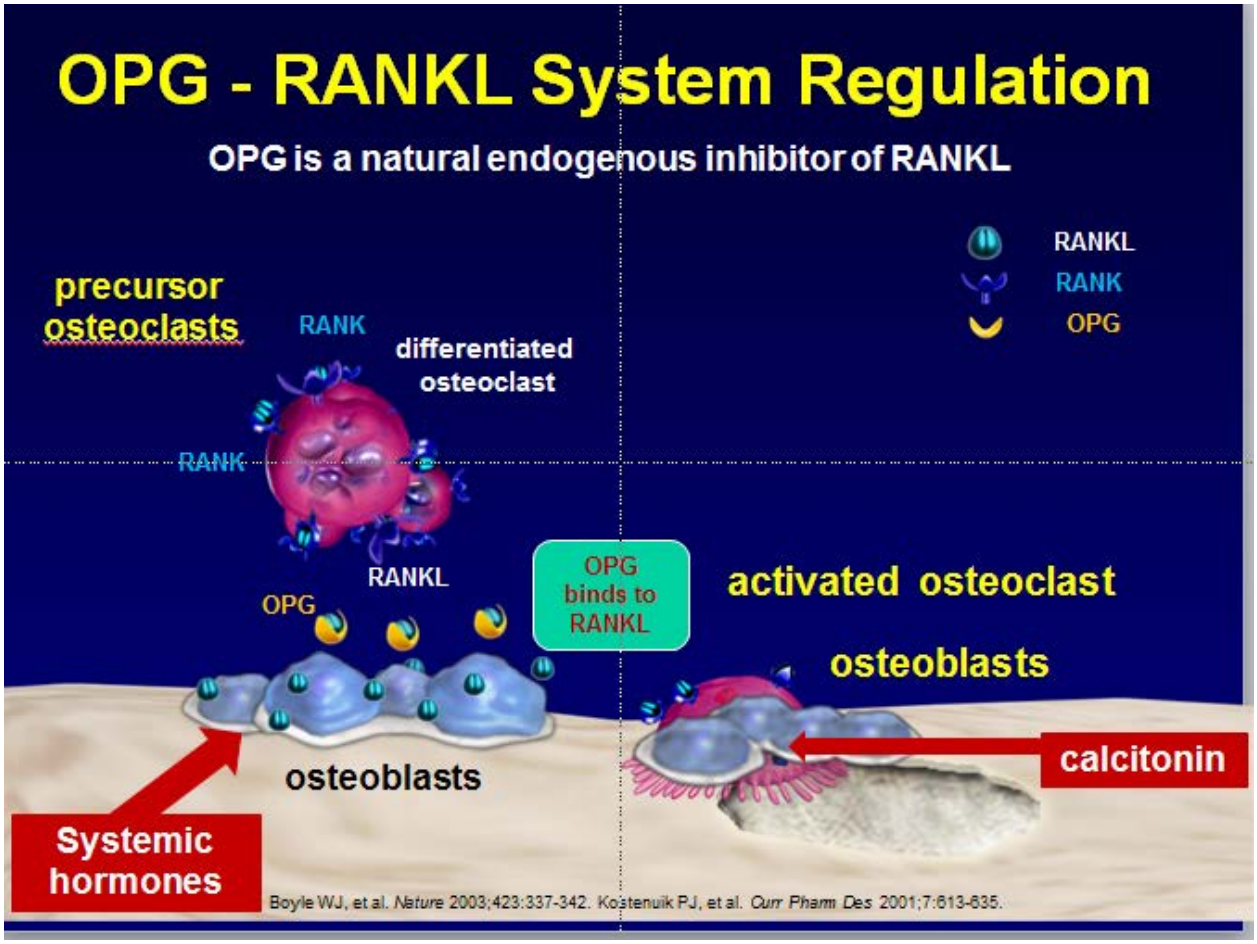


Figure 23.7. Regulation of the osteoblasts activity – systém OPG-RANKL

Apart from osteoblasts, RANKL and OPG are also produced by immune system cells, which confirm the connection between bone remodelling and immune system activity.

In addition to the low sclerostin level, osteoblasts are stimulated by TGF-ß1 released by osteoresorption activity of osteoclasts from the bone matrix. Bone remodelling is the result of a multifactorial effect of hormones, cytokines and other agents which finely balance the system to maintain stable bone mass while enabling an intense exchange of mineral components (Ca, Mg, phosphates) with the extracellular fluid as well as repair processes.

23.1.2.4. Hormonal Factors in Regulating Bone Remodelling

Parathyroid hormone – has an indirect effect on osteoclasts through receptors for the PTH on osteoblasts; this results in *increased production of RANKL* and *decreased OPG* followed by increased osteoresorption, all this assuming continuous stimulation of PTH like in primary hyperparathyroidism.

Intermittent administration of the PTH once daily not only induces the stimulation of osteoresorption but above all leads to greater stimulation of osteoblast differentiation. This results in an anabolic effect on the bone, where new bone formation is stimulated more than osteoresorption.

Sex hormones:

- Oestrogens – also act through osteoblasts; suppress RANKL production and stimulate OPG secretion, acting on mesenchymal cells at the same time, thus suppressing the production of other osteoclast-stimulating cytokines such as M-CSF, IL-1, IL-6 and TNF-α, meaning they act at the new bone formation phase – during matrix formation.

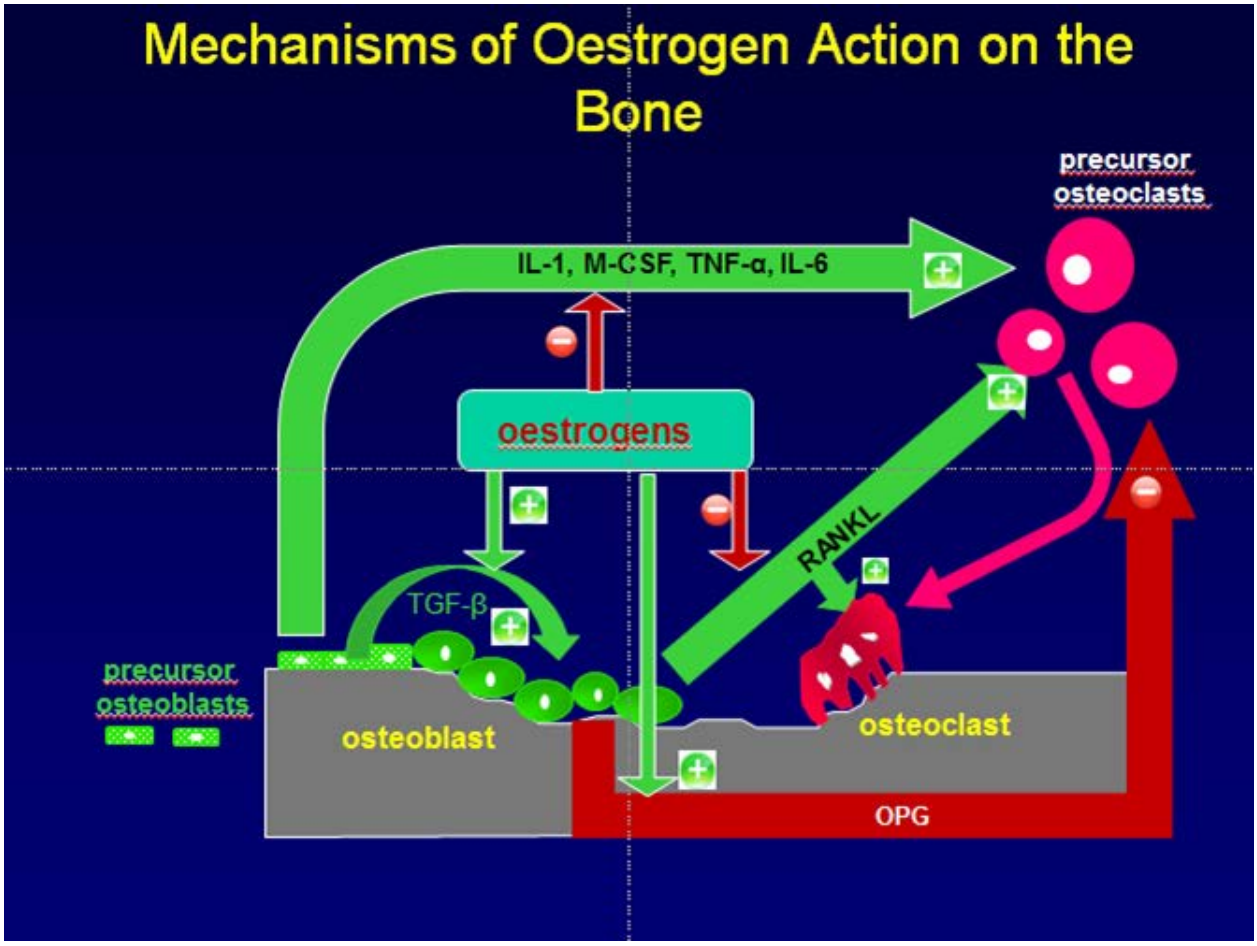


Figure 23.8. Mechanisms of oestrogen action on the bone

- Androgens primarily act on bone mineralisation and support periostal bone apposition. Differences in hormone concentrations between genders lead to different bone shape remodelling. Through the action of androgens, men have a larger diameter of supporting bones and a thinner cortical bone with predominant periostal apposition than women.
- IGF1 stimulates osteoblast differentiation.
- Thyroid hormones stimulate bone remodelling with a stronger effect on osteoresorption.
- Corticoids – stimulate osteoresorption through increased RANKL production and decreased OPG production and decrease collagen I formation at the same time.
- Calcitonin – as the only hormone directly inhibits osteoclast activity and induces their apoptosis.

23.1.3. Summary – Key Information

The bone has the following *3 basic functions* in the human body: *structural* – biomechanical support of the limbs and the protection of body cavities, points of attachment for muscles and tendons; *haematopoietic* – the bone provides a spatial matrix for haematopoietic cells, and *metabolic* – as a deposit for calcium, magnesium, sodium and phosphates, the contributes to regulating calcium and phosphate homeostasis and phosphate buffer systems.

- The cortical (compact) bone makes up roughly 80% of overall bone mass and less than 50% of metabolic activity, while the trabecular (cancellous) bone makes up less than 20% of the bone mass, but more than 50% of metabolic activity.

Bone is made up of water (8%), and inorganic (70%) and organic (22%) components. The basic part of the inorganic component is hydroxyapatite $Ca_{10}(PO_4)_6(OH)_2$ and mineral salts of Ca, Mg, Sr, F and Zn. 90% of the organic component (osteoid) is made up of collagen I, however also contains other non-collagen molecules such as osteocalcin (20%).

The bone tissue contains 3 cell types: 1. Osteoblasts, 2. Osteocytes, 3. Osteoclasts.

- The osteoblast is the basic cell for new bone formation and a cell regulating the bone remodelling process. The osteoblast produces osteoclast-stimulating cytokines (RANKL) and its natural inhibitor, OPG. Most systemic hormones (PTH, oestrogens, androgens, cortisol etc.) control bone remodelling by influencing on the production of RANKL and OPG in osteoblasts.

- Osteocytes are formed from osteoblasts located between cortical lamellae and inside trabeculae. They create a cross-linked spatial network acting as mechanosensors. Osteocytes are the source of signals triggering the bone remodelling process in the given area.
- Osteoclasts come from the monocytic line and are actually specialised macrophages. The osteoclast is a multinucleate cell able to adhere to the bone surface and degrade bone tissue beneath the membrane. The osteoclast needs RANKL for its activity; conversely it is inhibited by OPG or calcitonin as the only systemic hormone.
- Bone remodelling is a continuous process of activation of bone resorption and new bone formation involving all the three bone cell types. This results in a new bone protein matrix (osteoid), which is subsequently mineralised secondarily.
- The remodelling process lasts about 180 days and can be divided into some phases:

Activation phase – the action of stimuli activates osteocytes, and subsequently osteoclasts and osteoblasts.

Osteoresorption phase – lasts around 10 – 14 days, osteoclasts resorb the bone.

New bone formation phase - active osteoblasts produce the bone protein matrix (osteoid); this is followed by the passive process of primary osteoid mineralisation (50% of minerals). This process lasts for around 150 days.

Resting phase – the newly formed bone further mineralises. This is a secondary mineralisation process with an unknown duration.

One site in the bone where the remodelling cycle takes place is referred to as the bone remodelation unit (BRU). Millions of BRUs at different stages of activity are concurrently present in the human body.

- Bone remodelling regulation: Osteocytes activate bone remodelling and osteoblasts control it using the RANKL/OPG-RANK system. The RANKL – OPG - RANK system is the key regulatory mechanism of the remodelling unit's activity. The system consists of the RANK receptor on osteoclasts. When the receptor is stimulated, pre-osteoclasts differentiate into osteoclasts and their metabolic activity is maintained. RANKL stimulates its activity, while OPG is a competitive inhibitor. Both RANKL and OPG are produced by osteoblasts, and their proportion is decisive for the activation or inhibition of osteoclasts.

Systemic hormones transfer their signal to the bone by inducing the production of RANKL and OPG in different ways.

23.2. References

9. R.Eastell, Bone Markers, Martin Dunitz Ltd, London 2001
10. I.Sotorník, Š. Kutílek a kol., Kostní minerály a skelet při chronickém onemocnění ledvin, Galén, Praha 2011
11. R. Rizzoli, Atlas of Postmenopausal Osteoporosis, 2nd Edition, Current Medicine Group, London 2005
12. C.J. Rosen, Primer on Bone Metabolic Diseases and Disorders of Mineral Metabolism 7th Ed, ASBMR, Washington, 2008
13. V. Vyskočil, Osteoporóza a ostatní nejčastější metabolická onemocnění skeletu, Galén, Praha , 2009
14. L. de Groot, J.L.Jameson, Endocrinology 4th Edition, Vol2W.B Sounders, 2001
15. O. Růžičková, Ovlivnění signální cesty RANK/RANKL/OPG , Osteologický bulletin 17, 2012/1, 21 - 35
16. M.Cirmanová, L. Stárka, Sklerostin- nový regulační marker kostního obratu a klíčový cíl v terapii osteoporózy, Osteologický bulletin 16, 2011/1, 16-19
17. http://www.osteoscoop.com

23.3. Bone Metabolism

Author: MUDr. Richard Pikner, Ph.D.

Reviewers: RNDr. Zdeněk Švagera, Ph.D., MUDr. Věra Ploticová

23.3.1. Options for Laboratory Detection of Bone Remodelling – Bone Markers

Laboratory methods offer many possibilities of testing the state of bone remodelling. In addition to the **genetic analysis** which helps confirm some congenital diseases such as *osteogenesis imperfecta*, Ehlers-Danlos syndrome and other diseases, such methods primarily include tests for molecules circulating in the plasma (serum). Their concentrations are subsequently proportional to bone cell activity. Bone remodelling tests are usually performed together with tests for calcium-phosphate metabolism parameters (PTH, 25 OH vitamin D, calcitonin) which have a major impact on the rate of bone remodelling. The body prefers to maintain normocalcaemia and normophosphataemia to maintaining bone integrity. The differential diagnosis also employs tests for systemic hormones (thyroid hormones, cortisol, oestrogens, androgens, IGF-1, etc.).

Commonly used or research methods are used to assess bone remodelling. Research methods include determining local regulation molecule concentrations (RANKL, OPG, sclerostin). Commonly used bone remodelling parameters are divided into osteoresorption markers or new bone formation markers, and are discussed in this chapter.

23.3.2. Nomenclature for Routinely Used Bone Markers (Abbreviations)

- New bone formation markers:
  - → Osteocalcin (OC)
  - → Bone alkaline phosphatase isoenzyme (bone ALP or bALP)
  - → Procollagen type I N-terminal **propeptide** (PINP)
- Bone resorption markers:

Collagen type I telopeptides (products of type I collagen degradation in the bone)

- → N-terminal cross-linking telopeptide of type I collagen (NTX-I)
- → C-terminal cross-linking telopeptide of type I collagen (CTX-I b, Crosslaps)
  - Type I collagen telopeptide (ICTP)
  - Tartrate-resistant acid phosphatase 5b (TRACP 5b)

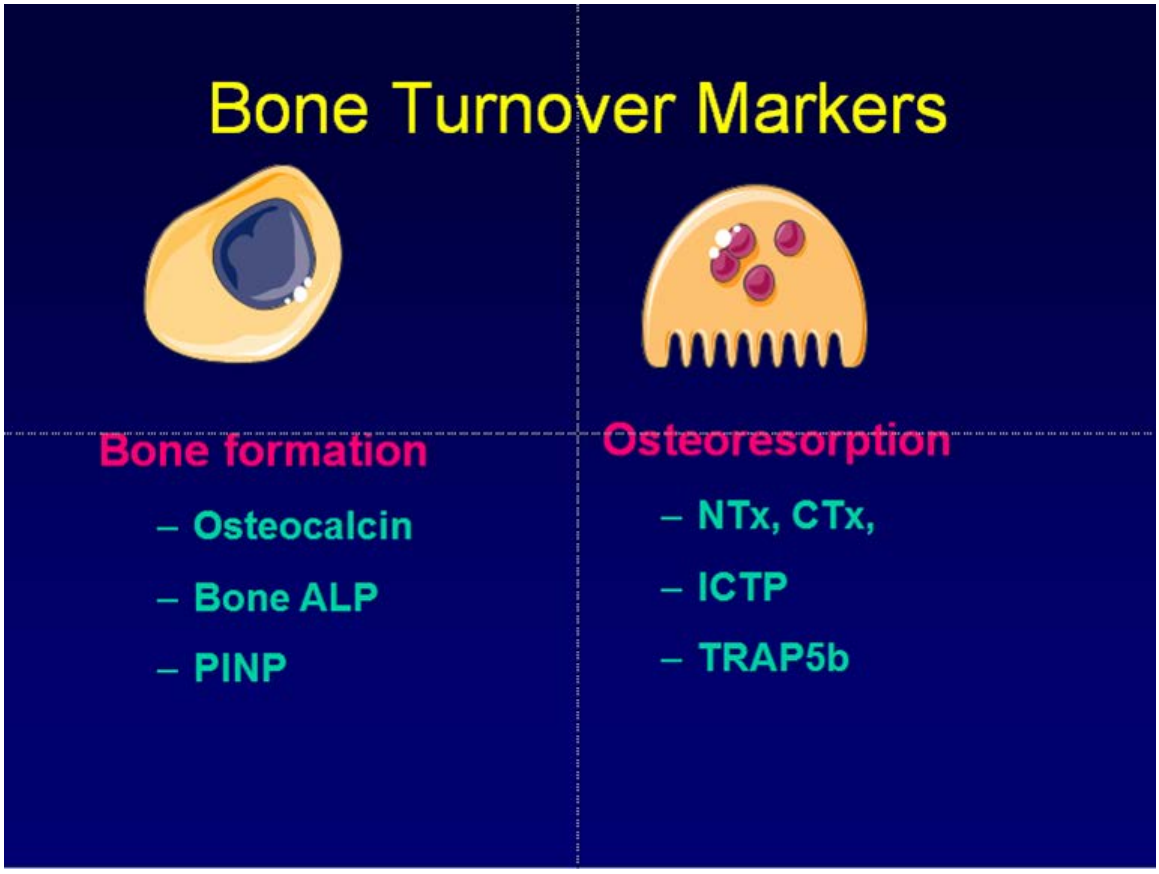


Figure 23.9. Bone Turnover Markers – Graphical summary

23.3.2.1. NEW BONE FORMATION MARKERS

Osteocalcin

Osteocalcin is the most abundant non-collagen protein in the bone matrix (osteoid). It is synthesized in osteoblasts and its subsequent carboxylation is dependent on vitamin K. Osteocalcin deposited in the bone tissue has a high affinity to calcium and contributes to bone matrix mineralisation. Serum osteocalcin levels are proportional to its production by osteoblasts. The latest research has revealed information about osteocalcin as a systemic hormone released from the bones, which contributes to regulating the leptin-adiponectin system and androgen synthesis.

Bone alkaline phosphatase isoenzyme (bone ALP, ostase)

Alkaline phosphatase comprises 4 isoenzymes: placental, carcinoplacental, intestinal, and the hepatic/renal/bone group. These isoenzymes differ in their primary structure and are coded by different genes. In an adult the bone isoenzyme produces less than a half of the total ALP activity (most of it is produced by the hepatic isoenzyme). The bone isoenzyme is a product of osteoblasts, and its quantitative test is a function of their activity.

N-terminal type I procollagen (PINP)

It is released to circulation during type I collagen production. Osteoblasts produce procollagen, which releases the N and C-terminal ends (PINP, PICP) by extracellular proteolysis. The circulating N and C-terminal ends are subsequently quantified. The PINP or PICP values provide information about the level of type I collagen production not only in the bones but also in all tissues producing collagen I (skin, connective tissue, blood vessels). Cartilage contains type II collagen.

23.3.2.2. BONE OSTEORESORPTION MARKERS

N-terminal cross-linking telopeptide of type I collagen (NTX-I), C-terminal cross-linking telopeptide of type I collagen (CTX-I)

Type I collagen is degraded by the action of proteolytic enzymes osteoclasts. The main proteolytic enzyme is cathepsin K, which cleaves collagen fibre into various fragments. The fragments cleave off from both the C and N-terminal ends of the type I collagen molecule. In the circulation they are detected as the N-terminal cross-linking telopeptide of type I collagen (NTX-I) or C-terminal cross-linking telopeptide of type I collagen (CTX-I). They are finally excreted in the urine (accumulation in patients with chronic kidney failure). Their concentrations then indicate the degree of type I collagen degradation.

Type I collagen telopeptide (ICTP)

Another kind of type I collagen degradation prevails in some pathological situations – degradation by matrix metalloproteinases (MMPs), MMP-9 in this case. MMP-9 is produced mainly by tumour cells of bone metastases and causes local pathological degradation of type I collagen, and a fragment other than CTX or NTX is formed, type I collagen telopeptide (ICTP). ICTP concentrations are an indicator of the degree of pathological bone degradation by MMP-9. In the serum, ICTP is subsequently degraded to smaller fragments excreted through the kidneys; ICTP accumulates in chronic kidney failure.

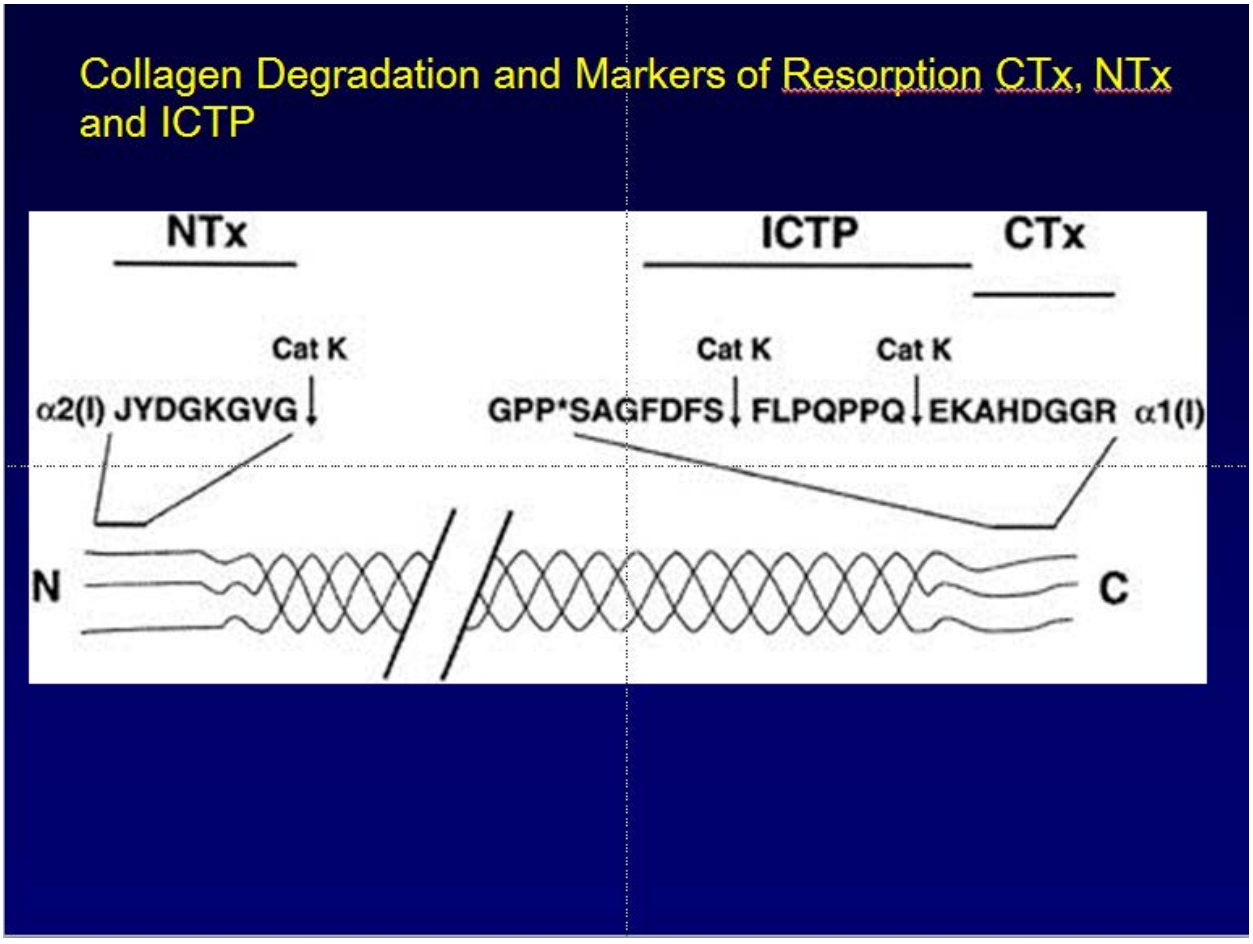


Figure 23.10. Collagen degradation and Markers of Resorption CTx, NTx, ICTP

Tartrate-resistant acid phosphatase 5b isoenzyme (TRACP 5b)

It is one of the 6 ACP isoenzymes. It is abundant in osteoclasts. Quantitative tests of serum TRACP 5b indicates the degree of osteoclast activity.

23.3.3. Indications for Bone Marker Tests

The aim is to assess the rate of bone remodelling. It should be kept in mind, however, that the plasma or serum parameters tested provide a picture of the overall activity of all bone remodelling units in the whole body. They are in various phases of bone remodelling. This results in an average activity of all bone remodelling units. It is thus logical for minor local processes (solitary metastasis) not to reflect in the values until their activity has significantly exceeded the average activity of other remodelling units (BMUs).

A big advantage of bone markers is that sampling is minimally stressful for patients as they only give a normal blood sample.

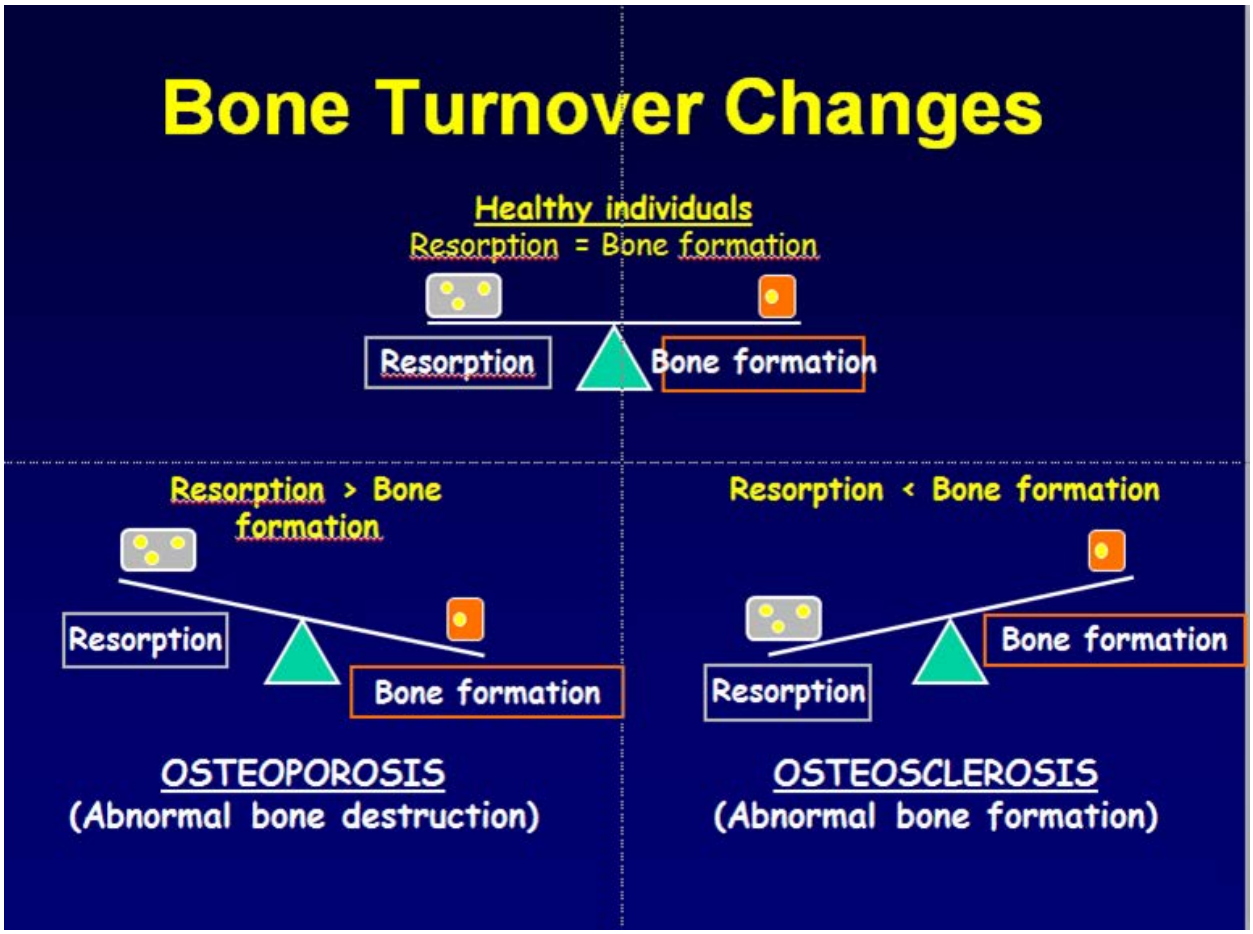


Figure 23.11. Bone Turnover Changes – Balance between the bone resoption and formation

Bone markers are used for:

1. Monitoring the effect of metabolic bone disease therapy
2. Predicting the risk of fractures and bone metastases
3. Differential diagnosis
4. Personalized therapy???

23.3.4. Monitoring the effect of metabolic bone disease therapy

This is probably the most common indication for bone marker tests in post-menopausal osteoporosis. Bone markers practically react immediately to the therapy applied, while the imaging methods (densitometry) are able to detect a change after about 12 months. The effect on bone markers can be expected depending on the therapy selected:

- Bisphosphonates (alendronate, ibandronate, risedronate, zolendronate, pamidronate) reduce bone remodelling – both resorption and new bone formation, and support secondary mineralisation.
- SERMs (selective oestrogen-receptor modulators – raloxifene, tamoxifene) induce the pre-menopausal state of bone remodelling. They are useful for treating early post-menopausal osteoporosis with increased osteoresorption, which is normalised by SERMs; SERMs also reduce the risk of hormone-positive breast cancer.

- Calcitonin considerably inhibits osteoporosis for not longer than about 8 – 12 hours; 1/3 of patients are non-responders, whose bone marker levels do not respond to the administering of calcitonin.
- Intermittent administration of parathormone (once daily, s.c.) – intermittent PTH administration leads to a different profile of RANKL: OPG production than the profile typical of a continuously elevated PTH level. Both new bone formation and osteoresorption are stimulated. New bone formation is, however stimulated more (a higher percentage of change in the markers as compared with their level before treatment). The treatment has an anabolic effect on the bone.
- Denosumab is a subcutaneously administered human antibody with the same effect as OPG, leading to considerable suppression of bone resorption followed by a decrease in new bone formation. It has a similar effect on bone remodelling as bisphosphonates.
- Strontium ranelate is a drug, which presents another option for treating osteoporosis, however administration of strontium does not markedly change bone remodelling. Its effect probably uses another mechanism through the calcium-sensing receptor.
- The human antibody against sclerostin is only in the phase of clinical testing: New bone formation prevailing over osteoresorption, i.e. a considerable anabolic effect on the bone, can be expected.

Similarly, changes in bone remodelling can be monitored through successful therapy of other metabolic skeleton diseases (osteomalacia, Paget's disease, bone metastases, etc.).

In some cases it is helpful to assess the degree of patient's non-compliance, the site of the expected decrease/drop, or bone remodelling parameters changed/unchanged against expected values.

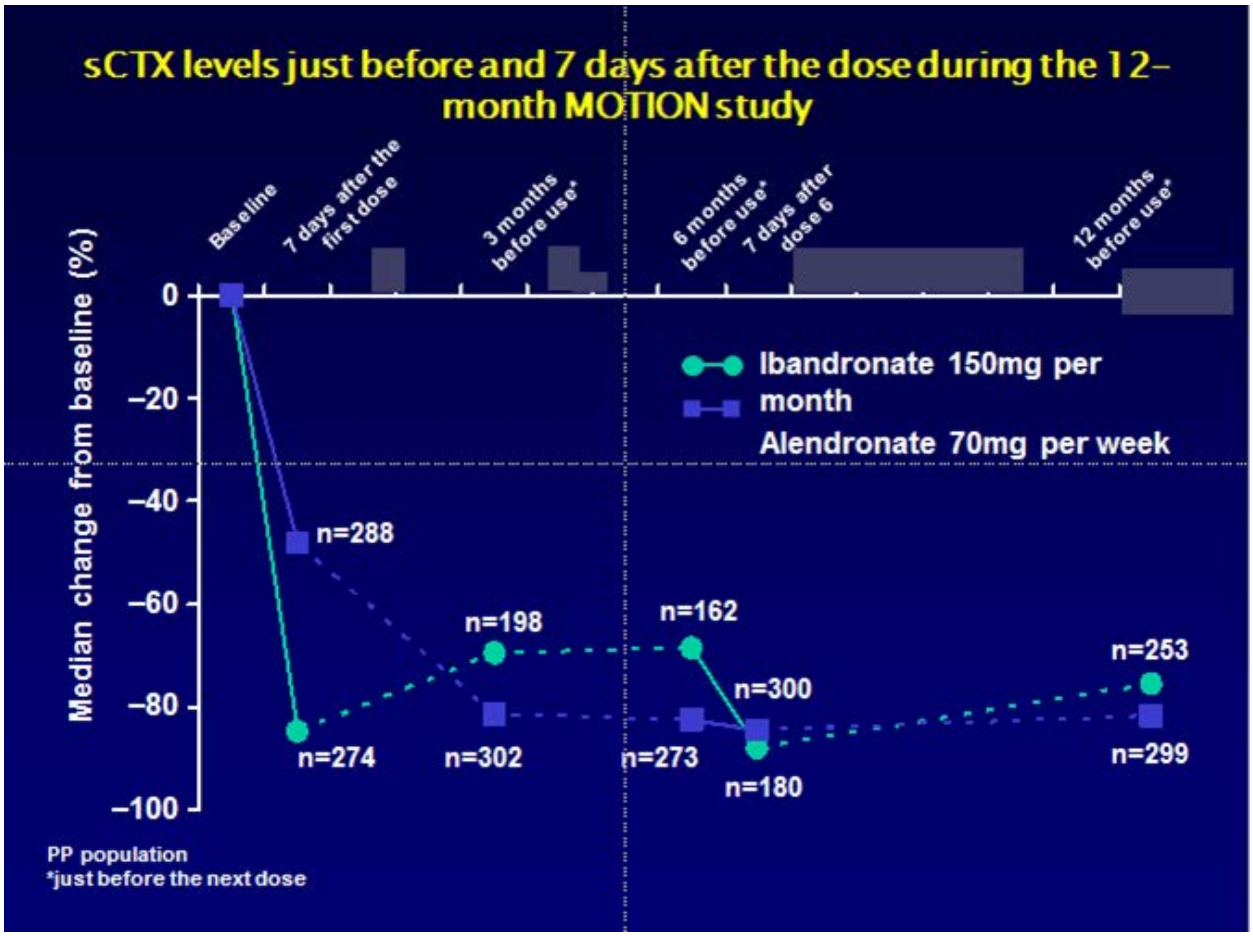


Figure 23.12. CTX levels in serum before and 7 days after the dose during the 12 – months MOTION study

23.3.5. Predictive Value of Bone Markers

The determination of bone remodelling markers is used to estimate the risk of osteoporotic fracture or bone metastases. The evidence shows that the higher bone markers, the higher risk of complications (fracture, metastasis, progression or relapse). Breast cancer patients show a negative correlation of bone resorption markers and survival time (the higher osteoresorption, the shorter time of survival).

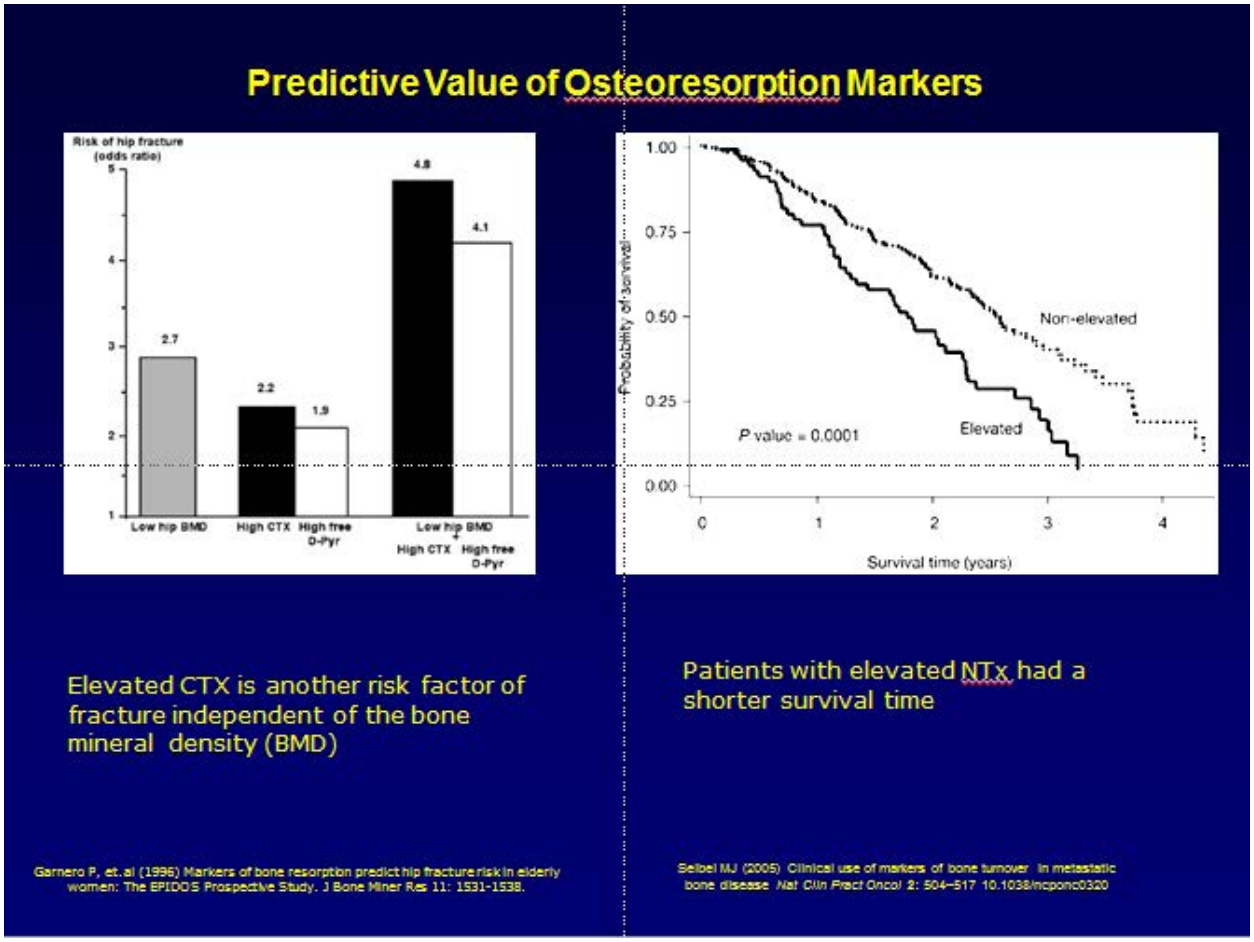


Figure 23.13. Predictive value of osteoresorption markers

23.3.6. Bone Markers in Differential Diagnosis

Bone remodelling markers are not very useful for differential diagnosis. This is determined by their non-specific character. Metabolic skeleton diseases affect bone remodelling, yet the inciting causes are different and changes are only quantitative.

Osteomalacia, for example, causes increased bone remodelling very similar to multiple bone metastases of a mixed character, as well as primary hyperparathyroidism and the condition after healing fractures.

Increased bone resorption alone cannot be interpreted as post-menopausal osteoporosis as it can have several causes such as osteolytic metastases or primary hyperparathyroidism.

However, some laboratory findings are quite typical and may speed up the diagnostic process.

- A sporadically elevated bone ALP value with hyper or normocalcaemia is typical of Paget's disease
- Elevated ICTP values are suspected to be caused by osteolytic or mixed metastases or osteomalacia
- Elevated CTX (NTX) and decreased PINP are typical of corticoid treatment.

23.3.7. Personalized Therapy

The issue of personalized therapy is very questionable. The main method of checking the effect of osteoporosis treatment at present is treatment based on evidence-based, multicenter, prospective, placebo-controlled studies. There is no space for personalized therapy here, though. Most of these studies are performed in patients with normal or slightly deficient vitamin D level. Is it therefore possible to start treatment in patients with severe vitamin D deficiency? Severe vitamin D deficiency is found in around 30 – 40% of the normal population with osteoporosis.

The point of personalisation is finding an optimum therapy model for a specific patient on the basis of known medical history, and an analysis of risk factors and laboratory tests, of course.

Personalized therapy is used for Paget's disease where the ALP level is decisive. Therapeutic bisphosphonates (zolendronate) are administered to normalize the ALP. An additional zolendronate dose is pointless until the level increases.

Another example is individual dosing of vitamin D based on initial and target vitamin D levels, body weight %, sunbathing, etc.

23.3.8. Test Methods

Bone remodelling parameters are measured almost exclusively using immunoassays. Except for TRAP5b and ICTP, most methods are currently available on fully automated instruments. TRAP5b and ICTP are measured using enzyme immunoassay on plates, or using radioisotope detection kits.

When measuring osteocalcin, it is important to use methods testing the intact form and N-MID fragment to ensure a lower variability of results.

23.3.8.1. Pre-Analytical Phase – Sample Stability, Diurnal Rhythms, Diet, Method Selection

Bone markers show significant variability depending on the age, gender, sampling time, food consumed and use of some drugs (corticoids), and also depending on the test method used. Bone markers should therefore be tested in one laboratory, if possible.

Changes can be in the order of tens to hundreds of %.

This means it is essential to take samples under standardized conditions and follow age, gender and race-specific reference limits.

TRAP5B is the least affected method of all methods (diurnal rhythm, diet, renal functions).

Standard sampling is in the morning between 08:00 and 11:00, fasting patient.

Sources of biological variability of bone markers

Uncontrollable sources:

Age – bone remodelling is age-dependent primarily in children. Bone remodelling should be assessed in correlation with the individual patient's growth phases (growth curve); levels are more or less stable in middle age; the same holds true for bone mass and mineralisation. Old-age changes (first a rise in osteoresorption and also reduced bone formation from 70 years) are not physiological and are very individual.

Diurnal rhythm – PTH is the main systemic hormone regulating bone metabolism. The PTH has a biphasic circadian rhythm with a slight increase between 4:00 and 7:00 p.m., peak values between 2:00 and 6:00 a.m. and lowest levels between 10:00 and 11:00 a.m. The bone remodelling markers follow the same diurnal rhythm.

Gender – bone remodelling values vary with the gender as they reflect different oestrogen and androgen levels.

Race – bone remodelling parameters are race-dependent.

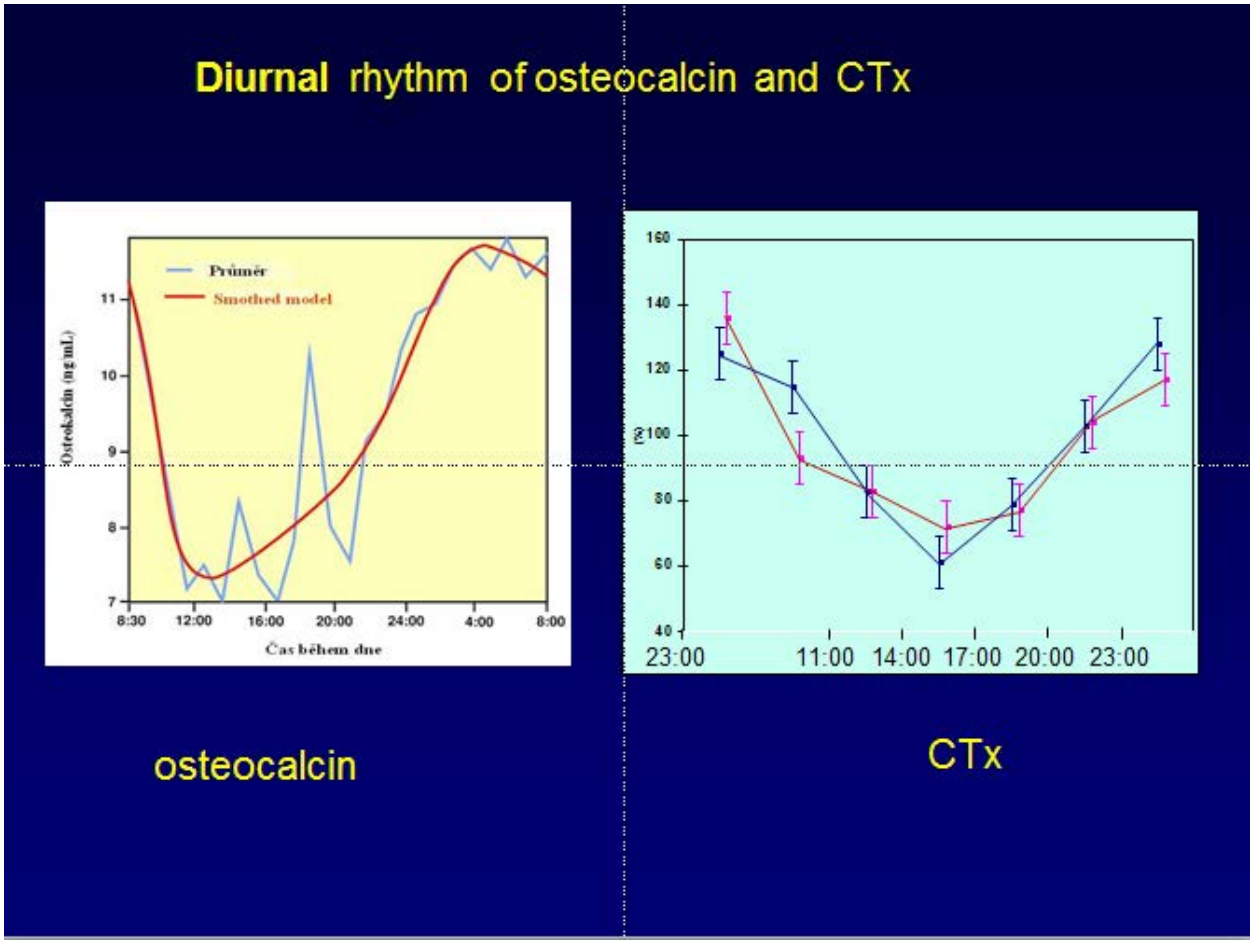


Figure 23.14. Diurnal rhythm of osteocalcin and CTx

Controllable sources

Physical strain – exertion impacts bone remodelling regulation by suppressing sclerostin production. Physical strain has to be prolonged and repeated; an extreme one-off load does not significantly affect the values.

Diet – ingestion of food leads to a dramatic drop in bone resorption and a slight increase in bone formation. Fasting samples are recommended.

The use of chondroprotective agents containing type I collagen affects collagen degradation marker CTX and NTX levels.

23.3.8.2. Sample Stability following Collection:

Bone markers such as osteocalcin, CTx, PINP, ICTP, bone ALP and TRAP5b are measured in the serum. If the serum is separated from blood elements within 8 hours from sampling, they are stable at 4 – 8 °C for at least 72 hours. Repeated freezing and thawing is not good for osteocalcin, but is possible for the other parameters.

23.3.8.3. Interpretation of Results

Reference Limits

Reference limits depend on the age, race and gender.

They are method-specific, so are not listed.

23.3.9. Specific Metabolic Bone Diseases

23.3.9.1. Bone mineralisation and age

Both sexes reach peak bone mass levels around the 2nd or 3rd decade of life. In the following period until 50 years of age, the bone mass is relatively constant on the provision of the „physiological state“: normal calcium intake, adequate physical activity, normal reproductive state and absence of diseases or influences causing secondary osteoporosis. Old age brings hormonal changes affecting bone mineralisation. In addition the prevalence of secondary hyperparathyroidism due to deficient vitamin D and calcium increases. This is followed by involutional changes such as decreases to physical activity, reduced hormone production etc. A disease called osteoporosis occurs.

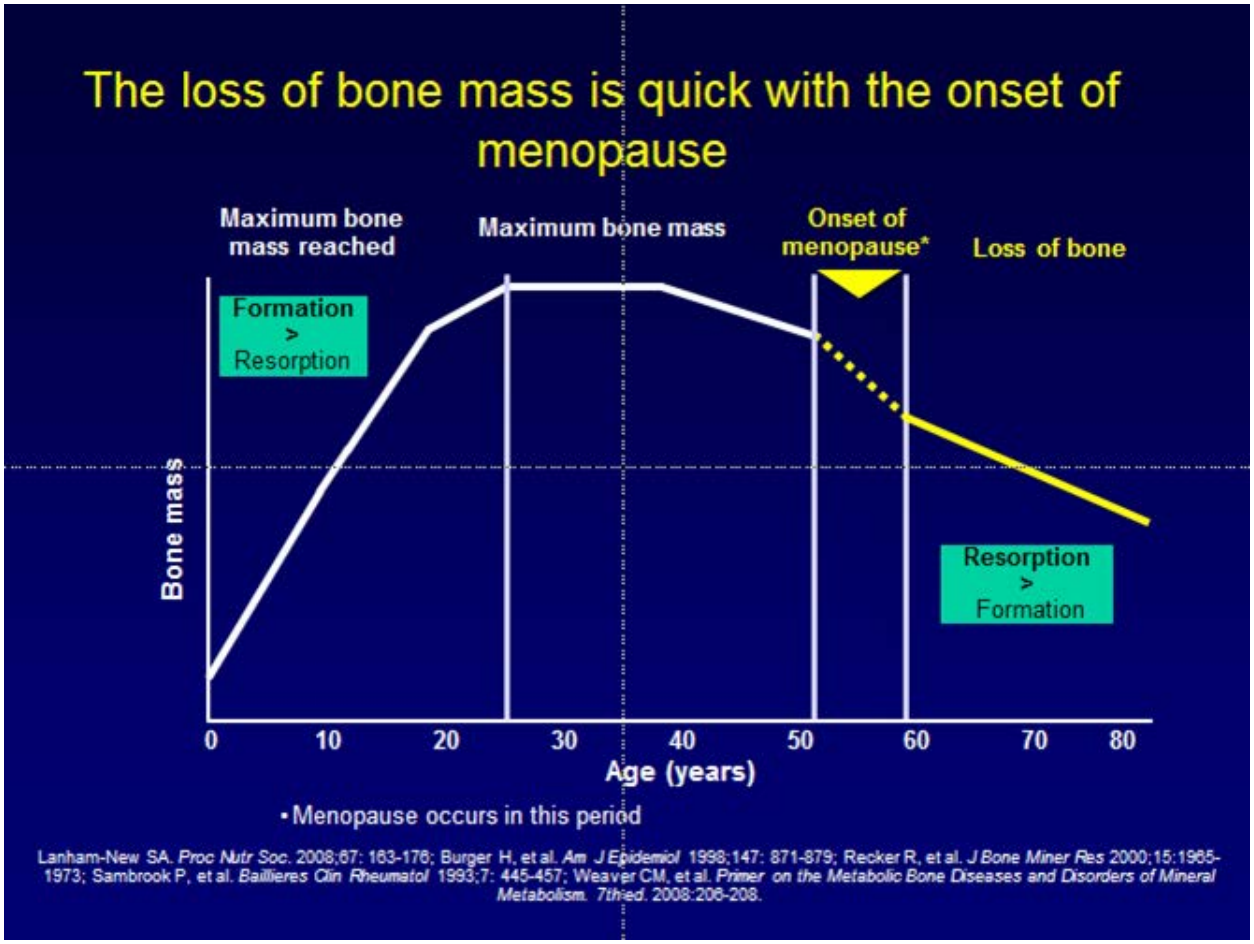


Figure 23.15. The loss of bone mas sis quick with the onset of menopause

WHO defines **osteoporosis** as a systemic bone disease characterised by a decrease in bone mass and damage to the bone tissue microarchitecture, with a subsequent increase in bone fragility and proneness to fractures. The loss of minerals and the osteoid is proportional.

Bone mineral density (BMD) is most commonly measured by dual-energy X-ray absorptiometry (DXA). Measured bone mineral density is then compared with the average value found in young, healthy individuals of the same sex (T--score) and with the average value in the same age category (Z-score).

Osteoporosis in post-menopausal women and men over 50 is defined as a T-score less than 2.5 SD (standard deviation minus average density).

Reduced bone mass corresponds to the T-score -2.4 to -1.0 SD, and normal bone density is defined as the T-score between -1.0 and +1.0 SD.

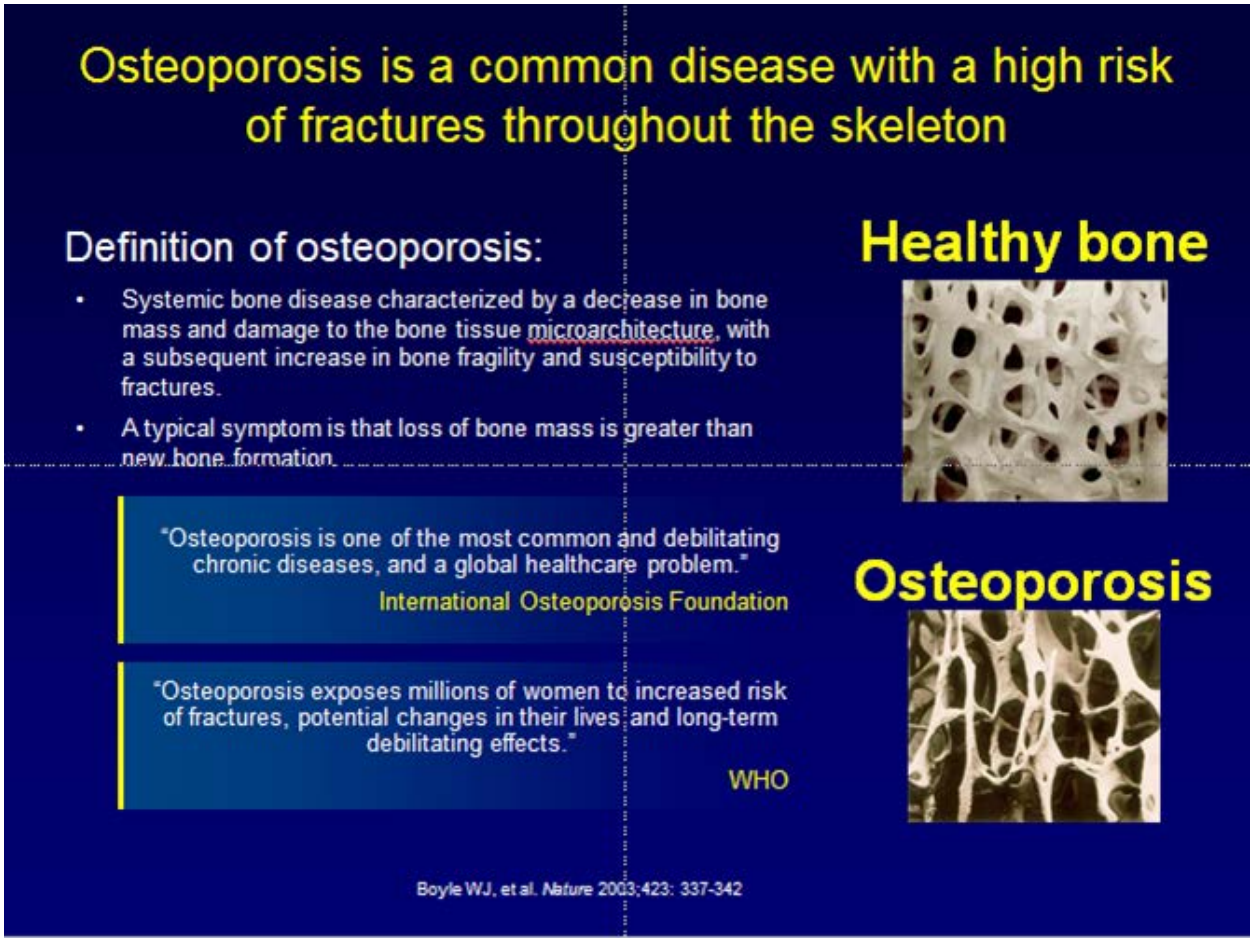


Figure 23.16. Osteoporosis is a common disease with a high risk of fractures throughout the skeleton

Post-menopausal osteoporosis due to deficient oestrogens

Oestrogen deficiency caused by the menopause leads to reduced secretion of OPG and increased production of cytokines (RANKL, M-CSF, IL6, IL-11). This leads to increased activation of osteoclasts and reduced activity of osteoblasts. Bone remodelling thus increases, osteoresorption in particular. Increased amounts of calcium released from bones leads to reduced secretion of PTH, thus increasing calciuria and reducing the production of 1.25 $(OH)_2$ vitamin D. The total daily loss of calcium rises about threefold. Administration of oestrogens normalizes the pathophysiological mechanism and prevents loss of bone mass. This problem afflicts women around 10 years after the menopause.

Lab: Increased osteoresorption and normal or decreased new bone formation

Involutional (senile) osteoporosis

Multifactorial aetiology: secondary hyperparathyroidism occurs due to vitamin D deficiency and conversion in the kidneys, decreased physical activity, lack of androgens and oestrogens and other anabolic hormones (IFG1); commonly due to poor diet – protein malnutrition, and coinciding drugs and diseases.

Lab: Osteoresorption and new bone formation are decreased in the lower quartile of reference limits

Glucocorticoid-induced osteoporosis (GIOP)

Chronic oral or intravenous administration of glucocorticoids (does not apply to replacement therapy for adrenal insufficiency) leads to a fast induction of osteoporosis. Major losses are in the trabecular bone (most frequent fractures in the axial skeleton – vertebrae); peak loss in the first 12 months of treatment. The loss depends on the dose, drug type, age, gender, BMD and individual sensitivity.

Mechanism of osteoporosis induction:

1. Direct effect

- Reduced new bone formation
- Number and differentiation of osteoblasts and apoptosis of osteoblasts
- Reduced synthesis of osteoid (collagen I)
- Increased bone resorption

2. Indirect effect
   - Induced deficiency of calcium
   - Reduced calcium resorption in the intestine and increased degradation of 1,25 $(OH)_2$ D
   - Reduced renal tubular reabsorption - hypercalciuria
   - Reduced secretion of GH and IGF-1 (anabolic effect)
   - Reduced secretion of sex hormones

Lab: Increased osteoresorption, reduced new bone formation (PINP in particular), secondary hyperparathyroidism

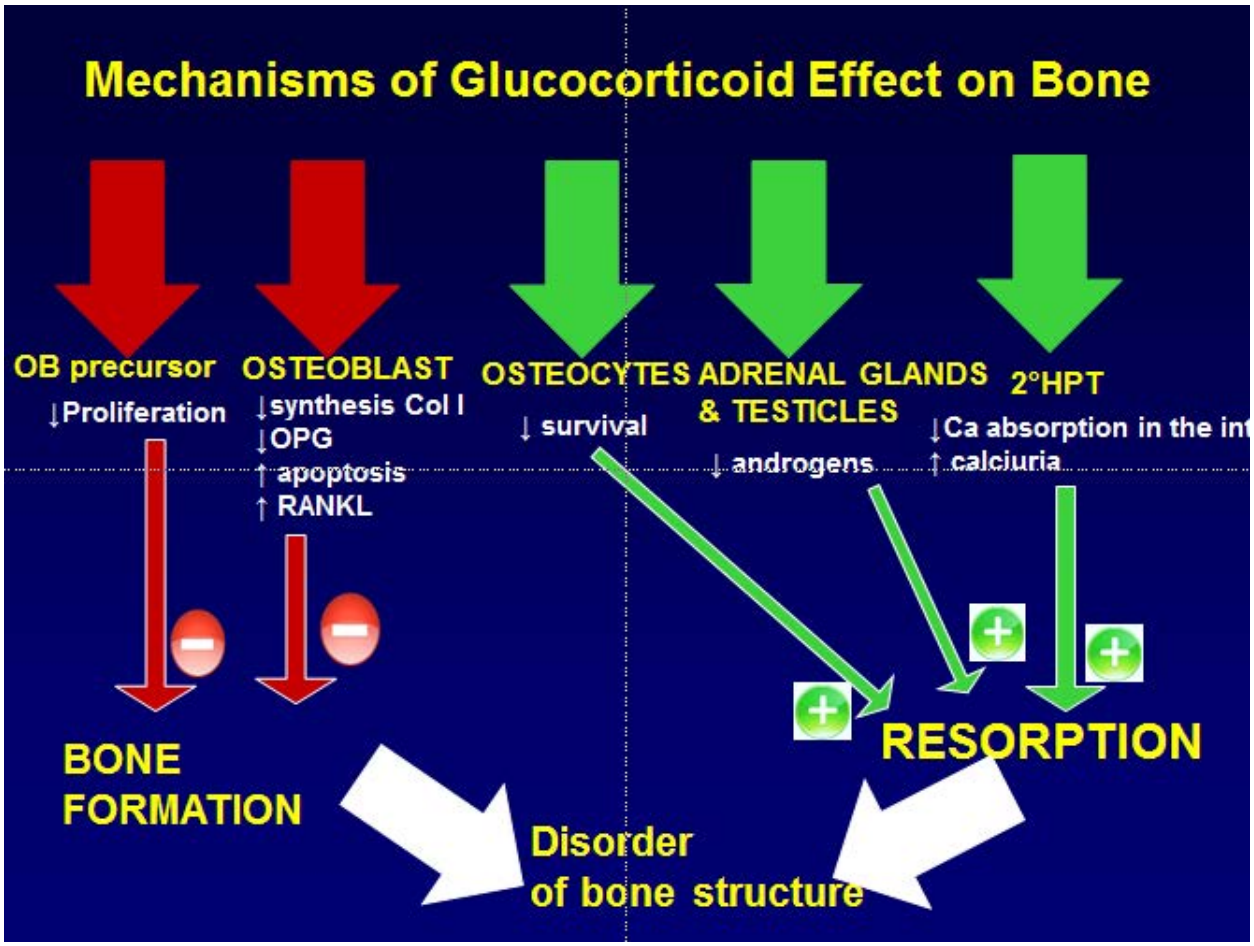


Figure 23.17. Mechanisms of glucocorticoid effect on bone

Thyrotoxicosis-induced osteoporosis

Excess thyroid hormones lead to increased bone turnover, negative calcium balance and loss of trabecular bone.

Lab: Osteoresorption is increased more than new bone formation

Immobility-induced osteoporosis

Increased sclerostin levels lead to increased osteoresorption on negative calcium balance, and new bone formation decreases.

Lab: Osteoresorption is increased much more than new bone formation, hypercalciuria

23.3.9.2. Osteomalacia, rickets (rachitis)

Osteomalacia is also a systemic bone disease, but this one afflicts osteoid mineralisation. This disorder is referred to as **rickets (rachitis - missing mineral and normal osteoid mass)** in children with open epiphyseal cartilages.

The most common cause is deficient vitamin D with subsequent malabsorption of minerals in the small intestine. This fact leads to abnormal mineralisation of the newly formed osteoid, increased formation and insufficient ossification. In addition to the bone, also the epiphyseal cartilage is afflicted in children, which leads to bone deformities.

Causes of osteomalacia

1. Vitamin D deficiency
2. Disorders of vitamin D metabolism (hereditary, anticonvulsant-induced)
3. Malabsorption syndrome
4. Renal acidosis (primary and secondary)
5. Depletion of phosphates

Lab: Highly increased bone remodelling (new bone formation and osteoresorption). In addition, calcium and phosphate metabolism disorder is present depending on the cause (low vitamin D, low or normal calcium, low phosphorus, increased PTH)

23.3.9.3. Other selected common osteopathies

Hyperparathyroidism, hypoparathyroidism see chapter Calcium, Phosphate and Magnesium Metabolism

Chronic kidney disease - mineral and bone disorder (CKD-MBD) – formerly renal osteodystrophy, renal bone disease

It is now defined as a disorder of 1 of the following 3 clinical components:

1. **Laboratory component =** impaired calcium-phosphate metabolism and bone remodelling
2. **Bone component =** histological bone tissue alterations
3. **Calcification component =** presence of off-skeleton calcifications

In its initial phase renal failure leads to increased requirements for phosphate excretion. The production of FGF23 increases, which leads to reduced production of active calcitriol in the kidneys and subsequently to reduced calcium absorption (hypocalcaemia or the lower quartile of standard values). This results in secondary hyperparathyroidism. This finding is typical of the histological presentation of **high-turnover osteopathy** (*osteitis fibrosa*).

On the other hand, around a half-sized group of patients present with **adynamic bone disease**. It is characterized by low PTH levels, low bone remodelling, while histology samples show a cell-deficient bone without signs of remodelling.

The remaining 10% is **mixed-type, high-turnover and low-turnover bone disease**. Osteomalacia due to deficient vitamin D with typically low bone remodelling is very rare today.

Prevalence of CKD-MBD in patients with CKD stages 3-5:

1. 32% ... Osteitis fibrosa
2. 20% … Mixed disease
3. 18% … Adynamic bone disease
4. 16% … Normal bone remodelling
5. 8% … Osteomalacia
6. 6% … Mild disease

Lab:

- **Osteitis fibrosa:** High bone remodelling, 2nd generation PTH often over 300 pg/mL, normal bone mineralisation

- **Adynamic bone disease and osteomalacia:** Low bone remodelling, 2nd generation PTH is decreased or corresponds to healthy values, impaired bone mineralisation
- **Mixed bone disease:** Normal or slightly increased bone remodelling, medium-increased 2nd generation PTH, impaired bone mineralisation

Hypercalcaemia in tumour diseases

Bone metastases are most common in breast (73%), prostate (68%), thyroid (42%), kidney (35%) and lung (36%) cancer. Primary bone marrow impairment is typical of multiple myeloma.

The most common cause of hypercalcaemia in tumour diseases is the overproduction of PTHrP (parathormone-related protein), permanently elevated levels of which, similar to the PTH, lead to osteoresorption and hypercalcaemia.

The formation of metastases is connected with the primary tumour's ability to create and release micrometastases into the blood stream. The bone then provides a suitable environment full of growth factors. Metastasizing tumour cells are able to stimulate osteoclasts or osteoblasts by means of cytokines produced by them. This results in increased osteoresorption leading to osteolytic metastases, or new bone formation leading to osteoplastic metastases. On the other hand, stimulated bone cells release many growth factors, which retroactively support tumour growth (vicious circle – *circulus vitiosus*).

Tumour micrometastases are not always histologically homogeneous. Mixed metastases consisting of osteolytic and osteoplastic foci are common.

Lab: The rate of bone remodelling is directly proportional to the scope of metastases and, of course, to the activity of tumour and bone cells. (Small solitary metastases do not usually affect bone markers.)

Osteolytic metastases – increased osteoresorption of ICTP and CTX in particular. ICTP is more specific for bone metastases as tumour cells more often eliminate bone collagen I by means of MMP-9, which produce ICTP.

Osteoplastic metastases – increased PINP, less bone ALP

Mixed metastases – increased bone remodelling (osteoresorption and new bone formation)

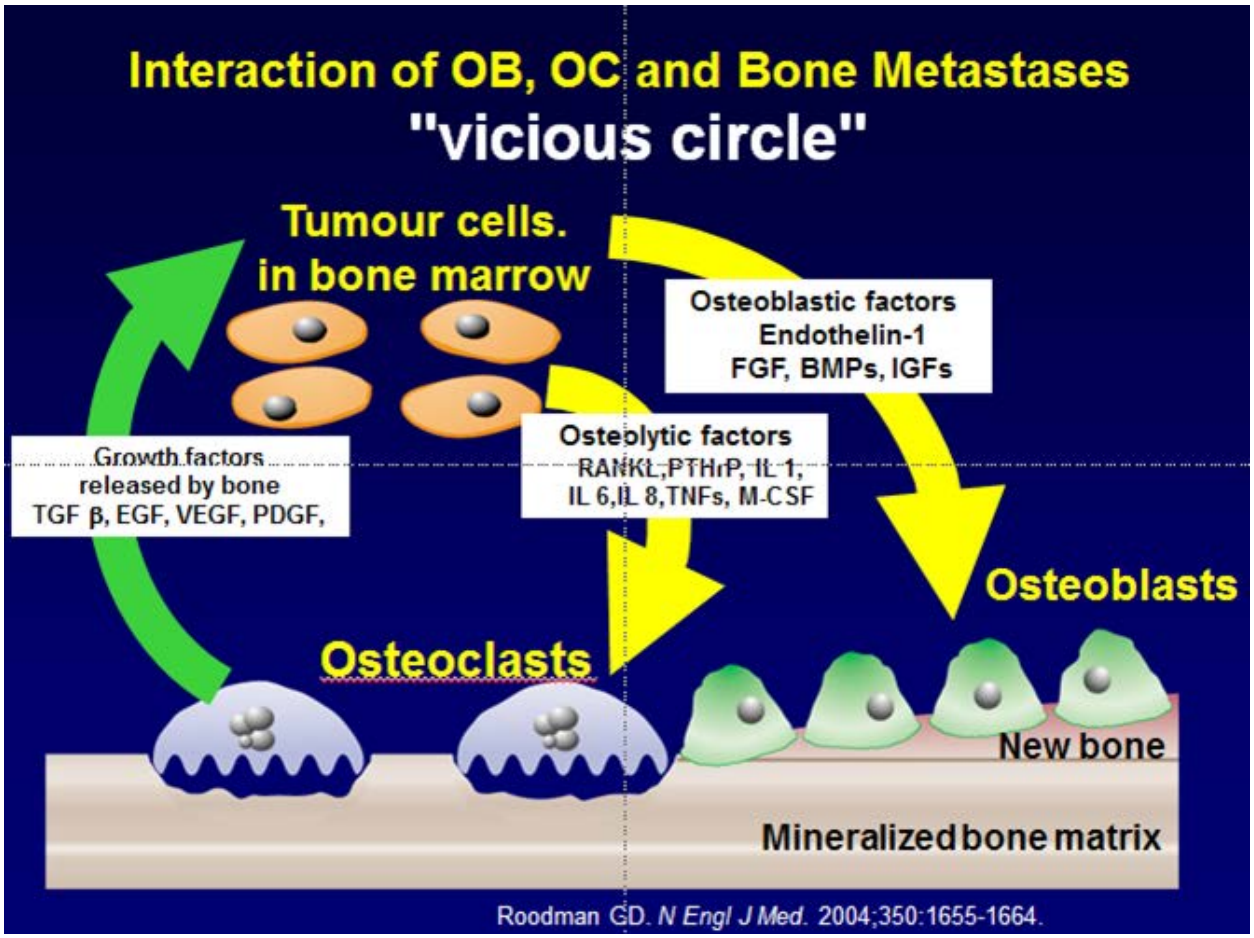


Figure 23.18. Interaction of osteoblasts, osteocytes and bone metastases – „vicious circle"

Paget's disease

Paget's disease is a localised bone disease with impaired bone remodelling regulation. The function of osteoclasts is impaired. Osteoclasts begin to resorb the bone locally and permanently, and a compensatory increase in new bone formation follows. This results in the formation of a disorganised mixture of trabecular and lamellar bone. This causes deformities and increases proneness to pathological fractures.

Lab: Increased bone remodelling, ALP in particular; calcium-phosphate metabolism is usually normal.

23.3.10. Summary – Key Information

- New bone formation markers are proportional to osteoblast (OB) activity: bone ALP, or type I collagen synthesis (PINP).
- Bone resorption markers are proportional to the amount of degraded type I collagen (CTx, NTx, ICTP) or proportional to osteoclast activity (TRACP5b).
- Measured plasma or serum bone remodelling parameters provide a picture of the overall activity of all bone remodelling units (BMUs) in the whole body. BMUs are found in various phases of bone remodelling. The result of the bone marker test is total average activity of all BMUs.
- Bone markers react practically immediately when the effect of therapy is checked, while the imaging methods (densitometry) are able to detect a change after about 12 months. The corresponding effect on bone markers can be expected depending on the therapy selected:
  - → bisphosphonates, denosumab and calcitonin decrease bone remodelling
  - → PTH increases bone remodelling
  - → SERMs (selective oestrogen-receptor modulators) normalize bone remodelling
- Predictive value – bone remodelling markers are used to estimate the risk of osteoporotic fracture or bone metastases. The higher the bone resorption marker value outside the reference range, the higher the risk of accident. The higher the PINP values in patients with prostate cancer, the higher the likelihood of osteoplastic metastases.
- Bone remodelling markers are not very useful for differential diagnosis. It is due to their non-specificity, since metabolic skeleton diseases cause quantitative rather than qualitative changes in bone remodelling.
- Bone remodelling parameters are measured almost exclusively using immunoassays.
- Bone markers show significant variability depending on the age, gender, sampling time, food consumed, the use of some drugs (corticoids), and also depending on the specific immunoassay used. **Changes can be in the order of tens to hundreds of %. Standard sampling is in the morning between 08:00 and 11:00, fasting patient. Following blood cell separation, most parameters are stable at 4–8 °C for 24 hours.**
- **Osteoporosis** is a systemic bone disease characterised by a low bone mass and damage to the bone microarchitecture (the loss of minerals and the osteoid is proportional).
- **Typical laboratory findings – osteoporosis:**
  - *Post-menopausal osteoporosis* due to deficient oestrogens – bone markers: Increased osteoresorption and normal or decreased new bone formation
  - *Involutional (senile) osteoporosis* - Osteoresorption and new bone formation are decreased in the lower quartile of reference limits.
  - *Glucocorticoid-induced osteoporosis* – Bone markers: Increased osteoresorption, reduced new bone formation (PINP in particular), secondary hyperparathyroidism
  - *Thyrotoxicosis-induced osteoporosis* – Bone markers: Osteoresorption is increased more than new bone formation
  - *Immobility-induced osteoporosis* – Bone markers: Osteoresorption is increased much more than new bone formation, hypercalciuria
- Osteomalacia is impaired osteoid mineralisation in adults; this disease is referred to as rickets (*rachitis*) in children with open epiphyseal cartilages. **Osteomalacia is characterised by insufficient mineralisation, however the osteoid amount is normal**.
- *Osteomalacia* – Highly increased bone remodelling (new bone formation and osteoresorption). In addition, calcium and phosphate metabolism disorders are present depending on the cause (low vitamin D, low or normal calcium, low phosphorus, increased PTH).

- **Chronic kidney disease-mineral and bone disorder (CKD-MBD) – formerly renal osteodystrophy, renal bone disease**:
  - *Osteitis fibrosa*: High bone remodelling, 2nd generation PTH often over 300 pg/mL, normal bone mineralisation
  - *Adynamic bone disease and osteomalacia*: Low bone remodelling, 2nd generation PTH is decreased or corresponds to healthy values, impaired bone mineralisation

Mixed bone disease: Normal or slightly increased bone remodelling, medium-increased 2nd generation PTH, impaired bone mineralisation

- **Hypercalcaemia in tumour diseases** – Bone markers: The rate of bone remodelling is directly proportional to the scope of metastases and, of course, to the activity of tumour and bone cells.
  - Hypercalcaemia is caused by PTHrP (PTH-related protein) overproduction
  - Osteolytic metastases – increased osteoresorption of ICTP and CTX in particular. ICTP is more specific for bone metastases than CTx.
  - Osteoplastic metastases – increased PINP, less bone ALP

Mixed metastases – increased bone remodelling – osteoresorption and new bone formation

- **Paget's disease** – Bone markers: Increased bone remodelling, ALP in particular; calcium-phosphate metabolism is usually normal.

23.4. References

1. RE.Coleman et al, *Handbook of Cancer Related Bone Disease*, BioScientifica, Bristol, 2010, UK
2. R.Eastell: Bone Markers , Martin Dunitz Ltd, London 2001
3. I.Sotorník, Š. Kutílek a kol.; *Kostní minerály a skelet při chronickém onemocnění ledvin*, Galén, Praha 2011
4. R. Rizzoli, *Atlas of Postmenopausal Osteoporosis*, 2nd Edition, Current Medicine Group, London 2005
5. C.J. Rosen, *Primer on Bone Metabolic Diseases and Disorders of Mineral Metabolism 7th Ed*, ASBMR, Washington, 2008
6. V. Vyskočil, *Osteoporóza a ostatní nejčastější metabolická onemocnění skeletu*, Galén, Praha, 2009
7. L. de Groot, J. L. Jameson, *Endocrinology 4th Edition*, Vol2W.B Sounders, 2001
8. S. Dusilová Sulková, *Diagnostika, prevence a léčba kostní a minerální poruchy při chronickém onemocnění ledvin (CKD-MBD):KDIGO doporučení a navazující iniciativy národních a nadnárodních pracovních skupin*, Osteologický bulletin 17, 2012/1, 14-20
9. M.Luchavová, V. Zikán: *Cirkadiánní rytmus PTH a jeho role v kostním metabolizmu*, Osteologický bulletin 15, 2010/3, 95-101
10. KDIGO Clinical Practice Guidelines for the Diagnosis, Evaluation, Prevention, Treatment of Chronic Kidney Disease –Mineral and Bone Disorder (CKD-MBD), Kidney International vol 76, supl 113, 2009
11. http://www.osteoscoop.com

CHAPTER 24

24. Laboratory Diagnostics in Gastroenterology

Authors: MUDr. Kocna Petr CSc.

Reviewer: MUDr. Milan Dastych

Laboratory diagnostics in gastroenterology includes targeted diagnostic methods, specific screening programmes and non-invasive function test programmes employing modern methods of analyte detection in faeces or using breath tests.

24.1. Screening programmes

Screening programmes focus on early diagnosis of diseases which would otherwise remain unrecognized at this early stage. Screening programmes are carried out in asymptomatic individuals (i.e. individuals without signs or symptoms of the disease). Defined high-risk groups are subject to preventive medical care programmes. Gastroenterology screening currently includes two basic programmes: screening for celiac disease and screening for colorectal tumours.

24.2. Function tests

They form an important part of the clinical and diagnostic process in gastroenterology. They supplement imaging examination results with an essential piece of information, namely, how the organ is functioning, i.e. its ability to react to stimulation. Function tests include precisely defined stimulation and the result is interpreted as the organ's response to such stimulation, taking baseline analyte values prior to stimulation into account. In addition, the interpretation of indirect function test results must also consider the function of other organs or systems that take part in the process.

24.2.1. Modern, non-invasive function diagnostics

Modern diagnostics focuses on breath tests based on measuring hydrogen or carbon ^{13}C concentrations in exhaled air. H_2-breath test applications cover a wide area, from differential diagnostics of malabsorption syndrome, syndrome of small intestinal bacterial overgrowth (SIBO) and gastrointestinal motility testing, to oral-caecal transit time (OCTT) or the quality of colon preparation before endoscopy. The time of transit through the gastrointestinal tract is an important piece of information for evaluating and interpreting other function tests, and it is therefore often combined with additional breath tests such as the $^{13}C/H_2$ - lactose test, where enzymatic cleavage of lactose is evaluated (the marker is carbon ^{13}C), and at the same time bacterial cleavage in the colon is used to correct motility and transit (the marker is H_2). The reliability and evaluation of H_2 breath tests are increased not only by hydrogen assays, but also combinations with methane assays, enabled by GC (Gas Chromatography) analyzers and NDIRS (Nondispersive Isotope-Selective Infrared Spectroscopy) technologies.

24.3. Laboratory diagnostics of gastric pathologies

This domain includes serological tests of the gastric mucosa: gastrin-17 level, pepsinogens I and II (pepsinogens A and C) ratio, Helicobacter pylori infection, Hp-antibodies, CagA and VacA antigens, the gastric acidity (HCl) function test and the gastric motility breath test (OABT).

24.3.1. The diagnosis of Helicobacter pylori infection

The diagnosis of HP includes invasive tests, requiring the collection of gastric or duodenal mucosa biopsy specimens, or non-invasive tests. The culture test has the highest sensitivity and specificity. However, the considerable sensitivity of the Helicobacter pylori bacterium to oxygen is a drawback since the test requires special conditions for

collection and transport. The rapid urease test (CLO) is based on the chromogenic detection of urease activity (the surface marker of Helicobacter pylori), and is a routine test performed at the time of endoscopy. More recent tests determine Helicobacter pylori in biopsy specimens by immunological detection – the iRUT method. Molecular biology and PCR methods are facilitating other modern methods of detecting Helicobacter pylori in biopsies and stool samples. The breath test with ^{13}C - labelled urea (UBT) is the gold standard of Helicobacter pylori infection diagnostics; the test can also be performed by endoscopy with the administration of 20 mg labelled urea and the collection of gastric gas content using the bioptic channel of the endoscope. A reliable variant of Hp detection is non-invasive Hp antigen detection in the stool.

24.3.2. Gastric acidity tests

Gastric acidity tests are based on the stimulation of parietal cells, withdrawal of gastric juice and the determination of free and total HCl. Pentagastrin is optimal for stimulation, but also histamine (Lambling test) and insulin can be used. Functional test evaluation depends on the type of stimulation, which means that results vary depending on gastrin, histamine or insulin administration. Hypochlorhydria (hypoacidity) or even achlorhydria (anacidity) is a significant symptom of pernicious anaemia or suspected malignancies (gastric carcinoma at the early stages exhibits hyperacidity or normal acidity, however). The gastric acidity test is important for diagnosing Zollinger-Ellison syndrome, where high basal and maximum secretion is determined (BAO > 15, MAO > 60); a BAO/MAO index > 0.60 is found in more than 50% of Zollinger-Ellison syndrome cases. Pentagastrin stimulation in the test can also be used for mucin analysis. Latest studies use the bQRT test (Blood Quininium Resin Test) to detect hypochlorhydria. An endoscopic variant of the function test is stimulation by 4 µg tetragastrin/kg body weight subcutaneously with a 10 - minute secretion collection (20 - 30 minutes after stimulation), and theacidity in mEq/10 minutes is measured by titration. The correlation with MAO-BAO is r = 0.92 and the reproducibility of the test is CV = 5.6%. A breath test with labelled calcium carbonate was the latest development in 2009.

24.3.3. Serum gastrin level

Gastrin level determined by RIA or ELISA immunochemical detection is 25 - 100 ng/L. A 10 to 1000 - fold increase in gastrin level is detected in Zollinger - Ellison syndrome (gastrinoma, pancreatic tumour with gastrin overproduction); however, this level fluctuates widely even during the day; normal gastrin level can be found in 20 - 40% of cases. As there are three forms of gastrin, the test result depends on the type of antibody used in the test. Gastrin test methods have been standardized to synthetic gastrin G-17; G-34 and G-13 tests depend on cross-reactivity with the relevant antibody. The normal proportion of the forms G-13:G-17:G-34 is 8:2:1; G-34 is higher in a fasted state, and G-17 and G-13 are higher after meals. The gastrin test is part of the gastrin stimulation test, where a 90 - minute profile is determined (in 9 serum specimens) following insulin, secretin or Ca-gluconate stimulation.

24.3.4. Pepsinogen A

Pepsinogen A is a mucosal atrophy marker and is used in genetic studies as a subclinical marker of duodenal ulcer disease. Pepsinogen C is used as a marker for the state of gastric mucosa (or the PG-A/PG-C ratio), and also as a marker of Helicobacter pylori infection eradication. Electrophoresis permits the separation of 8 gastric mucosa proteases in agar gel: pepsinogens PG1 - PG5 form a group of immunologically identical proteins – pepsinogen I (PG-I, PGA), pepsinogens PG6 and PG7 form a group of pepsinogen II (PG-II, PGC), and the last protein is cathepsin E (SMP, Slow Moving Proteinase). A decrease in the pepsinogen A level is tested in patients with achlorhydria, for example in pernicious anaemia. The latest studies have proven a significant decrease in pepsinogen-I with a concurrent increase in IgA antibodies against Helicobacter pylori in gastric carcinoma. The detection of pepsin in saliva/sputum is also indicated as a simple screening for gastroesophageal reflux disease (GERD) with extra-gastric signs of the disease. The PG-I/PG-II ratio decreases significantly depending on the histological risk or the presence of a vacA+ positive Helicobacter pylori infection. The combination of pepsinogen-I/II, gastrin-17 and Helicobacter pylori antibody tests is used as a „serological biopsy", referred to as the GastroPanel, in the differential diagnosis of gastritis. Screening testing of the risk of atrophic gastritis or gastric carcinoma in connection with positive Helicobacter pylori is another area of screening for gastrointestinal tract tumours. Pepsin determination in the insulin test is also clinically important.

24.3.5. The ^{13}C-octanoic acid breath test

The acid breath test is a non-invasive test of gastric evacuation, which can be used for a differential diagnosis of functional dyspepsias, reflux diseases and for the indication of some modern drugs (prokinetics). Octanoic acid does

not absorb in the stomach, but is absorbed quite quickly in the duodenal mucosa. Metabolic activity in the liver tissue produces $^{13}CO_2$, which is then determined in exhaled air. The range of normal values is 110 - 160 minutes for solid food and 91 - 155 minutes for semi-solid food. Gastric evacuation rate assessment using the ^{13}C-octanoate breath test exhibits a high correlation with the scintigraphic method.

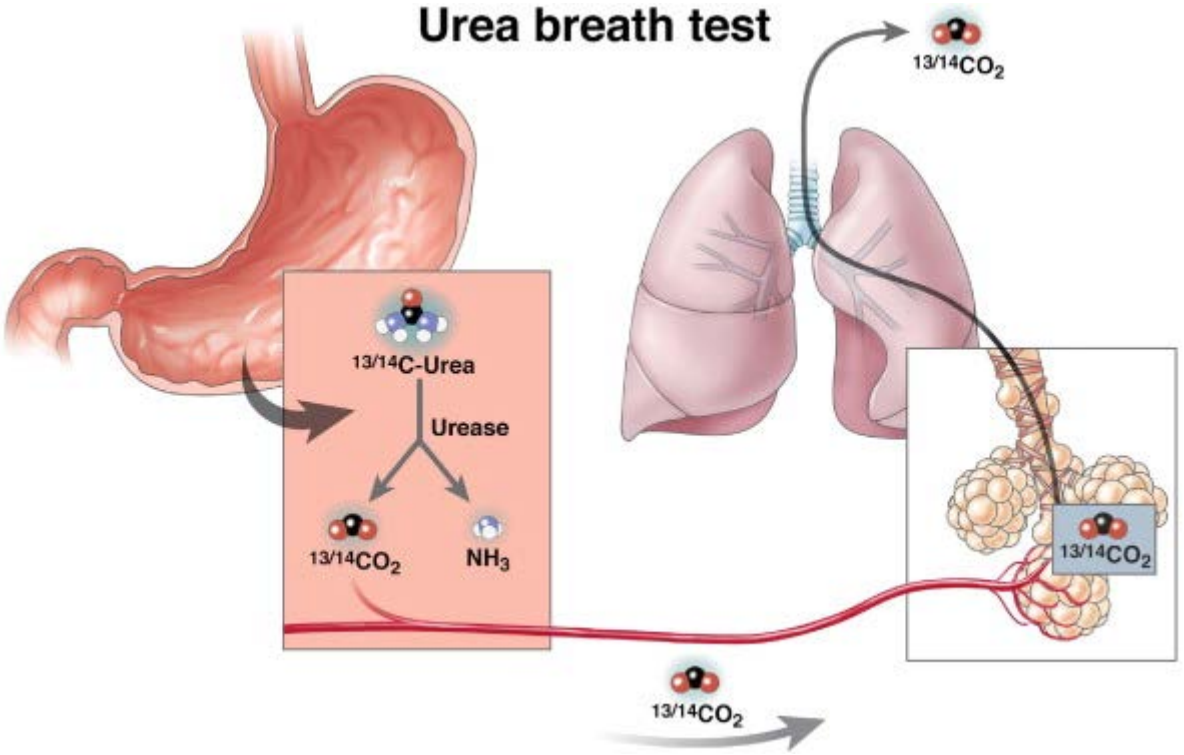


Figure 24.1. A positive answer offers conclusive evidence that the patient is infected with Helicobacter pylori. In the absence of Helicobacter pylori, the administered urea is absorbed from the gastrointestinal tract and subsequently voided.

24.4. The laboratory diagnostics of malabsorption syndrome

The diagnostics of malabsorption includes function tests aimed at clarifying the cause and degree of malabsorption (primary and secondary malabsorption syndrome), and serological tests in connection with the screening and long-term monitoring of patients with celiac disease. Function tests include challenge/tolerance tests with oral administration of D-xylose or vitamin A, and intestinal permeability tests such as the lactulose/mannitol (LA/MA) test. The lactulose/mannitol test can also be combined with the administration of D-xylose, sucralose and other sugars. Each sugar is analyzed using gas chromatography for relative indices of permeability. The average LA/MA index value is 0.016 ± 0.008 and LA/XY 0.013 ± 0.009. The intestinal permeability LA/MA test can be used for monitoring post-operative chemotherapy.

Function tests may also use other substances such as iron and vitamin B12 (Schilling test) or secretion function tests with labelled albumin (^{51}Cr-albumin test). When collecting samples following the administration of tested substances and interpreting results, one should consider not only the transit time but also the different locations in the small intestine (from the duodenum to the terminal ileum) where the tested substrate is absorbed. Non-invasive breath tests include the H_2-lactose breath test or ^{13}C-xylose breath test suitable for detecting small intestinal bacterial overgrowth.

24.4.1. A comprehensive stool analysis

Stool analysis from macroscopic evaluation, determination of volume, structure and pH to an analytical determination of laboratory markers, is considered the basic testing process if malabsorption syndrome is suspected. The qualitative determination of undigested sugars, fats and proteins – a microscopic evaluation of stained faeces is presented by the latest publications of internal medicine, gastroenterology and clinical biochemistry.

24.4.2. ß-carotene and vitamin A

Carotene and vitamin A are fat-soluble; ß-carotene is a retinoid – a vitamin A precursor, and their serum level

therefore depends on lipid digestion and absorption. Of the ß-carotene in the circulation, 80% is bound to LDL, 8% to HDL and 12% to VLDL. ß-carotene represents about 25% of serum carotenoids. The very short half-life of ß-carotene, i.e. its rapid conversion into vitamin A, is important from the clinical point of view. The determination of ß-carotene is clinically important primarily as a screening test if malabsorption syndrome is suspected. Increased ß-carotene levels have been described in hypothyroidism, diabetes mellitus, myxoedema, nephrotic syndrome, hyperlipoproteinaemia and pregnant women. Reference values depend on the test procedure; the range commonly specified for the extraction method (i.e. total serum carotenoid assessment) is 0.90 - 4.60 µmol/L, the narrower range of reference values is 1.12 - 3.72 µmol/L. Only the lower limit 0.93 µmol/L is specified for malabsorption syndrome screening in adults, and the reference range 0.37 - 74 µmol/L is described for the HPLC technique specifically assessing ß-carotene. The challenge/tolerance test with vitamin A is evaluated based on serum level increase 3 and 5 hours after administering the test challenge. Normal values 3 hours later are within the range of 3.6 - 12.6 µmol/L, and 7.2 - 24.6 µmol/L 5 hours later. Values < 3.6 µmol/L 3 hours later, or < 7.2 µmol/L 5 hours later indicate a pathological test result.

24.4.3. The D-xylose tolerance test

This test is usually indicated to confirm intestinal malabsorption in gluten-sensitive enteropathy (celiac sprue) and tropical sprue. The urine output result depends on the renal function. A falsely positive result may be found in conditions such as vomiting, dehydration, myxoedema, ascites or oedema. Many drugs such as aminosalicylic acid, acetylsalicylic acid, digitalis, indomethacin and neomycin decrease D-xylose excretion by the kidneys. The ^{14}C-D-xylose breath test and the ^{13}C-D-xylose breath test are variants of the xylose absorption test, and have been used more commonly than urine xylose determination for diagnosing intestinal malabsorption over the last few years. The serum level is 1.40 - 3.80 mmol/L 1 hour after D-xylose administration, 2.13 - 3.86 mmol/L in 2 hours, 1.27 - 2.80 mmol/L in 3 hours, 0.73 - 1.93 mmol/L in 4 hours and 0.40 - 1.20 mmol/L in 5 hours. Values < 1.67 mmol/L 2 hours after the administration of 25 g to adults, and < 1.33 mmol/L after the administration of 5 g to children, indicate a pathological result. In 5-hour urine collection, pathological values after the administration of 5 g D-xylose to children aged 5-12 are < 0.8 g/5 hrs (the broader range of normal values is 0.5 - 1.65 g/5 hrs), and < 4 g/5 hrs after the administration of 25 g to adults (some authors find results < 5 g/5 hrs pathological). The threshold drops to 3.5 g/5 hrs in individuals over 65. If D-xylose is administered based on weight (to small children in particular), the normal range of xylose excreted in the urine over 5 hours is 10 - 33% of the administered amount.

24.4.4. Celiac sprue (CS)

Celiac disease – gluten-sensitive enteropathy, is the primary malabsorption syndrome. Celiac disease is an autoimmune disorder with a genetic predisposition (HLA-DQ2/DQ8) and a specific humoral response to the triggering factor – wheat gliadin (gluten) peptides or storage proteins (prolamins) of the related cereals barley, rye and oats. The diagnosis of celiac disease is based on the histological or histochemical assessment of small intestinal mucosal biopsy specimens. Parameters assessed in celiac disease include inflammatory mucosa alterations, raised intraepithelial lymphocyte count, small intestine mucosa lesions with reduced or lost intestinal villi, hypertrophy of Lieberkühn's crypts, abnormal enterocyte maturation, and a remission of these alterations when on a gluten-free diet.

Figure 24.2. Damage to the intestine is due to a specific immune response triggered by certain peptides derived from gluten proteins in wheat, rye and barley, and sometimes oats. Damage to the intestine generally disappears over months to years after the harmful gluten proteins are completely excluded from the diet. If gluten is accidentally or deliberately reintroduced into the diet, the immune response to gluten is reactivated; many patients experience acute food poisoning and the intestine can be injured if gluten exposure continues.

Diagnostic tests for celiac disease focus primarily on celiac disease screening and include IgA and IgG class anti-gliadin antibodies (AGA-A, AGA-G), IgA and IgG class anti-endomysial antibodies (EmA) and IgA and IgG class anti-tissue transglutaminase antibodies (atTG-A, atTG-G). Tissue transglutaminase is an enzyme that plays a key role not only in the diagnosis but also in the pathogenesis of the disease itself. Gliadin, gliadin fragments and peptides contain a very high percentage of glutamine (glutamine and proline constitute 30 - 40%), and so gliadin is a very good substrate for the tissue transglutaminase enzyme, which creates a bond to the substrate (gliadin), modifies these peptides, and the neoepitopes that are produced then bind to surface glycoproteins of HLA-DQ2/DQ8-positive immunocompetent cells and induce an immune response in the small intestine mucosa. Studies from 2008 - 2010 recommend detecting antibodies against synthetic gliadin-specific nonapeptides or deamidated gliadin peptides. The sensitivity and specificity of ELISA tests for deamidated gliadin peptides (DGP) is comparable to the reliability of anti-transglutaminase antibodies; DGP antibodies in combination with IgA and IgG are even more reliable. The latest studies recommend tests in combination with IgA atTG and IgG DGP. Molecular biology methods such as PCR permit the detection of specific HLA-DQ2/DQ8 markers. Screening is particularly important in patients with another autoimmune disease. The risk of asymptomatic celiac disease (without clinical signs) in patients with type 1 diabetes is 10 times as high as the risk in the normal population, i.e. the incidence is not 1:200 but 1:20. The increased risk is similar for other autoimmune diseases (autoimmune thyreopathy, hepatitis, rheumatoid arthritis and others).

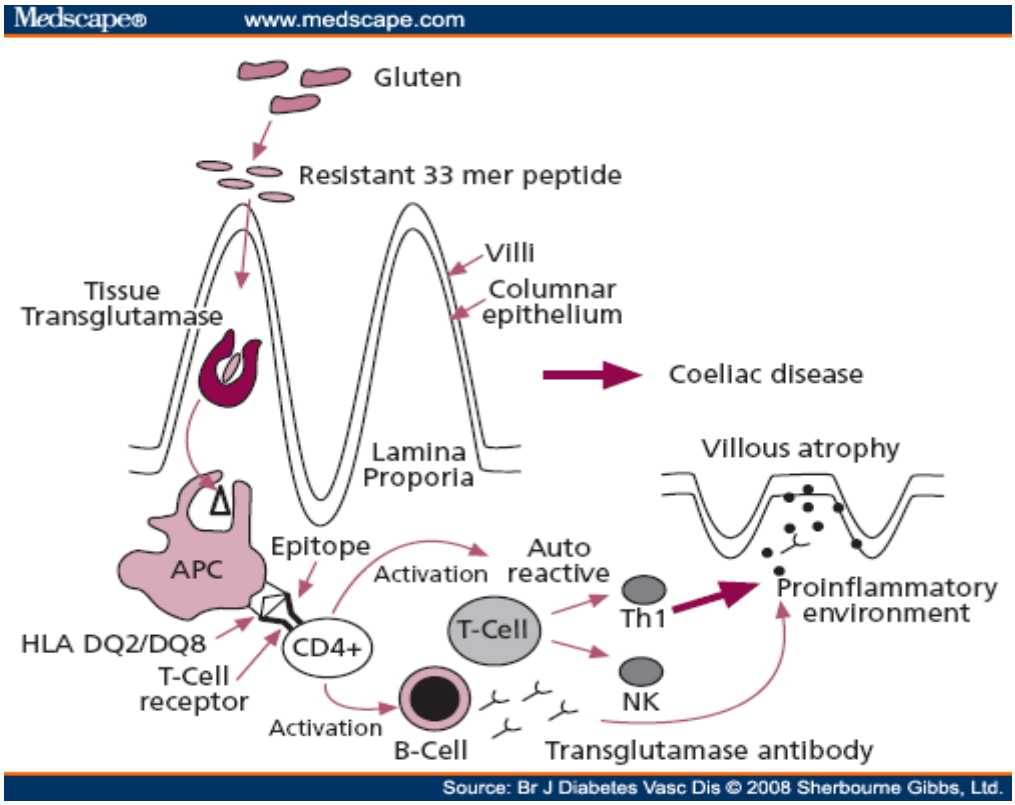


Figure 24.3. Gluten is degraded by gastrointestinal enzymes to a 33 amino acid (33-mer) peptide. The 33-mer peptide is absorbed across the small bowel epithelium to the subepithelial layer in the lamina propria. Tissue transglutaminase deaminates the 33-mer peptide. The deaminated peptides are processed by APC to three epitopes that bind to the HLA-DQ2 or DQ8 molecules. The T-cell receptor on T-cells then cross-react with the HLA molecule leading to the initiation of an autoreactive immune response with subsequent activation of B-cells, CD4+ Th1 cells and NK cells. The resultant proinflammatory environment results in further immune activation and migration of lymphocytes resulting in the characteristic pathological finding of increased intraepithelial lymphocytes and villous atrophy.

24.4.5. The lactose tolerance test

The lactose tolerance test is an indirect measurement of intestinal lactase activity for the differential diagnosis of the malabsorption syndrome lactose intolerance. The traditional test is blood sugar measurement 15, 30, 60 and 90 minutes after oral administration of 50 g of lactose. Lactase deficiency is demonstrated by a rise in blood sugar of less than 1 mmol/L. A newer lactose test consists in breath tests such as the H_2 test, measuring a rise in hydrogen concentration in exhaled air as a result of bacterial decomposition of unsplit lactose in the colon, or carbon ^{13}C detection after the administration of ^{13}C-labelled lactose. Very accurate results are provided by the combined $^{13}C/H_2$-lactose test. The lactose intolerance diagnosis can be supplemented by specific DNA genotyping – the 13910 T/C variant. Lactose intolerance can also be diagnosed by a duodenal biopsy-based rapid test, which is similar to the CLO rapid test for Helicobacter pylori diagnosis. The incubation medium in the cell for biopsy includes the glucose-oxidase and peroxidase

enzymes and the chromogenic oxidation-reduction substrate. The test takes 20 minutes.

24.4.6. Acute pancreatitis (AP)

Acute pancreatitis is a disease with variable clinical symptoms. Important issues in diagnosis are early diagnosis, severity of the disease, prognosis and monitoring the therapy process. Two stages can be distinguished in severe AP. The first is an extensive, sometimes necrotising inflammation of the pancreas with subsequent development of the Systemic Inflammatory Response Syndrome (SIRS), which may lead to Multiple Organ Dysfunction Syndrome (MODS). The next stage involves bacterial infection of the pancreatic necrosis and a deepening of systemic complications, renal failure, coagulation disorders and other manifestations. Laboratory diagnostics includes the assessment of many specific and non-specific markers, pancreatic enzymes and their precursors in the serum or urine, inflammatory markers and immune reaction markers such as cytokines, interleukins and TNF. The most commonly used parameter is the serum total α-amylase level, elevated values of which may revert to normal 3 - 5 days from the onset of the disease. The total α-amylase level is reliably elevated in 100% of acute pancreatitis cases, but it is also elevated in 80% of all acute abdomen pain cases. Diagnostically much more valuable is the pancreatic isoenzyme assay (P-type, P-AMS), whose level is also increased in 100% of acute pancreatitis cases, but only in 10% of acute abdomen pains. The diagnostic value of a serum pancreatic lipase assay is considerably higher than the α-amylase assay. The serum lipase level remains increased after an acute pancreatitis attack considerably longer than the amylase level (increased lipase activity 14 days later has been described). As serum lipase originates primarily in pancreatic cells, the assay provides considerably higher specificity, comparable with the specificity of the pancreatic α-amylase isoenzyme. A reliable early marker for acute pancreatitis is the pancreatic elastase-1 level post-ERCP and pancreatic carcinoma, where elastase-1 (determined by ELISA techniques) has the highest specificity and sensitivity to pancreatic carcinoma of all pancreatic enzymes. The earliest markers for rapid diagnosis of acute pancreatitis are the detection of TAP and CAPAP activation peptides in the urine. Serum procalcitonin has been studied over the last few years as a marker of bacterial infection of pancreatic necrosis and sepsis.

24.4.6.1. Trypsinogen, trypsinogen activation peptide (TAP) & carboxypeptidase activation peptide (CAPAP)

TAP can be detected in the serum and urine. Clinically, the most commonly used is the urinary trypsinogen-2 level; values 5600 – 10,000 µg/L correspond to a severe, serious form of acute pancreatitis and values 130 – 890 µg/L to a medium to mild form of AP. The product of trypsinogen conversion to active trypsin, trypsinogen activation peptide (TAP), is also determined in the urine. Increased TAP levels are clinically significant for evaluating the seriousness of acute pancreatis: urinary TAP values over 15 nmol/L detect medium pancreatitis, and values over 40 nmol/L a severe form of the disease. The CAPAP is longer than the other peptides released in pancreatic proenzyme activation. It is therefore more stable and more useful for laboratory diagnosis. The normal serum CAPAP level measured by the RIA is 0.8 nmol/L.

24.4.7. Chronic pancreatitis (CP)

CP is defined as a continuing inflammatory process of the pancreas, characterized by irreversible morphologic changes, causing characteristic pain and permanent loss of pancreatic function. The gold standard of functional diagnosis is the secretin-pancreozymin (secretin-CCK) test. Quantitative determination of faecal fat in a 72 - hour collection period is also used as a reference test. Of all the tests available, the secretin-CCK test provides the most accurate information about the secretory activity of the pancreas. Despite the non-standard arrangement, it is considered the “gold standard” of pancreatic function tests, in particular for pancreatic insufficiency assessment. Normal values depend on the method of stimulation, collection and analysis of the duodenum content. The values listed below are just one of the variants: secretin-stimulated secretion volume 165 - 536 mL/hr, HCO_3^- concentration 9.8 - 39.7 mmol/hr, trypsin activity 9.3 - 171 U/20 minutes, amylase activity 34 - 204 U/20 minutes. Foreign literature describes one variant, the Lundh test, in which the pancreas is stimulated by a standardized Lundh meal. The latest studies combine endoscopy, stimulation by CCK or secretin, and the analysis of lipase, bicarbonate and other enzyme concentrations in the pancreatic juice collected at endoscopy, or they recommend assays for other markers such as zinc, which is a more stable analyte compared with pancreatic enzymes. The technological advances made in imaging methods in gastroenterology today offer a combination of morphological imaging with the simultaneous evaluation of some aspects of the tested organ function. The ERCP and MRCP methods can be performed with concurrent stimulation by i.v. administered secretin or cholecystokinin, and are the originators of modern combined testing procedures such as S-MRCP, MRCPQ or ePFT (endoscopic pancreatic function test).

24.4.7.1. Human pancreatic elastase-1

Elastase-1 is synthesized by pancreatic acinar cells. The enzyme is secreted by pancreatic juice to the duodenum, and the protein sequence selected for immunochemical detection is not degraded during intestinal transit. Elastase-1 assessment therefore exhibits greater diagnostic benefit compared to a chromogenic assay for chymotrypsin in stool. The reference range is 200 - 500 µg/g of stool, the borderline range is 100 - 200 µg/g, and serious pancreatic insufficiency is found if values are < 100 µg/g of stool. The immunochemical assay for elastase-1 is not affected by transit through the colon, substitution therapy or other factors affecting the enzyme assay for chymotrypsin in stool. The specificity of the method is 93%, sensitivity is 100% for severe pancreatic insufficiency and 87% for medium and mild forms. This test is commonly used in paediatrics to demonstrate cystic fibrosis with almost 100% specificity and sensitivity. Falsely reduced values can be caused by dilution (water content) during diarrhoea.

24.4.7.2. The breath test with ^{13}C-mixed triglycerides

The principle is the cleavage of this substrate by pancreatic lipase. The ^{13}C-MTG substrate is a triglyceride with ^{13}C - labelled octanoate in position 2 and stearate in positions 1 and 3. Pancreatic lipase cleaves triglyceride, and ^{13}C--octanoate is further oxidized in the liver (beta-oxidation of fats). The amount of $^{13}CO_2$ in exhaled air is measured with a breath test analyzer. The ^{13}C-MTG test is clinically important in the differential diagnosis of malabsorption syndrome, as a function test of exocrine function of the pancreas, and for the long-term monitoring of patients with chronic pancreatitis. ^{13}C breath tests with mixed triglycerides (MTG) or hiolein are indirect function tests of exocrine pancreatic function, and can be used to monitor the success of pancreatic substitution therapy.

24.4.8. Colonic pathologies

Laboratory diagnostics of colonic pathologies focuses primarily on screening for colorectal tumours (CRCA), faecal occult blood tests, the haemoccult guaiac screening test (gFOBT), the more sensitive immunochemical test (iFOBT), and the quantitative faecal occult blood test (qi-FOBT). The activity of inflammatory diseases and tumours can be monitored by many faecal markers, the detection of calprotectin, lactoferrin, pyruvate kinase M2 and the protein S100A12. Progress in molecular biology allows the isolation of DNA from a stool specimen and the assessment of many genetic markers.

The risk of colon carcinoma in the Czech population continues to rise, with 80 cases per 100,000 inhabitants in 2010, which is the highest incidence in Europe. Colorectal carcinoma (CRCA) develops as an adenoma malignization by a sequence of gene mutations over the course of ten years (on average) and most of the tumour tissue is localized intraluminally, at least at the initial stage of the disease. A small proportion of colorectal carcinomas develop in the presence of risk factors such as non-specific inflammatory bowel diseases, familial adenomatous polyposis, hereditary nonpolyposis colorectal cancer and other hereditary syndromes. The carcinoma (and the adenomatous polyp) already intermittently bleeds at the asymptomatic stage of the disease. Screening programmes include primarily laboratory methods of occult blood detection such as faecal occult blood test (FOBT) (*test okultního krvácení ve stolici* – TOKS – in Czech), the detection of genetic markers, and imaging methods such as sigmoidoscopy, colonoscopy, irrigography, virtual computer-aided colonography using computed tomography (CT) or magnetic resonance (MR). Progress in molecular biology and the application of PCR in routine diagnosis are also opening up brand new avenues in GIT tumour screening. The latest screening methods are based on the detection of specific mutations of the DNA isolated from a stool specimen using PCR methods or biochip array technology. APC, K-ras, p53, microsatellite instability (MSI) and other genetic markers can be assessed for the detection of colon tumours from stool specimens. Detection of specific protein (vimentin) methylation may also be used as molecular markers of colorectal cancer, and these DNA alterations can be monitored in the stool and serum of patients with CRCA.

For many years, the only test recommended for screening was the Haemoccult - guaiac test. The test is based on the pseudoperoxidase reaction of haemoglobin and contains a test medium (strip) impregnated with guaiac resin or, in former tests, a derivative of benzidine such as dimethylbenzidine (o-tolidine). Given the chemical principle of the oxidation reaction, tests are affected by the presence of other oxidizers (vitamin C), the presence of haemoglobin from food (meat, blood), and a falsely positive result may also be caused by the presence of vegetable peroxidases (some root vegetables). Defined dietary limitations are therefore recommended depending on how the test is arranged. The sensitivity of first generation gFOBTs is very low, 26 - 30% for cancer detection; however, gFOBTs have almost zero false positivity. Qualitative immunochemical tests (iFOBT) of the second generation are almost twice as sensitive, but their specificity is considerably lower and false positivity is almost 25% for some tests. The third generation of FOBTs, qiFOBT with quantitative faecal Hb detection, provide 90 - 95% sensitivity and specificity by optimizing cut-off values. This is

why quantitative immunochemical tests, qiFOBT, are recommended specifically for CRCA screening at present.

24.4.8.1. Calprotectin

Calprotectin is a 36.5 kDa calcium binding protein with antimicrobial activity; it is composed of two heavy chains and one light chain, and is primarily derived from monocytes and neutrophils. Calprotectin as a marker detected in the stool exhibits lower variability than haemoglobin and is a suitable indicator in the diagnosis and monitoring of therapy for inflammatory bowel diseases, ulcerative colitis and Crohn's disease, or necrotising enterocolitis in children. The determination of concentration in a stool specimen with a cut-off of 30 mg/L exhibits 97% specificity and 100% sensitivity for the differential diagnosis between acute Crohn's disease and irritable bowel syndrome (IBS). Faecal calprotectin is also tested as a colorectal cancer marker. Another inflammatory disease marker measured in the stool is lactoferrin, which has a superior sensitivity to calprotectin.

For more detailed information about each laboratory method, test procedures, links to current NLM Medline abstracts and 100 colour diagrams and photos, refer to the up-to-date on-line GastroLab at the First Faculty of Medicine website - *http://glab.zde.cz*.

CHAPTER 25

25. Diabetes Mellitus

Authors: prof. MUDr. Jaroslav Racek, DrSc.,

Reviewer: prof. MUDr. Richard Průša, CSc.

25.1. Definition and Incidence of the Disease

Diabetes mellitus is a **metabolic disease with two major causes**:

- Absolute deficiency of insulin
- Resistance to insulin (i.e. tissues do not respond to the action of insulin)

The incidence of diabetes has an ever growing tendency in developed as well as developing countries. While Czechoslovakia had about 4% of diabetics 30 years ago, the incidence in the Czech population is almost 8% today; another few percent of people have diabetes without knowing it. The number of diabetics in developed countries is expected to rise by another 20% and in developing countries by up to 70% in the next 20 years. With the present trend unchanged, about 33% of the US population will suffer from diabetes in 2050. This is especially due to unhealthy lifestyle and a growing number of obese people. *Figure 1* illustrates the trend in the worldwide incidence of diabetes in absolute and relative figures and its approximation to 2030.

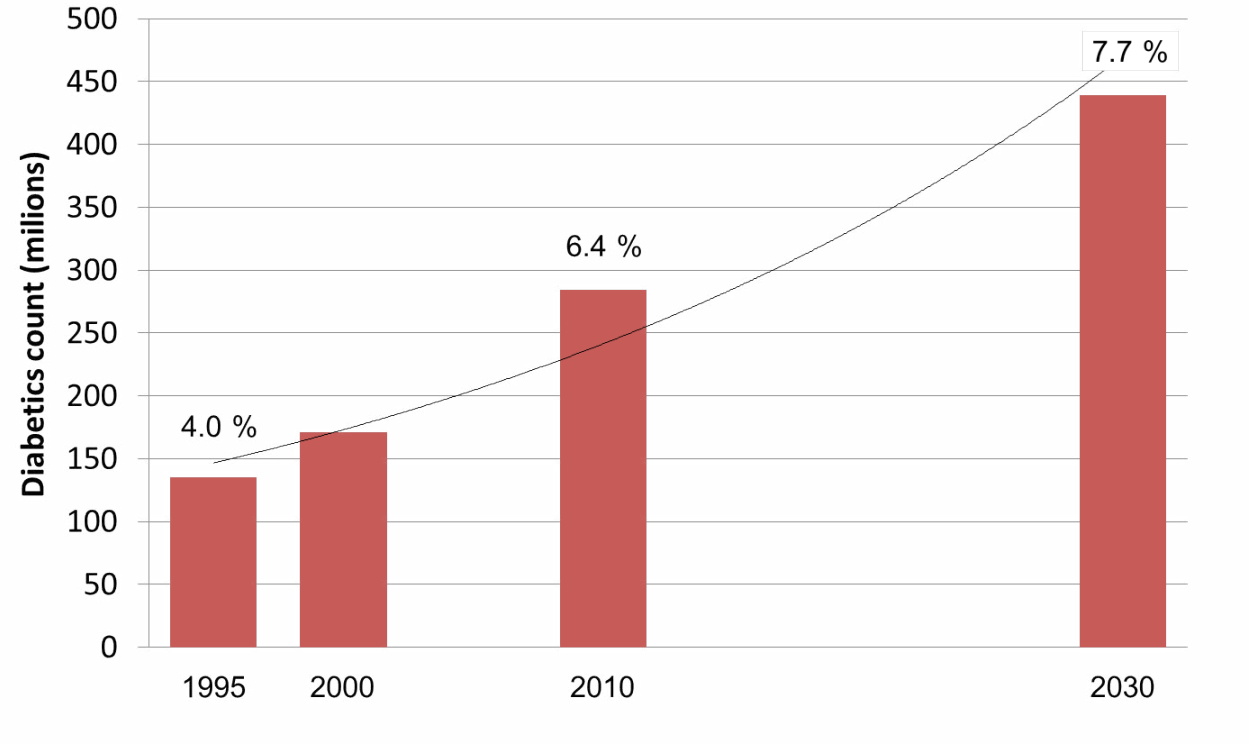


Figure 25.1. Trend in worldwide incidence of diabetes

Diabetes mellitus is a very serious disease due to potential complications that shorten the patient's life and are exceedingly expensive for the healthcare system. Early diagnosis of the disease is of paramount importance: 20 - 30% of newly diagnosed patients are estimated to have late diabetic complications. As complications develop slowly, over seven or more years of the disease, diabetes could have been identified in these people and adequately treated years before.

25.2. Clinical and Laboratory Signs of Diabetes

Clinical signs of diabetes are based primarily on glucose losses in urine accompanied by osmotic diuresis. The patient then has **polyuria** (increased volume of urine) followed by **polydipsia** (excessive drinking) and a **loss of weight** due to a loss of energy - rich substances (glucose and often ketone bodies) in the urine.

Laboratory signs of diabetes can be derived very easily from the knowledge of insulin effects (or the absence of it):

Hyperglycaemia (glucose is unable to penetrate into cells due to the lack of insulin or insulin resistance);

Glycosuria (when the renal threshold for glucose is exceeded, the offer of glucose to renal proximal tubular cells exceeds their maximum absorptive capacity and glucose gets into the definitive urine);

Ketonemia and ketonuria (cells lack glucose, so they have to get energy by cleaving fats or fatty acids, which the human body is unable to burn completely without concurrent metabolism of glucose. This results in the production of ketone bodies – acetoacetic acid, β-hydroxybutyric acid and acetone.) These acids cause **metabolic (keto)acidosis**.

25.3. Blood Glucose (Glycaemia) Testing

Blood sugar (glucose) testing is one of the basic diabetes tests used to diagnose and monitor treatment. Generally, two biological materials are used:

Venous blood; an addition of anticoagulant (K_3EDTA, Na_2EDTA) prevents blood clotting, while sodium fluoride (NaF) is used as a glycolysis inhibitor and provides a stable glucose concentration for at least 24 hours. As fluoride inhibits final glycolysis reactions, initially we see a slight decrease in glycaemia; this is not observed, though, if the blood is acidified by citrate at the same time. Venous blood plasma is necessary for the diagnosis of diabetes. The glucose concentration is sometimes measured in venous blood serum – usually when additional laboratory tests are required. This method requires timely separation of the serum from red blood cells;

Capillary blood (usually from the fingertip) is used for monitoring diabetes management. Patients may measure blood sugar by themselves with their own glucose meter; the same method is used for POCT blood sugar measurement in ICUs, operation theatres or acute admission units in hospitals. Capillary blood can also be sent for laboratory testing or collected at home for glucose profile monitoring; in this case, blood is collected in plastic „pointed" micro test tubes containing anticoagulant and NaF; a volume of blood adequate to the added preservative has to be taken and then mixed in the test tube. Another option is collecting an exact amount of capillary blood (usually 20 µL) in a short capillary tube and transferring the blood to a little test tube with a lysis buffer; this ensures a stable glucose level while being transported to the laboratory.

Needle-free **transdermal glucose monitoring** is in the phase of clinical studies; this facilitates patient care in intensive care units where repeated measurement of blood glucose level is required.

25.4. Diagnosis of Diabetes

As mentioned in the preceding chapter, the only acceptable biological material recommended for the diagnosis of diabetes by the Czech Diabetes Society and the Czech Society of Clinical Biochemistry is venous blood plasma.

The following glucose levels in the venous blood plasma are indicative of diabetes mellitus:

> 7.0 mmol/L, if the patient was fasting

> 11.0 mmol/L, if it is unsure whether the patient was fasting

However, this value has to be confirmed by a new blood sample collected on another day. If the result above is found only once, typical clinical signs of diabetes have to be present to establish the diagnosis.

If the above glucose level is found only once and the patient does not have clinical signs of diabetes, or if fasting venous blood plasma glucose is within the range of 5.6 – 7.0 mmol/L (slightly elevated), the diagnosis of diabetes has to be confirmed (or ruled out) using a challenge test – oral glucose tolerance test (oGTT).

25.4.1. Glucose Tolerance Test (oGTT)

Indications for oGTT have been listed in the paragraphs above. The patient has to be instructed about preparations for the test. At least three days before the test, the patient should eat normal food without carbohydrate limitations and be fasting from the evening.

The test itself consists of challenging the patient by 75 g of glucose administered *per os*. The patient has one minute to drink a solution with 75 g of glucose, usually flavoured with lemon or citric acid. The patient then rests and another venous blood sample is taken two hours later. The tested individuals are classified into the following groups based on the venous blood plasma glucose level:

Glucose level < 7.8 mmol/L: *normal glucose tolerance*

Glucose level > 11.0 mmol/L: *diabetes mellitus*

Glucose level 7.8 – 11.0 mmol/L: *impaired glucose tolerance*; repeating the oGTT is recommended in a few months

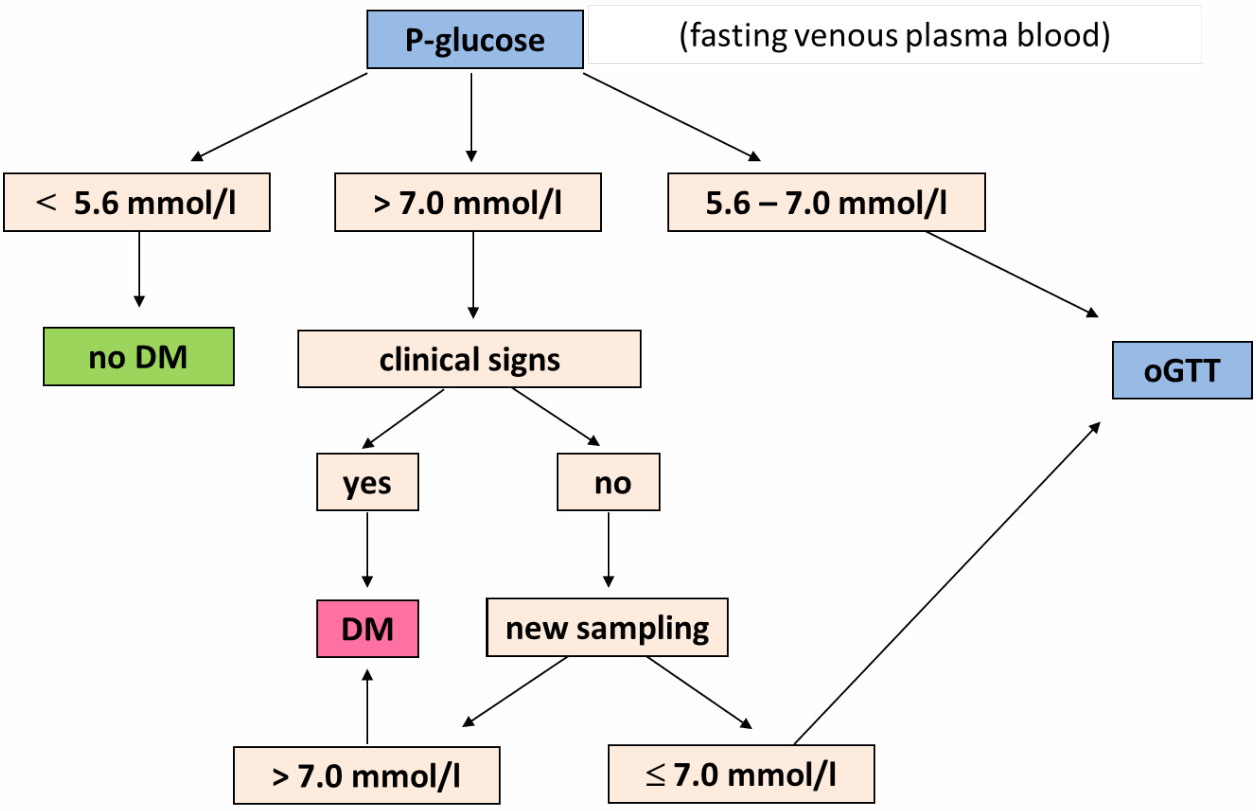


Figure 25.2. Diabetes diagnosis algorithm

Notes:

oGTT is less common in the USA. The condition, which is an indication for the oGTT in the CR (i.e. fasting venous blood plasma glucose > 7.0 mmol/L found once in an individual without signs of diabetes or with glucose level within 5.6 – 7.0 mmol/L), is called *impaired fasting glucose*. Together with impaired glucose tolerance (i.e. venous blood plasma glucose level within 7.8 – 11.0 mmol/L 2 hours after the oGTT), it is sometimes referred to as *prediabetes*. There is no doubt that these individuals are at high risk of developing diabetes but also cardiovascular diseases and even some malignant tumours.

The criteria for diagnosing diabetes have changed, or rather tightened over the years. Original fasting venous blood plasma glucose for the diagnosis of diabetes was 7.8 mmol/L. Based on the finding that diabetic retinopathy develops in a significant percentage of people with levels under 7.8 mmol/L, the decision limit was lowered to 7.0 mmol/L in 1997. Similarly, the lower decision limit for fasting hyperglycaemia was lowered from 6.1 to 5.6 a few years ago. These changes are aimed at timely capturing patients at risk of developing future diabetes and, in particular, late diabetic complications.

The oGTT procedure for pregnant and non-pregnant women is the same (challenge with 75 g of p.o. glucose) but

the evaluation of results is different, being more stringent. The fasting glucose level should not exceed 5.5 mmol/L, and 7.7 mmol/L 2 hours after the challenge. The glucose level may optionally be evaluated 60 minutes after the challenge and should not exceed 8.8 mmol/L. Gestational diabetes has been described in greater detail chapter Laboratory examinations during pregnancy.

25.5. Laboratory Monitoring of Diabetes

If a patient is diagnosed diabetes mellitus, laboratory tests are required. Three levels of laboratory monitoring are distinguished depending on the time covered by the laboratory testing:

25.5.1.1. One-off glucose test

The test shows current blood glucose concentration at the time of collection. It is important especially if serious hyper or hypoglycaemia is suspected. The patient can make a self-test on fingertip blood using diagnostic strips and read the result on the personal glucose meter; an immediate reaction (i.e. taking insulin or conversely carbohydrates) is possible based on the result.

25.5.1.2. Blood glucose profile

The blood glucose profile reports blood glucose changes during a period of 24 hours. Fingertip blood is taken at predefined times (5 to 11 collections per day). Blood is taken before and after meals and, in indicated cases, also at 2 a.m. to reveal potential night-time (nocturnal) hypoglycaemia that could be masked by consequent reactive hyperglycaemia in the morning; this is sometimes referred to as the „dawn phenomenon".

This is not an acute test; adequately preserved samples (with sodium fluoride) can be stored and sent to the laboratory all at once. Insulin therapy is managed depending on the levels found (insulin type, intervals, doses).

25.5.1.3. Long-term diabetes compensation – glycated haemoglobin

The attending physician needs to know what the patient's compensation has been over a longer period of time. The glycated haemoglobin test is typically used for this purpose.

25.5.1.4. Glycation – explanation and principle

Glycation is a non-enzymatic bonding process of carbohydrate molecules to free amino groups in biomolecules. The molecule of glucose most often binds by its aldehyde group to a free amino group in a protein, ε-amino group of lysine in particular. The Schiff base (chemical name: aldimine) is formed in the condensation reaction. The reaction is fully reversible: the reaction goes from the left to the right during hyperglycaemia; if the glucose level drops, the glucose molecule cleaves off again and the reaction proceeds in the opposite direction.

However, this holds true only if hyperglycaemia is temporary. In the case of long-term hyperglycaemia, the aldimine molecule is restructured in a process called intramolecular rearrangement (called *Amadori* rearrangement after the person who discovered it). This reaction is irreversible and gives rise to ketoamine, a derivative of fructosamine. Fructosamine remains bound to the protein molecule until the molecule has been entirely degraded and eliminated from the body. *Figure 3* shows the glycation diagram.

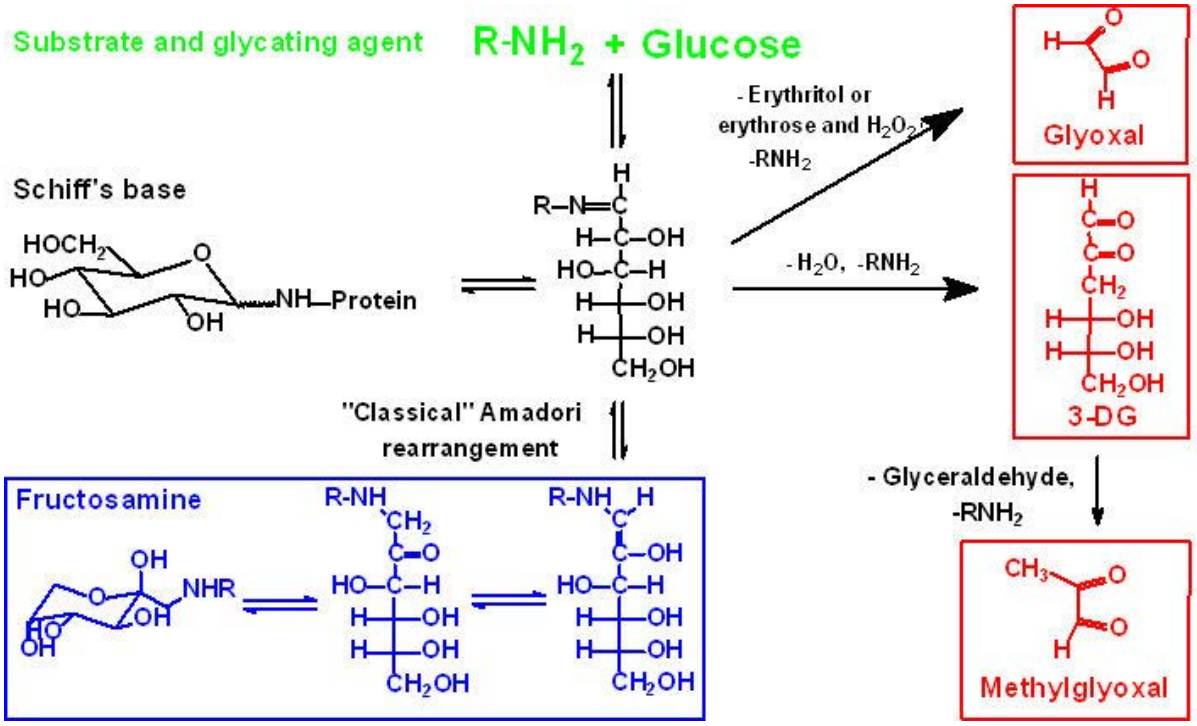


Figure 25.3. Diagram of protein glycation (explained in the text)

25.5.1.5. Glycated haemoglobin as an indicator of long-term compensation of diabetes

If suitable protein is chosen, protein glycation intensity may be indicative of long-term average glucose level, i.e. diabetes compensation. Such a molecule is a molecule of haemoglobin. The beta-chain in haemoglobin A is readily glycated; the biological half-life of haemoglobin is very long, and although enclosed inside red blood cells, it is available to glycation in diabetics – glucose penetrates red blood cells even without the effect of insulin. The measured parameter is then *glycated haemoglobin*, usually referred to as HbA_{1c} or only *A1c* for short.

Glycated haemoglobin shows the patient's compensation (i.e. average glucose level) for about 2 – 3 months before blood collection. It is usually expressed as a total haemoglobin percentage, with an upper reference limit of 4% for healthy individuals. In a well-compensated diabetic, the glycated haemoglobin ratio should not exceed 4.5%; compensation levels under 6% are satisfactory, and higher levels are unsatisfactory.

The figures above apply to glycated haemoglobin test calibration using IFCC-certified calibration, applicable in the CR since 2004. Many other countries (e.g. USA and many European countries) use the former National Glycohemoglobin Standardization Program (NGSP) certified calibration with resulting figures of around 2% higher. Lately there has been a tendency to express the glycated haemoglobin concentration in SI units indicating the amount of fructosamine molecules bound to haemoglobin. The unit is mmol/mol. The relationship is roughly as follows: HbA_{1c} 4.75% (IFCC) = 6.5 (NGSP) = 47.5 mmol/mol (SI units).

The calculation of estimated average glucose (eAG) sometimes recommended in the USA is not used in the CR. Glycated haemoglobin is converted using a simple formula to a kind of average glucose level, which is easier to understand for the patient; this, however, does not allow for individual differences in the glycation rate.

25.5.1.6. Glycated haemoglobin in the diagnosis of diabetes

As recommended by the Czech Diabetes Society and the Czech Society of Clinical Biochemistry (2012), the glucose tolerance test (oGTT) result is decisive for the diagnosis of diabetes in cases of doubt. The glycated haemoglobin test is used only to evaluate long-term diabetes compensation.

Lately there have been efforts in the US to employ glycated haemoglobin concentration also for the diagnosis of diabetes. There are several reasons for this:

- Better analytical parameters
- Higher stability of the analyte in biological material specimens

- Lower biological variability of HbA_{1c} (under 2 %)
- Higher comfort for the patient (fasting and drinking the glucose solution is not required);
- Minimal influence of stress on the result
- Use of the same test for the diagnosis and monitoring therapy

A group of experts from the American Diabetes Association (ADA), European Association for the Study of Diabetes (EASD) and the International Diabetes Federation (IDF) recommends using 6.5% as a glycated haemoglobin cut-off for diagnosing diabetes (NGSP); the corresponding IFCC-calibrated cut-off is 4.75%, i. e. 47.5 mmol/mol in SI units. According to the NGSP, values from 5.7 to 6.5% are subsequently indicative of an increased risk of developing diabetes or late diabetic complications. Glycated haemoglobin and the risk of late complications, diabetic retinopathy in particular, have been found to correlate fairly well. As this is a continuous function and it is impossible to find a value from which the increase in the risk is steeper, 5.7% was chosen by agreement in order to find a sufficient number of patients to whom qualified care could be provided.

The diagnosis of diabetes was repeatedly compared based on the classical criteria (fasting and postprandial glucose level, oGTT result) and based on the newly introduced criterion – the glycated haemoglobin level. That is to say that each of the criteria covers a somewhat different part of the population and the overlap is only partial (see *Figure 4*), while the glucose tolerance test seems to be more sensitive than the glycated haemoglobin test in detecting diabetes. Glycated haemoglobin should be used particularly in people who cannot guarantee proper preparation for the oGTT.

As recommended by the Czech Diabetes Society, glycated haemoglobin may be used in the screening of glucose homeostasis disorder, in respect of prediabetes in particular, but not for the diagnosis of diabetes.

Diabetes Test*	ADA Goal for	
	Diagnosis of Diabetes	Increased risk [Prediabetes]/IFG
HbA1c Using a method certified by NGSP and standardized to the DCCT assay. or	≥6.5%	5.7-6.4%
Fasting Plasma Glucose Fasting is defined as no caloric intake for at least 8 hours. or	≥126 mg/dL (7.0 mmol/L)	100-125 mg/dL (6.9 mmol/L)
2 Hour Plasma Glucose [OGTT] The test should be performed as described by the WHO, using a glucose load containing the equivalent of 75 g anhydrous glucose dissolved in water.	≥200 mg/dL (11.1 mmol/L)	140-199 mg/dL (7.8-11.0 mmol/L)

Figure 25.4. The diagnostic criteria for diabetes do not cover the same individuals (based on Škrha)

25.5.1.7. Glycated albumin

Glycated protein was once used as an indicator of medium-term diabetes compensation; the test was known as fructosamine. The test predominantly measured the glycated molecule of albumin. As the method was non-specific and had many interfering influences, it was gradually abandoned.

Several specific methods have been developed recently to determine solely *glycated albumin*. Glycated albumin is used as a patient's compensation indicator 2 - 4 weeks before blood collection and can be used in cases where glycated haemoglobin does not reliably indicate long-term diabetes compensation (shorter red blood cell survival, haemoglobinopathy etc.).

25.5.1.8. Point-of-care testing (POCT) and self-monitoring of diabetics

There is a possibility of testing the diabetic in the physician's office, in the operating theatre or at home by patients themselves. It is currently unimaginable to think of caring for a diabetic without being able to monitor the glucose level and immediate reaction to the reading.

Personal glucose meters providing an objective glucose evaluation using a diagnostic strip have been available for decades. State-of-the-art instruments work on the electrochemical principle and are able to store many results in the memory. Patient (or non-laboratory healthcare staff) instruction and external quality assessment of the glucose meter are required to ensure proper measurement. I deem it necessary to emphasize that glucose meters are used for monitoring diabetes and not to diagnose the disease; this test exhibits significantly worse reproducibility as compared with sophisticated methods for measuring the glucose level in clinical laboratories.

Additional options of POCT have recently been offered to diabetics. These mainly involve glycated haemoglobin. Unfortunately, many instruments manufactured for this purpose fail to show acceptable analytical parameters. The microalbuminuria test using a special strip (not a strip for a proteinuria demonstration!) is also possible; as this is by no means an acute test, it is more suitable to refer the test to the relevant clinical laboratory.

25.6. Other Laboratory Tests of Diabetic Patients

25.6.1. Microalbuminuria

Only a minimum amount of albumin and a certain amount of microproteins penetrate a healthy glomerulus (the glomerular membrane has a physiologically negative charge like albumin, which prevents a greater amount of albumin from penetrating it). These proteins are almost completely taken up by proximal tubule cells and are metabolized. However, the re-absorption capacity of the proximal tubule does not have a big functional reserve, and should the penetration of proteins through the glomerular membrane increase, non-absorbed proteins can be demonstrated in urine. This is also the case for the initial phase of diabetes, when temporary glomerular hyperfiltration may be seen; this is also partly due to the increased permeability of the glomerulus as a result of glycation or aseptic inflammation. An increased (greater than physiological) amount of albumin is then found in urine, although indemonstrable from protein test strips. This condition is referred to as *microalbuminuria* and the demonstration requires a more sensitive, immunochemical method. The diagram in *Figure 5* explains the concept of microalbuminuria; as the albumin concentration in urine depends on the patient's hydration, in practice the result is not given in mg/L but in units taking this fact into account (see below).

Table 28. Definitions of Proteinuria and Albuminuria

	Urine Collection Method	Normal	Micro-albuminuria	Albuminuria or Clinical Proteinuria
Total Protein	24-Hour Excretion (varies with method)	<300 mg/d	NA	>300 mg/d
	Spot Urine Dipstick	<30 mg/dL	NA	>30 mg/dL
	Spot Urine Protein-to-Creatinine Ratio (varies with method)	<200 mg/g	NA	>200 mg/g
Albumin	24-Hour Excretion	<30 mg/d	30–300 mg/d	>300 mg/d
	Spot Urine Albumin-Specific Dipstick	<3 mg/dL	>3 mg/dL	NA
	Spot Urine Albumin-to-Creatinine Ratio	<30 mg/g	30-300 mg/g	>300 mg/g
	Spot Urine Albumin-to-Creatinine Ratio (gender-specific definition)[v]	<17 mg/g (men) <25 mg/g (women)	17–250 mg/g (men) 25–355 mg/g (women)	>250 mg/g (men) >355 mg/g (women)

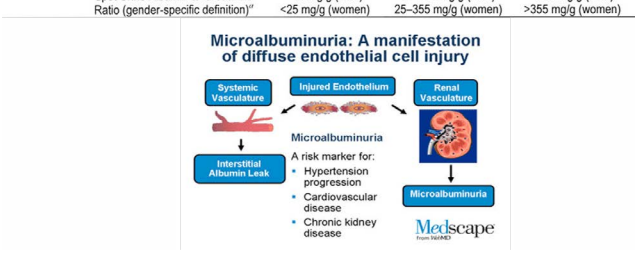


Figure 25.5. Explanation of microalbuminuria

The microalbuminuria test is used for early prediction of diabetic nephropathy in diabetics. Patients with diabetes should be tested for diabetic kidney disease on an annual basis. In type 1 diabetic patients, the screening should start 5 years upon being diagnosed with diabetes, and immediately after diagnosis in the case of type 2 diabetic patients. The screening includes an albumin/creatinine ratio test in the first morning urine specimen. A 24 - hr urinary albumin test is not recommended but albuminuria can be tested in a sample of urine collected while resting on the bed overnight; the result is then given in µg/min. One-off urine tests are preferred in practice, though; the result is related to the urinary creatinine concentration (ACR = albumin/creatinine ratio; the result is given in mg/mmol creatinine); *Table 1* shows the criteria for microalbuminuria. Given the high intraindividual variability (up to 30%), at least 2 out of 3 urine samples taken over 3 - 6 months should be positive to diagnose microalbuminuria; the test should not be carried out while concurrent urinary tract infection is present, after enduring higher physical stress and during menses. Diabetic kidney disease is always present in the form of incipient diabetic nephropathy if microalbuminuria is demonstrated.

Improved metabolic compensation of diabetes, satisfactory correction of hypertension and treatment with angio-

tensin-converting-enzyme inhibitors (ACEI) reduce the risk of renal injury progression. Patients with microalbuminuria (also non-diabetic patients) have a considerably increased cardiovascular risk. Microalbuminuria seems to be a marker for generalized endothelium dysfunction in these patients. Recently it has transpired that people with „high normal albuminuria" are also at increased renal and cardiovascular risk.

Urinary albumin concentration testing is also important in patients with arterial hypertension, where it may signalize incipient kidney injury.

Table 25.1. Criteria for microalbuminuria

Type of Sample	Microalbuminuria
Random specimen	2.6 – 29.9 mg/mmol creatinine (males)
	3.6 – 29.9 mg/mmol creatinine (females)
Collected urine	30 – 299 mg/24 hrs
	20 – 199 µg/min

Note: The preferred criterion is ACR (mg/mmol creatinine)

25.6.2. Autoantibody Tests

Type 1 diabetes mellitus is an autoimmune disease caused by β-cell destruction by cytotoxic T-lymphocytes. In addition, different types of diabetes-associated autoantibodies can be demonstrated in the blood serum of these patients; these do not have cytotoxic effects but they do signalize an ongoing autoimmune process. The effect of cytotoxic lymphocytes is very adequately compared to fire; autoantibodies form only smoke, which is produced by fire but does not have destructive effects itself.

Autoantibodies can be demonstrated months to years before the onset of clinical symptoms: More than 80 - 90% of β-cells have to be destroyed to have insulin production so low to induce symptoms of diabetes. Autoantibodies often fade away as diabetes proceeds.

The following DM1-related autoantibodies are demonstrated in practice:

- Islet cells autoantibodies (ICA)
- Insulin autoantibodies (IA)
- Glutamic acid dexarboxylase autoantibodies (GADA)
- Insulinoma 2-associated autoantibodies (IA-2A)
- Zinc transporter 8 protein islet autoantibodies (ZnT8A)

Except for insulin, all antigens, against which the autoantibodies are produced, are localized intracellularly; that is why β-cells have to be destroyed to make antigens accessible to immunocompetent cells.

ICA and GADA can be identified in 70 - 80% of people with DM1 at the onset of the disease; other types of autoantibodies have a slightly lower sensitivity. The detection of autoantibodies signalizes an increased risk of DM1 development in the future; the risk rises with the number of detected antibody types. About 3 – 16 – 40 and 50% risk is described in 1 – 2 – 3 and 4 types of autoantibodies found. The risk further increases if relatives have DM1 or if a high-risk HLA genotype is present.

It is to be expected that assays for autoantibodies in relatives with DM1 will be very important at the time when we are able to delay the onset of the disease by dampening the autoimmune process through preventive treatment.

As mentioned above, type 1 diabetes mellitus is an autoimmune disease. As autoimmune diseases often group together, considering DM1 is recommended also in people with another diagnosed autoimmune disease (celiac disease, autoimmune thyroiditis, etc.). And vice versa, the possibility of other autoimmune diseases should be considered in patients with DM1.

25.6.3. Insulin Resistance Quantification

Insulin resistance is typical not only of type 2 diabetes mellitus, but also of *metabolic syndrome*. Insulin resistance

should sometimes be objectified and its grade evaluated to assess therapeutic measures. This is, however, carried out using several methods that are not performed on a routine basis.

The *HOMA* (Homeostatic Model Assessment) *index* is computed from the serum glucose concentration and fasting insulin: HOMA = (glucose x insulin)/22.5. Normal values are under 1, probable insulin resistance is over 2, and very probable over 2.5.

Insulin curve – it is actually an oral glucose tolerance test prolonged to three hours after the oral administration of glucose. Insulin resistance will manifest itself by insulin overproduction, i.e. hyperinsulinism.

The hyperinsulinemic-euglycemic clamp is the most complex test to perform. The tested person is administered an i.v. infusion of insulin at the same time with concurrent glucose infusion. The glucose infusion rate is changed to have the glucose level within 5 – 5.5 mmol/L. Insulin resistance will manifest itself by lowered amount of glucose which should be administered to maintain euglycaemia.

25.6.4. Insulin and C-Peptide Test

Both insulin and C-peptide can be tested by immunochemical techniques using specific antibodies, but the tests are not commonly used in practice. As β-cells of the pancreas secrete insulin and C-peptide in equimolar amounts, C--peptide can be considered to be an indicator of endogenous insulin secretion. Unlike the insulin test, the C-peptide level is not affected by exogenous insulin or present anti-insulin antibodies. The C-peptide test is therefore important for the evaluation of residual insulin secretion in patients with type 1 diabetes, and for patients with hypoglycaemia of unclear aetiology. *Figure 6* shows C-peptide and insulin generation from proinsulin.

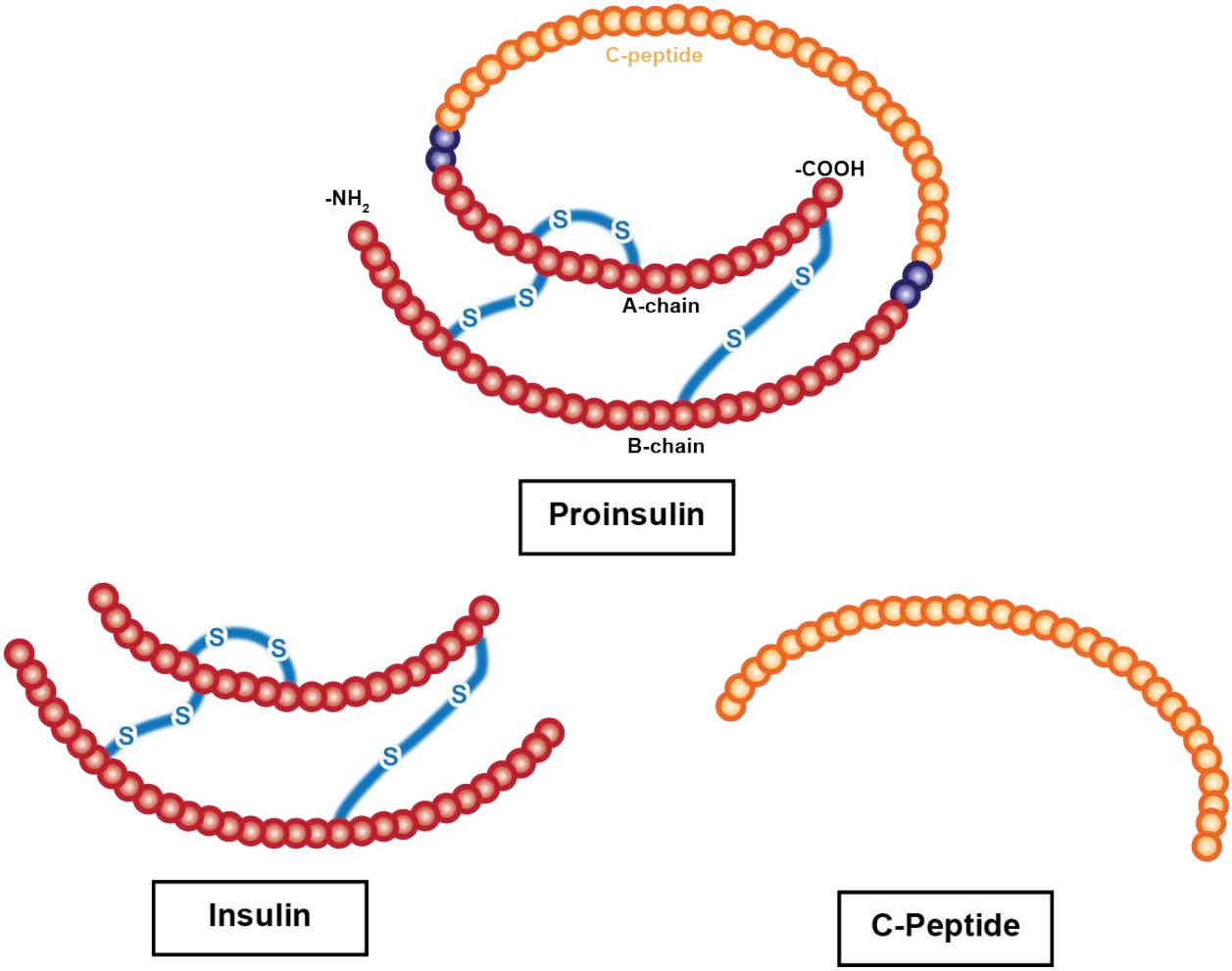


Figure 25.6. Diagram of proinsulin splitting into insulin and C-peptide molecules

25.7. Complications of Diabetes

Diabetes mellitus is serious primarily due to its complications which can be even life threatening. They can appear

quickly and develop within hours to days, sometimes even tens of minutes (acute complications of diabetes); in other cases the disease lasts for decades before complications manifest themselves (late/chronic complications). We cover these complications in brief with respect to their relation to laboratory tests in the following paragraphs. Refer to Internal Medicine for details.

25.7.1. Acute Complications of Diabetes

These complications occur suddenly and may present an immediate danger to the patient's life. In general terms we can distinguish two acute complications of diabetes.

25.7.1.1. Hyperglycaemic Coma

Hyperglycaemic coma is caused by an insulin deficiency. Type 1 diabetics with absolute insulin deficiency are afflicted more often, but type 2 diabetics are not spared either. The patient has considerable hyperglycaemia and loses high amounts of fluids and minerals (osmotic diuresis) through urine, which may even lead to a shock. Plasma osmolality grows. The incapacity of cells to metabolize glucose leads to fatty acid oxidation and ketone body production. This is why hyperglycaemic coma is often accompanied by diabetic ketoacidosis with a considerable decrease in blood pH, which is the main cause of loss of consciousness. Ketone bodies are tested in urine. Tissue hypoxia in patients in shock may lead to the reduction of acetoacetic acid to β-hydroxybutyric acid, which does not react with the diagnostic strip for the ketone body test; the results are then falsely negative.

The compensation for ketoacidosis manifests itself by deep Kussmaul breathing; the patient's breath may smell of acetone. Lactate overproduction in hypoxic tissues also plays a role in the development of acidosis in patients experiencing shock. On rare occasions lactate acidosis may complicate diabetes treatment with biguanides, and also occurs in severe deficiency of vitamin B_1 (thiamine).

Following insulin treatment, ketone bodies are metabolized to hydrogencarbonate (bicarbonate); which is why the patient is usually not given alkaline infusions – this could consequently lead to alkalaemia.

25.7.1.2. Hypoglycaemic Coma

Hypoglycaemic coma is usually caused by overdosing on insulin (i.e. the administration of an excess dose, insufficient supply of carbohydrates or unusual physical strain with a normal dose of insulin). The higher the rate of glucose decline, the more severe the signs are. Sweating is induced by sympathetic nerve activation and is followed by the loss of consciousness or even coma – CNS cells need glucose in order to work. The condition is treated by means of an oral or i.v. (if the patient is unconscious) supply of glucose.

In rare cases, severe hypoglycaemia may be the result of overdosing on oral antidiabetic drugs from the group of sulphonylurea derivatives.

25.7.2. Late (Chronic) Complications of Diabetes

Late complications endanger both type 1 and 2 diabetics. They develop slowly, over decades, manifesting no signs to begin with. As mentioned in the introduction, 20 – 30% of newly diagnosed patients have already developed late complications. As the prevention of late complications is of paramount importance, it is clear that the earliest possible diagnosis of diabetes is desirable.

Late complications involve primarily blood vessels but also the eye lens (cataract); blood vessel complications are distinguished into two main types.

25.7.2.1. Diabetic Macroangiopathy

This complication involves large arteries and the disease does not differ from atherosclerosis in people who do not have diabetes. The only difference is that diabetics are about three times more endangered than people without diabetes. The patient is at risk of developing all the known complications of atherosclerosis: different forms of ischaemic heart disease including myocardial infarction, cerebrovascular accident and lower extremity arterial disease complications.

Oxidative stress is the main factor for aetiopathogenesis of the atherosclerosis; hyperglycaemia in diabetics is accompanied by overproduction of free radicals, glycation of LDLs facilitates their oxidation with the following inability to bind to LDL receptors.

25.7.2.2. Diabetic Microangiopathy

This complication is typical of diabetes. This affects the small arterioles – their wall grows thicker, leading to tissue oxygenation disorders. The most serious condition involves the eye ground blood vessels (i.e. diabetic retinopathy even leading to blindness), glomeruli (i.e. diabetic nephropathy with ultimate renal failure), and lower extremity blood vessels with consequent gangrenes. The involvement of vessels supplying blood to the nerves contributes to the development of diabetic neuropathy. At least 40% of patients are included in chronic dialysis programmes due to renal failure caused by diabetic nephropathy; diabetic retinopathy is the most common cause of blindness in people of working age.

The aetiopathogenesis of this complication includes glycation of proteins with a long biological half-life and subsequent oxidation of the compounds produced. The result is the production of highly reactive aldehydes called *advanced glycation end-products (AGEs)*. AGEs react with other proteins and the glucose forms bridges between protein molecules. Cross-linked proteins stimulate the production of cytokines with inflammatory and proliferative effects; the vessel wall character and thickness change due to their activity, and diabetic microangiopathy develops.

The aetiopathogenesis of the process clearly shows that the most effective prevention means the best possible compensation of the diabetic to avoid excessive glycation and overproduction of free radicals.

25.8. Diabetes Management Options

These paragraphs do not strive to substitute the internal medicine textbook; they rather provide a brief outline of options for managing diabetes, emphasizing each drug's mechanism of action.

25.8.1. Diet

Diabetics at the initial stages of DM2 have to live on a diet with a certain restriction to carbohydrates. Weight-loss diet in patients with central obesity considerably increases sensitivity to insulin through the reduced cytokine overproduction by the fatty tissue. Increased physical exercise is a matter of course in the therapy.

Food containing carbohydrates can be divided by their glycaemic index reflecting the glucose growth rate and the time of glucose level persistence following consumption of the food. Glucose levels are measured within two hours from eating food containing 50 g of carbohydrates (in any form) and administering 50 g of *per os* glucose. The ratio of areas under the curve for the tested food relative to the pure glucose solution in percent equates to the glycaemic index. Foodstuffs with a high glycaemic index (fruit, honey, white bread) cause a rapid increase of glycaemia and increase the secretion of insulin, tissue resistance to insulin and the triacylglycerol concentration. By contrast, low-glycaemic index foodstuffs (rice, brown bread) do not have the metabolic effects listed above and seem to be more appropriate for a diabetic diet.

25.8.2. Pharmacotherapy

Many drugs with different indications and different mechanisms of action are available.

25.8.2.1. Insulin

Human insulin produced using biotechnology methods is the only form of insulin used today. It is indicated to patients with DM1 and patients with more advanced stages and complications of DM2. Many preparations differing by the onset rapidity and duration of action are available. One common method of administration is subcutaneous, often using insulin pens or insulin pumps, if necessary.

A preparation for inhalation is also available but is not used in practice due to its high cost and many limitations.

25.8.2.2. Insulin Analogues

This refers to an altered form of insulin whose molecules differ from human insulin by the amino acid sequence. The changed insulin structure leads to a changed rapidity and duration of insulin action. Short-acting insulin analogues last for a shorter period of time and have a more rapid onset than regular human insulin. They are not only used for injections but also for insulin pumps. The benefit of long-acting analogues is their long effect (up to 24 hours).

25.8.2.3. Biguanides

Biguanides are used for all patients with DM2 who do not have contraindications and tolerate them. They increase insulin sensitivity of the liver and muscle tissue, and damp gluconeogenesis and glucose absorption in the small intestine. They can induce lactate acidosis through the stimulation of glycolysis. Metformin is the main type in this group.

25.8.2.4. Sulphonylurea Derivatives

Sulphonylurea derivatives belong to secretagogues – they increase the secretion of insulin from β-cells in the islets of Langerhans. The basic mechanism of action is the blockade of ATP-sensitive potassium channels leading to cell membrane depolarization. If overdosed, they may induce hypoglycaemia, increase appetite and sometimes lead to increased weight. Glipizide and glimepiride are examples of sulphonylurea derivatives used in practice.

25.8.2.5. Inhibition of glucose production from starch in the small intestine

The main type is acarbose, an intestinal α-glucosidase inhibitor. This enzyme continues the action of α-amylase and finishes starch break-down into glucose. Acarbose therefore slows down the production of glucose from polysaccharides in the food and reduces postprandial hyperglycaemia.

25.8.2.6. Glitazones

Glitazones (thiazolidinediones) through PPAR-γ nuclear receptors activate the transcription of genes responsible for metabolic effects of insulin (i.e. reduce insulin resistance). Only pioglitazone is used.

25.8.2.7. Glinides

Glinides as well as sulphonylurea derivatives block the ATP-sensitive potassium channel in β-cell membranes. They act quickly, so are ideal to be used together with food to compensate postprandial hyperglycaemia. Repaglinide and nateglinide are examples of these substances.

25.8.2.8. Incretin-Modulating Drugs

Incretins are peptides released from the digestive tract, which facilitate maintaining glucose homeostasis. The main incretin hormones are glucagon-like peptide-1 (GLP-1) and glucose-dependent insulinotropic polypeptide (GIP). GLP-1 stimulates insulin secretion in response to eating food, inhibits glucagon secretion and slows down the emptying of stomach. This leads to a marked decrease in postprandial glycaemia fluctuations.

Synthetic GLP-1 analogues such as exenatide are used in practice. Another possible way of increasing the GLP-1 level is to slow down its elimination by the dipeptidyl peptidase-4 (DPP4) enzyme. The benefit of DPP4 inhibitors (gliptins) consists of their oral administration while GLP-1 analogues have to be injected. Examples of gliptins are sitagliptin and vildagliptin.

25.8.2.9. Biological therapy of type 1 diabetes mellitus

The aim of this treatment is to slow down the autoimmune process and thus also β-cell death in patients with type 1 diabetes mellitus. There are two monoclonal antibodies in the phase of clinical trials: anti-CD3 antibodies (modulate T-lymphocyte response) and anti CD20 antibodies (influence β-lymphocyte maturation and thus reduce presentation of antigens to T-lymphocytes). Certainly it would be best to influence the autoimmune process as soon as possible, when diabetic symptoms are still not present; nevertheless, it is still a long way for scientific research teams.

25.8.3. Islet Cell Transplantation

In general, transplantation options include the pancreas itself (in rare cases), combined transplantation of the kidney and pancreas in patients with diabetes and renal failure, or islet cell transplantation. This is preceded by complex separation, purification and multiplication of these cells in the tissue culture. The transplantation itself is simpler – the cells are injected to the portal vein of the liver, so no operation is required.

In an experiment, laboratory animals received an intraperitoneal implant of β-cells encapsulated in a polymer allowing the penetration of small (glucose) and medium-sized molecules (insulin), but preventing the entry of antibodies and immunocompetent cells. Experimentally induced diabetes mellitus was thereby compensated for many months.

25.9. Stress-Induced Hyperglycaemia

Insulin is the only hormone that lowers blood glucose. Contrary to that, there are many hormones with the opposite, hyperglycaemic effects. When these hormones prevail over insulin, hyperglycaemia occurs. Hyperglycaemia can be seen in hyperthyroidism or Cushing's syndrome, although hyperglycaemia induced by stress hormone overproduction is most common.

A stressed body requires the abundance of readily usable energy for muscle and other cells, i.e. the abundance of glucose. This is ensured by catecholamines (adrenalin), which supports glycogen breaking down in the liver, and glucocorticoids (cortisol), which also stimulate gluconeogenesis. Every patient in a severe condition (patients after undergoing surgery, sustaining an injury, in intensive care units, after myocardial infarction or ictus, with severe inflammatory disease) is stressed. Therefore, hyperglycaemia can be expected in all of these patients. Glucose tolerance tests should be performed at least 6 weeks once the acute condition has subsided.

A positive effect of blood glucose control aimed at reaching a glucose level of 6 - 10 mmol/L has been proven in patients in a severe condition with stress-induced hyperglycaemia. Efforts to reach lower glucose levels lead to an increased occurrence of hypoglycaemia, which increases patient mortality rates.

25.10. Causes of Hypoglycaemia

In conclusion and for the sake of completeness, there is a chapter which is only partially relevant to diabetes; the chapter deals with potential causes of hypoglycaemia. The most common causes of this condition include:

Prolonged fasting, particularly in people with reduced glycogen reserve in the liver; the condition often occurs in newborn infants (premature infants in particular) and cirrhotic patients, of course.

Insulin overdosing, either due to an excessive dose relative to the carbohydrate intake, or due to unusual physical strain which lowers the requirement for insulin.

Oral antidiabetic drug overdosing, namely sulphonylurea derivatives.

Alcohol intoxication – reduced form of the NADH coenzyme produced by alcohol metabolism inhibits gluconeogenesis, which ultimately leads to hypoglycaemia.

Nesidioma of the pancreas (= insulinoma) is an insulin-producing tumour of β-cells. The release of insulin may be sudden and paroxysmal, which manifests itself by sudden hypoglycaemia. High levels of insulin and C-peptide prevail in laboratory test results.

Hereditary disease such as some types of glycogenosis, monosaccharide metabolism disorders, gluconeogenesis disorders and fatty acid oxidation disorders. Patients with type 1 glycogenosis have severe fasting hypoglycaemia, because they are unable to break down glycogen in the liver due to a glucose-6-phosphatase deficiency. Contrary to that, patients with galactosaemia and fructose intolerance experience hypoglycaemia after eating the relevant sugar (after ingesting milk containing lactose in the former case, and after ingesting saccharose containing fructose in the latter). The most common cause of galactosaemia is a galactose-1-phosphate uridyltransferase deficiency; accumulating galactose-1-phosphate inhibits glycogenolysis, among others glycogen phosphorylase and glucose-6-phosphatase enzymes. Hereditary fructose intolerance features fructose-1-phosphate accumulation in the body (for a fructose-1-phosphate aldolase deficiency); this reduces the activation of hepatic phosphorylase and gluconeogenesis at the level of fructose-1,6-bisphosphate.

CHAPTER 26

26. Cardiac Markers

26.1. Cardiac Markers – Preparation

Authors: MUDr. Věra Ploticová; RNDr. Zdeněk Švagera, Ph.D.

Reviewer: MUDr. Roman Cibulka, Ph.D.

26.1.1. Ischaemic Heart Disease

Ischaemic heart disease (IHD) is an acute or chronic disease of the heart muscle caused by insufficient blood supply in coronary artery disease. The involvement of coronary arteries can be organic (narrowing), functional (spasm) or thrombotic (mural or obturating thrombus). Combinations of all these components can be seen in practice.

IHD is the most common cause of death in the Czech Republic.

26.1.1.1. Aetiology and Pathogenesis of IHD

The most common cause of IHD is atherosclerosis of the coronary arteries. At the beginning, atherosclerotic plaque build-up is usually associated with an enlarged arterial diameter (arterial remodelling), and the disease may be present without clinical symptoms for decades.

Only when the plaque narrows the arterial lumen to more than 70 %, does stenosis start to obstruct increased coronary flow under physical strain. This is how ischaemia is formed downstream the obstruction with its clinical symptoms in the form of exertional angina pectoris (chest pain on physical exertion). Another mechanism for initiating ischaemia is a spasm of the smooth muscle of the arterial wall. It is based on endothelial dysfunction induced by stimuli such as cold, tobacco smoke, physical and mental strain.

Should the plaque rupture, the thrombogenic (subendothelial) structures of the vessel wall become exposed, which leads to platelet activation and blood clotting which may even cause a sudden closure of the coronary artery. Bleeding in the plaque may also lead to a sudden arterial narrowing and a dramatic decrease in the blood flow. These conditions cause acute coronary syndromes (see chapter Acute Coronary Syndromes).

Atherosclerosis may affect all of the three major arteries evenly, however prevailing involvement of one of them is more common.

Heart arteries most commonly affected by atherosclerosis:

- **RIA**, ramus interventricularis anterior (ventralis) arteriae coronariae cordis sinistrae (ACS);
- **RCx**, ramus circumflexus arteriae coronariae cordis sinistrae;
- **ACD**, arteria coronaria cordis dextra.

Coronary atherosclerosis may build up evenly (concentrically) around the arterial perimeter, followed by the residual lumen is in the middle of the artery. However, the plaque may also be asymmetric (eccentric), outside the centre. The grade of atherosclerosis is assessed depending on the percentage of arterial lumen narrowing by sclerotic plaques, see Table 29p.1, measured by coronarography, intravascular ultrasound (IVUS) or found in the autopsy report.

Arterial Lumen Narrowing	Grade of Coronary Atherosclerosis
25 %	Grade I
50 %	Grade II

75 %	Grade III
over 75 %	Grade IV

Table 26.1. Grading atherosclerosis

Coronary artery diseases other than atherosclerosis are rare (e.g. Wegener's granulomatosis, syphilis).

The risk *factors* for coronary atherosclerosis are the same as the risks for atherosclerosis in general. They do not exactly explain the cause of atherosclerosis however do help estimate the risk of complications associated with atherosclerosis (cardiovascular risk). Therapy is used for prevention.

The most common cause of IHD is atherosclerosis of the coronary arteries.

26.1.1.2. Atherosclerosis

Atherosclerosis is a degenerative involvement of arterial walls characterized by the accumulation of lipids, inflammatory infiltration of the vessel wall and fibrous tissue proliferation. After birth, human arteries are smooth, elastic and not very resistant to the blood flow.

Aetiology and Pathogenesis of Atherosclerosis

For many years, atherosclerosis was regarded as a mechanical process typical of lipid accumulation in the vessel wall, and calcium incrustation at a later phase. Multifactorial aetiology of atherosclerosis is accepted at present. Risk factors have been shown to contribute to atherosclerosis. Literature mentions hundreds of them, however about ten are important in practice. Table 29p.2 lists the most commonly mentioned risk factors.

The first stages of atherosclerosis are known to be caused by increased penetration and infiltration of atherogenic lipoproteins and inflammatory cells from the blood through the endothelium, followed by their accumulation in the subendothelial space. At later stages, fibroproductive processes join the process of atherogenesis in reaction to depositing lipids and the inflammatory infiltration of the vessel wall. Endothelial damage, inflammatory mechanisms and atherogenic lipoproteins are involved to varying degrees in the development of these lesions. All of these factors are in complicated interactions and potentiate each other.

Uncontrollable Risk Factors	**Controllable Risk Factors**
Age (men over 45, postmenopausal women or over 55)	Smoking
Male gender	Dyslipidaemia, increased total plasma cholesterol (LDL--cholesterol in particular), increased triacylglycerols, decreased HDL-cholesterol)
Family history of premature IHD (manifested in men under 55 and women under 65), or other clinical manifestations in first-degree relatives (parents, siblings and children)	Hypertension
	Glucose metabolism disorders (diabetes mellitus, boundary fasting blood glucose, impaired glucose tolerance)
	Obesity (android type)

Table 26.2. Risk factors of atherosclerosis

26.1.1.3. Clinical Presentation of IHD

The clinical forms of IHD are shown in Table 29p.3 below.

Acute Forms	**Chronic Forms**
Unstable angina pectoris	Asymptomatic IHD
Myocardial infarction	Angina pectoris (exertional, mixed, variant)

Sudden cardiac death	Condition after myocardial infarction
	Dysrhythmic IHD
	Chronic heart failure

Table 26.3. Clinical forms of ischaemic heart disease

26.1.1.4. Acute Coronary Syndromes

Acute coronary syndrome is the clinical manifestation of a critical phase in IHD development, when the imbalance between the oxygen supply and requirements of the myocardium for the supply and removal of metabolic waste products becomes manifested.

The most common pathophysiological background for these conditions is an atherosclerotic plaque rupture and a blood clot formation in the coronary artery with potential presence of coronary spasm. The blood clot leads to a narrowing or closure of the coronary artery with varying degrees of seriousness, and embolism in its peripheral branches.

Acute coronary syndromes:

- ST-segment elevation myocardial infarction (STEMI);
- Non-ST-segment elevation myocardial infarction (NSTEMI);
- Unstable angina pectoris.

Clinical presentation

Sudden blockage or critical narrowing of the coronary artery causes ischaemia, manifested in most cases by chest pain while at rest. In typical cases, the pain is constricting or burning, affects a wide area, may radiate to the neck, lower jaw or the upper left extremity, and lasts more than 20 minutes. Patients may sometimes feel back pain or epigastric pain. The pain is usually accompanied by dyspnoea, sweating and nausea, and the patient has to stop working and wait for relief. In a small part of patients (about 20 %), myocardial infarction manifests itself only as dyspnoea, not pain. There can be a rise in blood pressure and heart failure may develop.

Patients with acute coronary syndrome have alternating pain attacks, are restless, anxious, have pale, sweaty and cyanotic skin, while the heart rate may be increased, normal or decreased.

Myocardial Infarction

According to the definition of the European Society of Cardiology of 2007, the term myocardial infarction is reserved for conditions with evidence of myocardial necrosis in a clinical setting consistent with myocardial ischaemia. Necrosis assumes a coagulative character, caused by high protein content. Around 20 hours from the start of ischaemia, the colour of the tissue turns to yellowish, and water loss shrinks the necrotic tissue volume. Fully developed necrosis is evident after 3 days.

Diagnostic criteria of myocardial infarction

Evidence of a rise and/or drop in laboratory markers of necrosis (troponin in the first place) with at least one value above the 99th percentile of the upper limit of the normal range with evidence of myocardial ischaemia with at least one of the symptoms of ischaemia listed below:

- Laboratory tests help establish diagnosis of acute MI;
- Symptoms of ischaemia;
- ECG alterations are indicative of new ischaemia (ST-T segment alterations or a new left bundle branch block);
- Development of pathological Q waves on the ECG;
- New loss of viable myocardium or new local disorders of myocardial kinetics evidenced by an imaging method.

Sudden and unexpected cardiac death including cardiac arrest, frequently with symptoms indicative of myocardial ischaemia, very often accompanied by new ST segment elevations. Sometimes a fresh thrombus is proved by a coronarography and/or section, when death occurred before blood could be taken and before a rise in laboratory markers

of necrosis.

In patients with percutaneous coronary intervention and normal baseline values of laboratory markers of necrosis, their increase over the 99th percentile of the upper limit of the normal range is a sign of periprocedural myocardial necrosis. By consensus, **periprocedural myocardial infarction** is defined as a rise in laboratory markers of necrosis more than three times the upper limit of the normal range with the presence of one of the signs listed below:

- Newly developed pathological Q waves;
- Left bundle branch block;
- Development of a new bypass or coronary artery closure proved by angiography;
- Evidence of a new loss of viable myocardium using an imaging method.

Myocardial infarction is currently classified in clinical practice into ST-segment elevation and non-ST-segment elevation myocardial infarction.

ST-elevation (STEMI) is indicative of necrosis of all myocardial wall layers; so this type of infarction is referred to as transmural MI. As a Q wave development can be seen on the ECG, also Q-infarction is used.

The extent of non-ST elevation MI is usually smaller, it affects subendocardial myocardial levels, and there is no ST--segment elevation or pathological Q-wave development. This is why it is called non-Q infarction or NSTEMI.

This classification is especially used on patient hospitalisation for rapid patient stratification. While STEMI patients go right to the operation theatre (without waiting for further test results), NSTEMI patients are examined in more detail to decide (taking other test results into account) whether or not myocardial infarction is involved.

Feature	**High probability** *Whichever of following:*	**Intermediate probability** *Absence of highly probable features and presence of whichever of following:*	**Low probability** *Absence of highly or intermediate probable features, but patient can show:*
Anamnesis	Thorax or left arm pain or dyscomfort as a main sign of before known angina pectoris Known anamnesis of ICHS including myocardial infarction	Thorax or left arm pain or discomfort as a main sign Age ≥ 70 years Masculine gender Diabetes mellitus	Probable ischemic signs and absence of whichever intermediate probable features Ancient cocain abuse
Examination	Transient murmur corresponding with MR, hypotension, excessive sweating, lungs oedema or crackles	Extracardial vascular disease	Thorax discomfort reproduced on palpation
ECG	New or most probably new transient ST part denivelation (≥ 1 mm) or T wave inversion in multiple precordial leads	Fix Q waves ST segment depression (0.5 – 1.0 mm) or T wave inversion > 1.0 mm	T wave flatteningor inversion < 1 mm in leads with dominant T waves Normal ECG
Cardiomarkers	Increase of cardial TnI, TnT concentration or CK-MB concentration	Normal	Normal

Figure 26.1. Marks and signs and probability of ACS

Unstable Angina Pectoris

Unstable angina pectoris is not a homogeneous nosological entity, but rather a syndrome with a name describing the acute condition.

Clinical definition of unstable angina pectoris – criteria:

- Newly developed angina pectoris; or
- Sudden worsening of existing angina pectoris; or
- Development of stenocardia in the subacute MI phase.

Worsening means more frequent anginous attacks (frequent stenocardia), more intense pain or change in the nature of attacks, stenocardia while at rest or protracted stenocardia. The difference from non-ST elevation MI consists in the absence of laboratory markers of myocardial necrosis. The ECG only shows signs indicative of subendocardial ischaemia.

Chronic Forms of Ischaemic Heart Disease

Exertional (Stable) Angina Pectoris

The typical clinical symptom is pain behind the breastbone in situations connected to increased oxygen consumption in the myocardium. Stenocardia usually lasts less than 20 minutes, radiates to the left arm, neck and back, and is usually not accompanied by vegetative symptoms. There is also a painless form of stable angina pectoris, most often manifested by dyspnoea.

Precipitating moments include physical exertion, heavy food or emotions. Discomfort is more pronounced in cold weather, or in the morning in some patients (quite typically when coming out from a warm environment to the freezing air). Discontinued exertion or nitrate administration lead to relief within minutes.

Variant Angina Pectoris (Prinzmetal's Vasospastic Angina)

Variant angina pectoris is a nosological entity, the pathophysiological background of which is spasms of epicardial coronary arteries caused by endothelial dysfunction.

The typical clinical symptom is stenocardia while at rest accompanied by transient ST-segment elevations and heart rhythm disorders. The spasms can be induced during coronarography examination by provocation tests. If no spasm is present on physical exertion, the patient is usually asymptomatic. Variant angina typically affects women at an age between 40–50 years with some risk factors of ischaemic heart disease.

Cardiac Syndrome X

This name includes a heterogeneous group of patients with typical exertional stenocardia and evidence of ischaemia on a stress ECG or thallium myocardial scintigraphy, but without the presence of stenotic alterations or spasm on coronary arteries. Microvascular dysfunction at the level of prearterioles with the impossibility to increase coronary perfusion on exertion is being considered as one of causes.

26.1.2. Heart Failure

26.1.2.1. Acute Heart Failure

Acute heart failure is a sudden severe heart disorder or a sudden worsening of the heart's function where the heart is unable to pump blood from the venous circulation to the lungs, or from the lungs to the arterial circulation. This results in congestion of the blood in some organs and insufficient tissue oxygenation.

Aetiology and pathogenesis

Many cardiovascular diseases present themselves as acute heart failure.

Acute heart failure is a syndrome and should never be a diagnosis in itself!

Depending on the dominant clinical symptoms, heart failure may be right-sided or left-sided.

Acute Right-Sided Heart Failure

Volume Overload	Pressure Overload	Lower Contractility
Tricuspid regurgitation	Pulmonary embolism	Right ventricular infarction
Pulmonic regurgitation	Left-sided heart failure	Right ventricular cardiomyopathy
Defects which cause shunting	Pulmonary artery stenosis	Perioperative right heart injury
Acute ventricular septal defect	Mechanical ventilation, pulmonary hypertension, chronic obstructive pulmonary disease, pulmonary embolism	

Table 26.4. Most common causes of right-sided heart failure

Clinical presentation

Typical subjective symptoms of acute right-sided heart failure include dyspnoea, reduced tolerance to exercise, fatigue and chest pain.

Acute left-sided heart failure

Aetiology and pathogenesis

Acute left-sided heart failure occurs as a result of sudden injury to the left ventricle by mechanical influences, direct injury or impaired diastolic filling.

This failure leads to sudden left ventricular pulmonary hypertension, which transmits back to the left atrium and pulmonary circulation. This leads to acute pulmonary hypertension and the amount of blood in the lungs rises. If pulmonary capillary pressure exceeds colloid osmotic pressure in the plasma, liquid penetrates into the interstitium followed by to the alveoli. *This is why interstitial and intra-alveolar pulmonary oedemas are recognized.*

Mechanical Influences	Direct Injury	Impaired Diastolic Filling
Pressure (hypertension, aortic stenosis)	Ischaemic heart disease	Cardiac tamponade
Volume (valvular regurgitation)	Cardiomyopathy	Arrhythmia
	Myocarditis	Exudative pericarditis

Table 26.5. Causes of acute left ventricular failure

Clinical presentation

Typical symptoms of acute left ventricular failure include dyspnoea and pulmonary oedema. Pulmonary oedema is a symptom of acute left-sided heart failure. The patient is extremely dyspnoeic even at rest, tachypnoic, anxious, restless, pale, perspiring and coughs up frothy sputum.

Cardiac asthma *(asthma cardiale)* is a condition of severe dyspnoea usually occurring at night when it wakes up the patient, who is first urged to sit up, then stand up and breathe while employing accessory muscles.

26.1.2.2. Chronic Heart Failure

Chronic heart failure is a disorder of the heart's performance, whereby the minute cardiac output decreases despite sufficient filling of the ventricles, and the heart is unable to cover the metabolic requirements of peripheral tissues. Depending on whether congestion prevails in the pulmonary or systemic bloodstream, heart failure is sometimes classified as right-sided and left-sided. This does not indicate, however, which of the ventricles is affected more.

Aetiology and pathogenesis

Chronic heart failure develops as a result of systolic and/or diastolic myocardial dysfunction. The most common cause is IHD, less frequently cardiomyopathies, myocarditis and toxic myocardial injuries. Worsening factors include hypertension, some arrhythmias, anaemia, hyperthyroidism or the iatrogenic effects of some drugs (antidepressants).

The basic pathophysiological cause of chronic heart failure development is a disorder of systolic or diastolic function of the ventricles with a subsequent drop in tissue perfusion, which induces the activation of sympathetic nerves, the renin-angiotensin-aldosterone system and the secretion of endothelin and vasopressin.

The myocardium undergoes remodelling (presented as dilation at the ventricular level) which worsens contractility and relaxation of cardiomyocytes, and ligament restructuring and collagen proliferation in the interstitium. Disorders of endothelial functions and microvascular fibrosis occur in the vascular blood stream. The heart is no longer able to increase its performance on exertion despite the fact that the ventricular dilation allows the stroke volume to be kept with an even lower ejection fraction (EF is less than 45 %).

Clinical presentation

Dominating clinical symptoms and objective pronounced signs allowing for right-sided and left-sided heart failure to be distinguished.

Left-sided heart failure

Most signs and symptoms occur as a result of blood congesting in the lungs, namely exertional dyspnoea, paroxysmal nocturnal dyspnoea (cardiac asthma), and pulmonary oedema. Objective signs include wet inspiratory crackles not changing their character after coughing, and also a one-sided or two-sided effusion.

Right-sided heart failure

Most symptoms occur as a result of blood congesting before the right ventricle, increased filling of the jugular veins, positive hepatojugular reflux, hepatomegaly and peripheral oedemas. Oedemas first develop around the ankles, form during the day and release at night (nocturia). They signal a 3–5-litre increase in the volume of extracellular fluid. Extreme retention of fluids manifests itself as hydrothorax, ascites, anasarca, and hydropericardium.

26.2. Laboratory Diagnosis of Acute Myocardial Infarction and Heart Failure – Use of Cardiac Markers

Authors: MUDr. Věra Ploticová; RNDr. Zdeněk Švagera, Ph.D.

Reviewer: MUDr. Roman Cibulka, Ph.D.

26.2.1. Laboratory Assays Used in Diagnosing Ischaemic Heart Disease

Myocardial infarction was very scarce in the nineteenth century. Each case was seen as a rarity. However the situation has considerably changed over the years. Dr. James Herrick published a study about the effects of coronary artery blockage in 1912. That was the time when symptoms of cardiac infarction were first mentioned in addition to sudden death and angina pectoris. Several years later, Dr. Herold Pardee first described and characterised the ECG record of acute myocardial infarction. After the Second World War, myocardial infarction became one of the most common causes of adult mortality. This was when people got interested about this disease and its causes. The Framingham study was published in 1960 and described the major causes of this disease. Even in the 1980s, the first attack of acute myocardial infarction (AMI) was regarded as waiting for the second AMI attack, the second as waiting for the third attack, and the third attack usually ended by death. The development of laboratory diagnosis and intense therapy in cardiology have led to a marked decrease in AMI mortality.

In general, diagnosing acute coronary syndrome (ACS) is based on an assay for intracellular proteins released from the ischaemic focus in the myocardium to the blood stream. Finding a sufficiently sensitive as well as specific parameter for the myocardium, which could point to the ACS fast, specifically and sensitively, has always been problematic.

Different, though not sufficiently cardiospecific parameters were used in the past. A change in aspartate aminotransferase (AST) concentration in acute myocardial infarction was described in the 1950s. Consequently, lactate-dehydrogenase (LD), creatine phosphokinase (CK), myoglobin and LD and CK isoenzyme tests were first used. In the 1980s, the "gold standard" for diagnosing AMI was the assay for creatine phosphokinase cardiac isoenzyme activity (CK-MB), which had its limitations (inadequate sensitivity, loss of activity in the circulation) and is no more satisfactory for practical use today.

As these assays were not specific enough for diagnosing cardiac muscle injury, new markers were sought, and in 1990 attention turned to structural proteins of the cardiac muscle, cardiac troponins, which have become the „gold standard" of today, and are currently firmly established in all recommendations of societies of cardiology with regard to diagnosing acute coronary syndrome (ACS)

26.2.1.1. Diagnosis of Acute Myocardial Infarction

According to the definition of acute myocardial infarction (2009), acute (developing or recent) myocardial infarction is defined as a typical increase and/or decrease in biochemical markers of myocardial necrosis (cardiac **troponin T or I**, or CK-MB) with concurrent presence of at least one of the following criteria:

α) Clinical symptoms of ischaemia;

β) Development of pathological Q waves on the ECG;

χ) ECG alterations indicative of ischaemia (ST elevation or depression);

δ) Relation to coronary revascularization (PCI, bypass);

ε) Evidence of a new regional disorder of kinetics (echocardiogram), or imaging evidence of a new loss of viable myocardium.

This recommendation clearly shows the importance of troponin assays as markers of myocardial necrosis.

26.2.1.2. Cardiac Troponins

Troponins are structural proteins of striated muscles, which means they can be found both in skeletal muscles and the myocardium. Troponin T (TnT) provides the bond of troponins C and I to tropomyosin; troponin I (TnI) inhibits actinomyosin-ATPase, which depends on the amount of calcium bound to troponin C. Troponins with tropomyosin create the troponin-tropomyosin complex, which performs a regulatory and structural role in the contractile apparatus of striated muscles (see the animation) **Animation – contractions!!!**

http://translate.google.cz/translate?hl=cs&langpair=en|cs&u=http://highered.mcgraw-hill.com/sites/0072495855/student_view0/chapter10/animation__action_potentials_and_muscle_contraction.html

Methods of molecular genetics helped reveal that troponins T and I in the skeletal muscle and myocardium have different primary structures and can be distinguished by immunoassays. This fulfilled the requirement for a cardiospecific method allowing for only cardiac troponins T (cTnT) and I (cTnI) to be determined, thus distinguishing myocardial involvement from damage to the skeletal muscle.

	Troponin I (cTnI)	Troponin T (cTnT)
Origin	Myocardium	Myocardium
Molecular weight (Da)	23,876	37,000
Biological half-life (minutes)	<120	120
Rise after AMI (hours)	3-5	3-4
Maximum concentration in blood (hours)	12-30	12-75
Persistence in blood (days)	4-7	4-12
Practicability for ACS diagnosis	Comparable	Comparable
Efficiency in cardiovascular risk stratification	Comparable	Comparable
Biphasic waveform during successful and early reperfusion	Less distinct	Distinct
Practicability for reperfusion evaluation	Yes	Yes
Diagnosis of reinfarction	Limited use due to long persistence in the blood	Limited use due to long persistence in the blood
Elimination through the kidneys	Yes	Yes
Rise in patients with stage 5 chronic renal disease (according to the National Kidney Foundation)	Less pronounced	Pronounced
Multiple of the upper reference limit at maximum	300x	300x
Estimate of AMI focus size	Yes	Yes
Timing of sampling for diagnosing AMI (CSKB recommendation, 2008)	On admission, 6–9 hours, if negative then 12–24 hours	On admission, 6–9 hours, if negative then 12–24 hours

Table 26.6. Summarizes some properties of cardiac troponins T and I

Advantages of Determining Cardiac Troponins (cTn) in ACS (AMI) Diagnosis

- Very high specificity for the myocardium – a rise in cTn concentrations always relates to myocardial injury;
- High diagnostic sensitivity (ability of the method to detect the disease);
- Multifold increase in concentrations in acute myocardial infarction;
- High sensitivity allows for even the slightest myocardial damage to be detected;
- Possible to estimate the focus size from the rise in concentration;
- Late diagnosis of MI due to the long persistence of increased values.

Disadvantages of Determining Cardiac Troponins in ACS (AMI) Diagnosis

A relatively slow increase after AMI – 3–4 hours on average (this drawback will be eliminated in the near future by the introduction of new, ultrasensitive methods for determining troponin).

Limited use in diagnosing reinfarction due to long persistence in the blood (this drawback should be eliminated by introducing ultrasensitive methods).

Different TnI cut-offs between laboratories – there are multiple manufacturers of TnI test kits using different antibodies, and therefore different cut-off values. TnT test kits are supplied by only one manufacturer, which guarantees uniform cut-off values and comparability between laboratories.

Literature describes the existence of a disease with an increased serum troponin level due to non-cardiac causes (renal failure, sarcoidosis, amyloidosis, critically ill patients with respiratory failure or diabetes, acute neurological diseases, burns, toxins, hypothyroidism, etc.). The increase in troponin in such diseases has different dynamics than those typical of AMI. In addition to this, the patient's clinical condition and other examination methods do not confirm AMI diagnosis.

Clinical Importance of Cardiac Troponin Assays

Cardiac troponins have been used in clinical diagnosis as organ-specific (cardiospecific) markers of ischaemic myocardial injury in acute coronary syndromes (ACS) and also myocardial injury with different aetiology and pathogenesis.

List of causes which may increase serum troponin levels:

- The most common cause is ischaemic necrosis due to limited or blocked perfusion by mural or occlusive thrombus in a coronary artery;
- However, elevated cTn concentration in the blood may also be a symptom of ischaemic and/or non-ischaemic myocardial injury not caused by atherothrombosis such as mechanical disruptions of cardiomyocytes due to myocardial injury, the cumulative effect of potentially cardiotoxic drugs, action of bacterial toxins, or as a result of a neurohumoral reaction;
- Other causes include cardiac muscle inflammations, circulatory collapse of varied aetiology (e.g. following polytraumas, sepsis with multi-organ failure, severe burns, etc.), when systemic hypotension may lead to myocardial ischaemia, a rise in troponin, and renal failure with slightly increased cTn (cTnT in particular) levels not indicative of infarction. Intoxications should also be mentioned (carbon monoxide poisoning is quite frequent: the reduced oxygen-carrying capacity of haemoglobin may lead to myocardial hypoxia and a rise in troponin);
- Doubtless and sometimes prognostically serious myocardial injuries with increased cTn concentrations in the blood occur for iatrogenic causes, during diagnostic or therapeutic procedures associated with coronary blood stream catheterisation and all cardiosurgeries;
- With corresponding clinical symptomatology, a rise in cTn concentration is considered a result of irreversible (definitive) myocardial necrosis, and should always be assessed in relation to diagnosing the disease and also in order to stratify the risk of further development of cardiovascular diseases;
- Long and permanently increased levels of cTn, cTnT in particular, are frequent (without clinical symptomatology of the ACS) in patients in the terminal stage of renal failure, and correlate significantly with the risk of their overall mortality; they should be continually monitored for cTn levels.

Other Tests for Diagnosing AMI

AMI diagnosis also employs other auxiliary tests which may supplement information obtained from cardiac troponin assays. These tests include the myoglobin assay and the creatine phosphokinase cardiac isoenzyme mass (CK-MB mass).

Parameter	Myoglobin	CK-MB mass
Origin	Striated muscle – skeletal muscle and myocardium	Myocardium-specific isoenzyme, below 5 % of total CK in skeletal muscle
Molecular weight (Da)	18,000	85,000
Biological half-life (minutes)	10 – 20	11 – 12
Rise after AMI (hours)	1 – 2	3 – 10
Maximum concentration in the blood (hours)	4 – 10	12
Persistence in the blood (hours)	24	48 – 72
Practicability for diagnosing ACS	For early phase after AMI	Where cTn cannot be determined
Practicability for reperfusion evaluation	Yes	Yes
Diagnosis of reinfarction	Yes	Yes

Elimination through the kidneys	Yes	No
Rise in patients with stage 5 chronic renal disease (according to the National Kidney Foundation)	Pronounced	Less pronounced
Estimate of AMI focus size	---	---
Timing of sampling for AMI diagnosis (CSKB recommendation, 2008)	2–6 hours from onset of disease, pointless after 12 hours	On admission, 6–9 hours, if negative then 12–24 hours

Table 26.7. Summarizes some properties of myoglobin and CK-MB mass

Clinical Importance of Myoglobin and CK-MB Mass Assays

Myoglobin is an early, yet organ non-specific marker of AMI. Myoglobin levels increase as a result of muscle injury of varied aetiology, when myoglobin is released into the blood. This parameter can be used at a very early stage after displaying ACS symptoms (3–6 hours) in order to rule out ACS diagnosis. Serum myoglobin levels have a very high negative predictive value (the probability that a person with the negative test is actually healthy – does not have the suspected disease). In practice this means that if this parameter is elevated, AMI cannot be ruled out (patient hospitalisation). Conversely, increased myoglobin does not mean that AMI is present since the elevation may be due to muscle damage with another aetiology.

The CK-MB mass assay is currently an alternative assay for the AMI diagnosis where the cTn assay is not available. This cardiac marker is still recommended, and is used to detect the myocardium (reinfarction) extent when high cTn levels persist in the blood. Elevated CK and CK-MB may also occur in non-cardiac conditions where AMI is not present. Examples include muscular dystrophy, surgery, trauma, intramuscular injections, collagenosis, hypothyroidism, hyperthermia, spasms, enormous physical strain, etc.

26.2.1.3. Less Common Cardiac Markers

Other parameters especially useful for detecting very early cardiac damage are being sought. One big goal in today's cardiology is to identify a patient with vulnerable plaque(s) in the coronary arteries prior to rupturing, and early identification of LV remodelling and dysfunction.

Present knowledge and basic characteristics of some new biomarkers in acute cardiology are potential candidates for clinical use for risk stratification of patients with ACS, early detection of left ventricular (LV) dysfunction development and possible monitoring of treatment.

These parameters can thus be divided based on the cause of their origin into:

- Markers of the vessel wall inflammation (interleukin 6, interleukin 10, highly sensitive CRP, lipoprotein-associated phospoholipase A2, etc.);
- Markers of destabilization or rupture of atheromatous plaque (soluble CD40 ligand, myeloperoxidases, etc.);
- Myocardial ischaemia markers (ischaemia modified albumin, free fatty acids, etc.);
- Cardiomyocyte necrosis markers (binding protein for fatty acids, glycogen phosphorylase).

26.2.2. Laboratory Assays Used for Heart Failure

Laboratory assays are also used to diagnose heart failure. Just like troponins reveal information about damaged cardiomyocyte (damaged structure), natriuretic peptides reveal information about overloaded myocardial cells (malfunction). In addition to standard patient history data, physical examination, cardiac and pulmonary X-ray, and above all echocardiography, natriuretic peptides have become an integral part of heart failure diagnosis.

26.2.2.1. Natriuretic peptides

Natriuretic peptides are hormones of a protein nature excreted mainly through the myocardium, also through the endothelium and kidneys. They have a protective function in the body – protecting the cardiovascular system against

pressure and volume overloading (pressure and volume overloading of the myocardium lead to increased BNP levels).

The main target organ of natriuretic peptides is the distal renal tubule where they inhibit reverse sodium transport (induce natriuresis, increased diuresis and the renin-angiotensin-aldosterone system inhibition). The history of natriuretic peptides dates back to 1981 when the extract from the atria was portrayed to cause natriuresis and vasodilatation. The chemical structure of atrial natriuretic peptide (ANP) was described in 1984, with the brain natriuretic peptide (BNP) being isolated from the swine brain in 1988, and with the central nervous system natriuretic peptide (C-natriuretic peptide, CNP) being isolated in 1990.

ANP is stored in the form of secretory granules in the cytosol of atrial cardiomyocytes; its release is immediate, in the order of tens of seconds to minutes.

BNP is not stored correspondingly, and an increase in its concentration in the blood is determined by the fast activation of its synthesis in ventricular cardiocytes. A rise in BNP concentration in the blood is very fast – in the order of minutes to tens of minutes.The N-terminal part of the propeptide of BNP (NT-proBNP), which is also present in the blood, is cleaved off in the metabolic activation of pre-prohormone proBNP with a concurrent release of BNP.

Routine tests exist for the active hormone BNP and the NT-proBNP. Advantages of the BNP or NT-proBNP assays have not been clearly defined. According to international recommendations published so far, the BNP and NTproBNP assay provide equal clinical information both in patients with dominating clinical symptoms of dyspnoea to differentiate between cardiac and extra-cardiac aetiology, and patients with suspected ACS for the stratification of risk of further cardiovascular complications. An advantage of NT-proBNP as compared with BNP can be longer stability in the blood against BNP.

The Clinical Importance of Natriuretic Peptide Assays

In clinical practice, natriuretic peptide assays, BNP and NT-proBNP assays in particular, are used as indicators of the heart's failure to work as a pump (regardless of etiology) and as a sign of activation of one of the basic neurohumoral systems regulating cardiac performance. The main areas where natriuretic peptides are used are:

- Differential diagnosis of acute dyspnoea with an unclear proportion of cardiac failure;
- Stratification of the risk of cardiovascular complications in patients with suspected ACS.

Differential Diagnosis of Acute Dyspnoea with an Unclear Proportion of Heart Failure

Samples are taken from patients with dominating dyspnoea immediately on being admitted. The result of the examination helps establish diagnosis of cardiac and extra-cardiac aetiology of dyspnoea (see the diagram below).

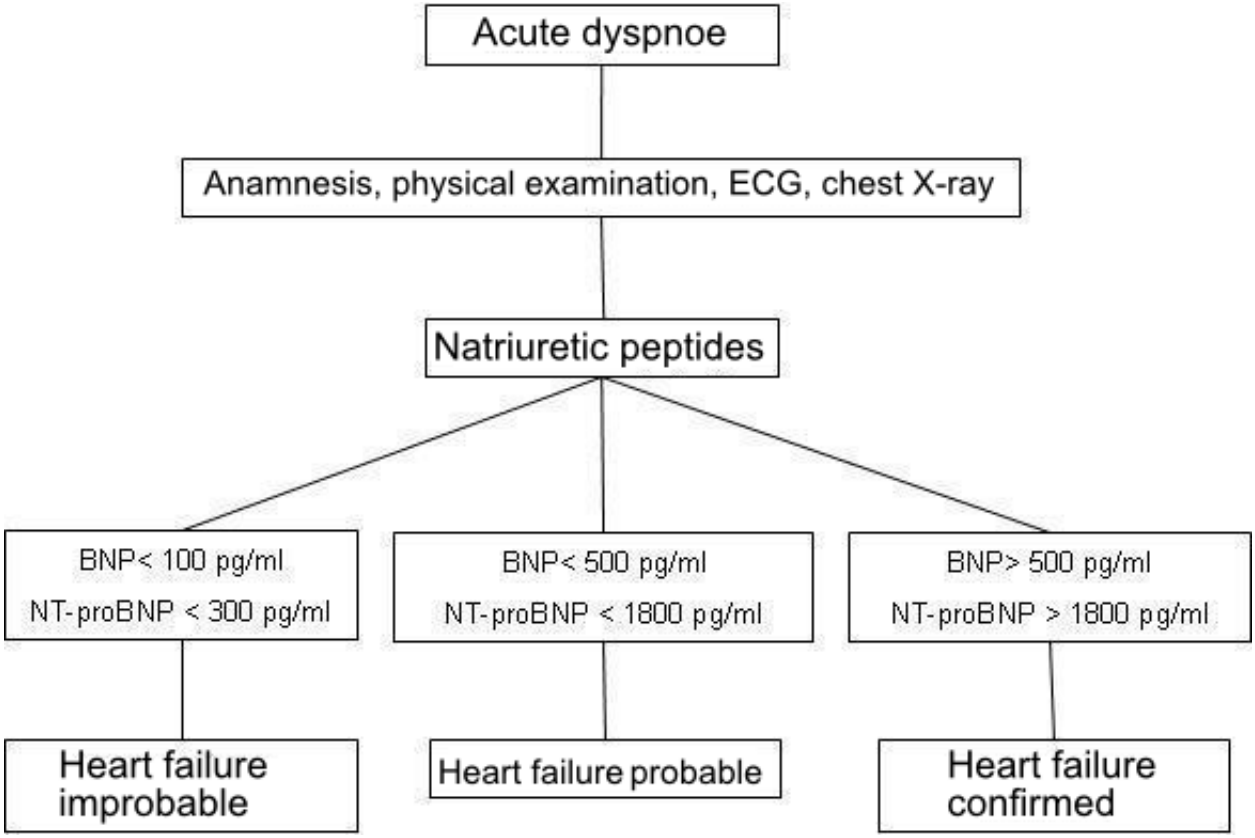


Figure 26.2. Acute dyspnoea and natriuretic peptides levels – scheme

The diagram shows that low natriuretic peptide levels indicate dyspnoea of non-cardiac origin, and pulmonary or other disease should be ruled out. Medium increased concentrations point to probable heart failure, however septic shock, pulmonary hypertension, ARDS, acute embolism or renal failure should be ruled out. Heart failure is very probable if natriuretic peptide concentrations are high; however, conditions such as septic shock, where high BNP levels are also detected, have to be ruled out.

Stratification of the Risk of Other Cardiovascular Complications in Patients with Suspected ACS

The recommended time of sample collection in patients with clinical symptoms of heart failure (primarily patients with suspected or diagnosed ACS) is 2 to 3 days upon being admitted , and a follow-up test after 1 to 2 weeks; the test results help stratify the risk of developing cardiac dysfunction and/or heart failure.

26.3. References

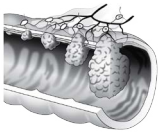
1. Racek, J. (2006). *Klinická biochemie* (2nd. ed.): Praha.,Galén
2. Diagnostika a léčba akutního infarktu myokardu s elevacemi ST. Cor Vasa 2009;51(10)
3. Tichý, M., Friedecký., B., Palička, V. ,Horáček, J., Jebavý, L., Pudil, R. *Současné názory na stanovení a klinické využití kardiomarkerů.* Klin. Biochem. Metab., 13(34), 2005, No.2,p. 98-102.
4. Engliš M., Šochman J. *Srdeční torponiny v klinické praxi*, Praha, TAC-TAC agency s.r.o., 2009
5. Zima, T. (2002). *Laboratorní diagnostika* (První vydání): Praha., Galén
6. Kettner,J. *Nové biomarkery v akutní kardiologii.* Interv Akut. Kardiol. 2008., 7(5): 193 – 199
7. Friedecký, B., Engliš, M., Franeková , J., Jabor, A., Kratochvíla, J., Schneiderka, P., Tichý, M., Zima, T. *Doporučení České společnosti klinické biochemie ke stanovení biochemických markerů poškození myokardu* schváleno 29.11.2007
8. Kelley WE, Januzzi JL, Christenson RH. Increases of cardiac troponin in conditions other than acute coronary syndrome and heart failure. Clin Chem. 2009 Dec;55(12):2098-112.
9. National Academy of Clinical Biochemistry Laboratory Medicine Practice Guidelines: *Clinical Characteristics and Utilization of Biochemical Markers in Acute Coronary Syndromes*, Circulation, April 3, 2007 vol. 115 no. 13 e356-e375

10. Žák, A., Petrášek, J. et al. *Základy vnitřního lékařství*, první vydání (2011): Praha, Galén

11. Doporučení pro diagnostiku a léčbu akutního srdečního selhání. Jindřich Špinar, Petr Janský, Jiří Kettner, Ivan Málek. Cor Vasa 2006;48(1):Kardio. K31

12. Jabor A. *Natriuretické peptidy*. Roche s.r.o., Praha 2010, ISBN 978-80-254-7929-2.

13. 2009 Focused Updates: ACC/AHA *Guidelines for the Management of Patients With ST-Elevation Myocardial Infarction* (Updating the 2004 Guideline and 2007 Focused Update) and ACC/AHA/SCAI Guidelines on Percutaneous Coronary Intervention (Updating the 2005 Guideline and 2007 Focused Update) J Am Coll Cardiol, 2009; 54:2205-2241, doi:10.1016/j.jacc.2009.10.015



CHAPTER 27

27. Laboratory Signs of Malignant Tumours

Author: Prof. MUDr. Marta Kalousová, PhD.

Reviewer: Doc. RNDr. Kristian Šafarčík, Ing. Vladimír Bartoš, PhD.

Laboratory tests in patients with tumours include both routine tests as in other groups of patients as well as special examinations. Routine laboratory tests provide basic information about the patient's condition, nutrition, organ involvement due to the tumour growth or metastasizing of the tumour with consequent impairment of organ function, about any ongoing inflammatory reaction or change in condition following the use of effective therapy. Special laboratory tests are also called "tumour markers" (note: the name is not entirely accurate as an increased tumour marker value does not always mean the presence of a tumour). Results may vary with time, either as a result of a curative action, tumour disease progression or the adverse effects of treatment.

Basic tests should always include blood count, basic urine tests and routine biochemical serum analysis. Cancer screening in patients over 50 includes blood in the faeces (faecal occult blood test) as part of the preventive medical check-up. Many alterations may already be seen in routine laboratory test results.

- The **blood count** may exhibit signs of anaemia as a result of bleeding, chronic disease, malnutrition, iron or other factor deficiency, or bone marrow suppression after chemo and radiotherapy. Leukocytes may be increased not only in blood cell tumours (leukaemias) but also as a result of consequent infection or non-specifically as a reaction to the tumour, or, conversely, may be decreased if bone marrow is suppressed.
- The **basic biochemical serum analysis** may also show various alterations:
  - Elevated inflammatory parameters due to infection or as a non-specific reaction to the cancer process
  - Signs of malnutrition – lowered albumin, prealbumin and cholinesterase
  - Elevated hepatic enzymes (ALT, AST, ALP, GGT) and bilirubin in liver injury, liver metastases and bile duct obstruction
  - Elevated urea, creatinine and potassium in kidney injury or as a result of urinary tract obstruction where the tract is compressed by the tumour
  - Elevated calcium in multiple myeloma or metastatic bone lesions
  - Elevated uric acid as a sign of tumour disintegration
  - Acid-base disorders in renal or pulmonary injury
  - Results of hormonal activity in endocrine-active tumours
  - and others

The **urine test** may show haematuria that may accompany kidney and bladder cancer or glomerulopathies and urinary tract infections, or the presence of a protein as in myeloma (paraprotein). Non-typical tumour cells may be detected in the urine sediment.

27.1. Tumour Markers – Definition and Classification

A tumour marker is a substance present in or produced by the tumour**,** or the host**,** in response to the presence of the tumour. The substance provides information about the biological properties and behaviour of the tumour.

Tests are either **qualitative**, i.e. histopathological tests for markers in the tumour tissue (i.e. cellular tumour markers), or **quantitative,** most commonly measuring the concentration of serum markers (i.e. humoral tumour markers), which is suitable for dynamic monitoring.

The **history** of tumour markers dates back to the 1930s, when human chorionic gonadotropin (hCG), physiologi-

cally produced by the placenta, was discovered in young men with seminoma (Zondek). α_1-foetoprotein, also physiologically produced by the placenta, was first described in mice with liver cancer (Tatarinov) in the 1970s, and in human hepatomas shortly after. Intensive research into these markers followed, and they were subsequently used in oncology and prenatal diagnosis. Based on existing knowledge, local and international expert associations provide recommendations for the rational use of tumour markers in clinical practice.

Tumour markers include:

- **Soluble markers – traditional tumour markers** tested in clinical biochemistry. These are substances of various chemical nature - enzymes, immunoglobulins, hormones, cytokeratins, glycolipids, glycoproteins, carbohydrates and receptor-type molecules.
- **Cellular markers** – receptor-type molecules, usually demonstrated histochemically directly in the tumour tissue. They are important for the selection and control of therapy and for the prognosis of the disease.
- **Circulating cellular elements** – circulating tumour cells, circulating endothelial cells and their precursors.
- **Genetic abnormalities** – oncogene and tumour suppressor gene mutations, protein products of oncogenes and other alterations.

Classical tumour markers can be divided into three groups:

- **Oncofoetal antigens** – substances produced physiologically by the foetoplacental unit (foetus or placenta) in the prenatal period; postnatally they are produced only in connection with certain diseases, usually cancer. Antigens appearing early in the postnatal ontogenesis are characteristic of less differentiated (i.e. more malignant) tumours. This group includes α_1-foetoprotein (AFP), human chorionic gonadotropin (hCG), carcinoembryonal antigen (CEA), placental alkaline phosphatase (PLAP) and several other substances currently of limited use in onkology.
- **Tissue or organ-specific antigens** – substances physiologically occurring inside healthy tissue or organs, outside they are released only in minimal amounts. Increased release is connected with pathological conditions (most often in a tumour disease but sometimes during inflammation or following trauma).Examples include prostate-specific antigen (PSA), neuron-specific enolase (NSE), S-100 protein, soluble cytokeratin fragments (TPA, TPS, CYFRA 21-1), most CA antigens defined using monoclonal antibodies (MAb), squamous cell carcinoma (SCC) antigens, thyreoglobulin (TG) and hormones, and hormone precursors, in tumours of glands physiologically producing them (such as C-peptide in insulinoma, vanilmandelic acid in urine in pheochromocytoma – tumour of the medulla of the adrenal glands).
- **Non-specific antigens** are enzymes and hormones produced by tumours (tumours of organs that do not produce them physiologically – paraneoplastic syndrome), or substances related to the presence of the tumour – markers of cellular proliferation, metabolism or a non-specific reaction of the body. This group includes ferritin, lactate dehydrogenase (LD), thymidine kinase (TK), β_2-microglobulin (ß-2M), some acute phase reactants (APR), lipid-associated sialic acid (LASA), and many other, less frequently used markers. Paraneoplastic hormone production is characteristic primarily of small-cell lung cancer. Various endocrine and metabolic syndromes occur with a corresponding manifestation typical of the relevant hormone (ACTH, parathormone, antidiuretic hormone, etc.).

27.1.1. Basic characteristics of selected tumour markers; use of markers depending on tumour site and type

AFP (α_1- foetoprotein): AFP is an albumin-like glycoprotein. At first it is produced physiologically by the yolk sac and later by the foetal liver. AFP is used during pregnancy for prenatal screening for congenital developmental defects in the 2nd trimester. The pathological occurrence of AFP is seen in tumours containing structures that produce AFP in the foetal period**,** such as hepatocellular carcinoma (but not cholangiocellular carcinoma), hepatoblastoma in children, and germ-cell tumours of the testicle (non-seminomas) and ovary. AFP may also be elevated in patients with hepatic cirrhosis (look out for the transition of cirrhosis to hepatocellular carcinoma!), in acute or chronic active hepatitis**,** where the maximum elevation is usually in the period of regeneration, in toxic liver injury, etc.

CEA (carcinoembryonic antigen): CEA is a glycoprotein (molecular weight: 180,000) with a high carbohydrate content. CEA can be found both in the embryonic tissue (foetal gut) and in adult epithelial tissue. It is used for monitoring colorectal carcinoma and also tumours in other locations (gastrointestinal tract, breast, lungs). CEA levels are higher in smokers and frequently in non-malignant diseases (elderly people, decreased kidney function, inflammatory diseases of the intestine and lungs).

hCG, β-hCG (human chorionic gonadotropin): hCG is a glycoprotein consisting of two noncovalently bound subunits, alpha and beta. The α-subunit is identical to the α -subunit of other hormones (LH, FSH and TSH), which enables cross-reactivity. The whole molecule or only the β-subunit is then tested depending on the indication. To be able to use hCG as a tumour marker, the antibody used must not cross-react with the other peptide hormones of similar structure.

This is why the β-subunit is tested (specific hCG). hCG is used to monitor germ-cell testicular tumours, choriocarcinomas and hydatidiform mole. In addition to the pregnancy test, it is used for the screening for congenital developmental defects in the 1st and 2nd trimesters of pregnancy.

CA 125: This is an antigen characterized by means of monoclonal antibodies. It is suitable for monitoring ovarian carcinomas. Elevated CA 125 levels may be found in ovarian cysts, inflammation of the adnexa, damage to serous membranes, ascites and pleural effusion.

CA 15-3: A glycoprotein from the mucin family**,** it is used primarily to monitor patients with metastasizing breast cancer.

CA 72-4: An antigen characterized by means of monoclonal antibodies and used to monitor stomach cancer.

CA 19-9: This is a glycolipid carrying the blood group Lewis(a) determinant (about 5% of the population does not produce CA 19-9). CA 19-9 plays a major role in monitoring patients with gastrointestinal tumours, pancreatic and bile duct cancer in particular.

CYFRA 21-1: A tissue-nonspecific soluble fragment of cytokeratin 19**,** CYFRA 21-1 is suitable for monitoring patients with non-small-cell lung carcinomas and patients with urinary bladder cancer.

NSE (neuron-specific enolase): NSE is an enzyme characteristic of nerve and neuroendocrine cells (APUD cells). Indications for the NSE test include neuroendocrine tumours, small-cell lung carcinomas and neuroblastomas. Elevated levels may also be found in metastasizing renal carcinomas. NSE is also increased in patients with CNS injury, and rising levels mean a poor prognosis for patients. Moreover, NSE is present in red blood cells and platelets**,** and may release from them; this is why blood should be processed as early as possible after collection in order to separate serum or plasma to avoid platelet haemolysis and disintegration, which would lead to an artificial increase in NSE.

PSA and fPSA (total and free prostate-specific antigen): PSA is a glycoprotein with enzymatic activity (serine protease) and is a physiological component of seminal plasma. If released to the blood, it quickly binds to antiproteases (antichymotrypsin in particular). Total PSA (free PSA plus PSA bound to antichymotrypsin) and free PSA can be tested. Prostatic cancer is an indication for the PSA test. PSA may also be elevated in benign hyperplasia or inflammation of the prostatic gland. The fPSA/PSA ratio**,** expressed as a percentage**,** can be assessed in addition to total PSA. In healthy individuals**,** fPSA represents more than 25% of total detectable PSA, and decreases in prostate carcinoma, often under 10%. Other tests include PSA velocity (PSA growth rate) and PSA density (PSA concentration relative to gland volume). Per-rectal examination, sexual activity, contusion of the prostate from cycling or horseback riding may release PSA from the prostate and increase serum PSA concentration. The histopathological examination of specimens from prostate biopsy is decisive for the diagnosis.

SCC antigen: A squamous cell carcinoma antigen, it is not very specific but is adequate enough for monitoring different types of squamous cell carcinomas, in particular of the head and neck, external genitalia and oesophagus carcinomas.

TPA (tissue polypeptide antigen) and TPS (tissue polypeptide-specific antigen): These are cytokeratins**,** closely related to CYFRA 21-1. TPA is a mixture of soluble cytokeratins 8, 18 and 19, and TPS is a soluble fragment of cytokeratin 18. They are classified as non-specific markers. Elevated values can be found in patients with advanced carcinomas of the breast, lung, colorectum, cervix, ovary and urinary bladder. TPS seems to be more specific for monitoring metastasizing breast carcinoma.

Thymidine kinase (TK): TK is an enzyme of DNA synthesis and therefore a cellular proliferation indicator. Extreme elevation is typical of leukaemia**,** where TK correlates with the severity of the disease. A certain increase can also be seen in solid tumours.

β_2-microglobulin: Exhibiting considerable sequence homology with immunoglobulins, β_2-microglobulin is identical to the light chain in HLAs. It is suitable for monitoring patients with multiple myeloma, non-Hodgkin lymphomas and B-cell chronic lymphocytic leukaemia. Increased β_2-microglobulin can also be seen in patients with chronic inflammatory and autoimmune diseases, acquired immunodeficiency syndrome (AIDS) and after organ transplantations. Renal function must be considered while interpreting results because β_2-microglobulin is excreted mainly through the kidneys (increase in serum due to lower glomerular filtration, and increase in urine due to tubular impairment when reabsorption worsens).

Ferritin: This is a high-molecular protein used to store iron, which is toxic in its free form. Indications for the ferritin test include haematological malignancies, in particular myeloblastic leukaemia and Hodgkin's disease. Ferritin is also elevated in iron metabolism disorders and non-specifically in many tumours such as hepatomas, pancreatic carcinoma, germ-cell tumours, as well as mammary gland and lung carcinomas.

Calcitonin (CT): CT is a peptide consisting of 32 amino acids and is secreted by the parafollicular C-cells of the thyroid. It is suitable for monitoring medullary thyroid cancer and for screening in families with medullary thyroid cancer (multiple endocrine neoplasia - MEN syndrome).

Thyroglobulin (TG): TG is produced by the thyroid and used for the synthesis and storage of hormones. It is used to monitor patients with well-differentiated thyroid cancer. It is not used for diagnosis as TG may also be increased in patients with goitre.

Chromogranin A (CgA): An acid glycoprotein found in secretory granules of normal and neoplastic neuroendocrine tissues, CgA is used in patients with neuroendocrine tumours.

Paraproteins: There are immunoglobulins and immunoglobulin components produced by a single proliferating plasma cell clone. Free light chains are filtered by the kidneys and transfered to urine (Bence Jones protein). Regrettably, test strips employing the protein error of acid-base indicators do not capture these globulins; this means that the protein found in urine is negative. Paraprotein is found in multiple myeloma, Waldenström's macroglobulinaemia, amyloidosis, monoclonal gammopathy of undetermined significance and various lymphoproliferative disorders.

S-100B (S-100B protein): This occurs physiologically in the nervous tissue. It is used to monitor patients with malignant melanoma.

Oestrogenic and progesterone receptors: These tumour tissue markers are tested in the breast cancer tissue**,** usually by immunohistochemical techniques**,** and help determine the tumour's reaction to hormonal therapy and estimate prognosis of the disease.

HER2/neu: The *HER2/neu* gene product is a transmembrane receptor possessing tyrosine kinase activity. It can be determined in a tissue specimen immunohistochemically or by using molecular genetics methods (fluorescence in situ hybridization, FISH) as a *HER2/neu* gene copy number**.** It is an important marker for invasive mammary gland cancer**,** where it is associated with a poor prognosis. If *HER2/neu* gene amplification is proved, the monoclonal antibody against the extracellular domain of this receptor (trastuzumab - HERCEPTIN) can be used for treatment.

In addition to the most common tumour markers listed above, many other markers do exist**.** Examples of less common tumour markers include vanilmandelic acid, which is gradually being substituted by catecholamines and metanephrines (markers for pheochromocytoma), or the test of the 5-hydroxyindoleacetic acid level for carcinoid. Newer soluble markers include proGRP (pro-gastrin-releasing peptide) for small-cell lung cancer and HE4 (human epididymis protein 4) for ovarian cancer monitoring.

Potential new tumour markers are proteins and oncoproteins – mutant gene products important for cell survival, division, differentiation, and metastasizing. Specifically, they interfere with cell cycle regulation (cyclins), apoptosis (e.g. the mutant gene p53 product), signal transduction (HER2/neu, epidermal growth factor), adhesion (adhesion molecules ICAM-1 and VCAM-1), angiogenesis (angiogenesis inhibitors – angiostatin, angiogenin and thrombospondin), or are associated with specific tumour cell properties (matrix metalloproteinases, urokinase-type plasminogen activator uPA and uPA inhibitor PAI-1).

Overview of tumour markers suitable for each cancer type

Gastrointestinal tract	• **Oesophagus** – superior third – ; – CA 72-4, CEA, CA 19-9 • **Stomach** – • **Colon** – • **Liver** - hepatocellular - ; cholangiocellular - • **Pancreas** –
Lungs	• small-cell cancer – NSE • CYFRA 21-1, CEA, SCC

Gynaecological region	• **Mammary gland** – CA 15-3, CEA, TPA/S, steroid receptors • **Ovary** – non-mucinous - ; mucinous – ; germinal – AFP, hCG • **Cervix uteri** – epidermoid - SCCA, CYFRA 21-1; adenocarcinomas – CEA • **Corpus uteri** – • **Vulva** – • **Chorion** (**choriocarcinoma, mola hydatidosa**) – hCG
Urogenital region	• **Kidneys** – TPA/S, CEA, NSE • **Urinary bladder** – TPA/S, CYFRA 21-1 • **Prostate** – PSA, fPSA, chromogranin • **Testes** – seminomas - hCG, AFP, NSE; non-seminomas – hCG, AFP
Haematological malignancies	– thymidine kinase, ferritin, lactate dehydrogenase – Hodgkin – β2-microglobulin, ferritin, lactate dehydrogenase; non-Hodgkin – thymidine kinase, β2-microglobulin, lactate dehydrogenase – β2-microglobulin, paraproteins
Other	• **Head, neck** – SCCA, CYFRA 21-1 • **Carcinoid** – 5-hydroxy, 3-indoleacetic acid, NSE • **Thyroid gland** – differentiated carcinoma – thyroglobulin, CEA; medullary - CT, CEA (NSE); anaplastic TPA/S • **Melanoma** – S-100B, NSE, thymidine kinase • **CNS** – neuroblastomas – NSE; gliomas – CEA; astrocytomas – thymidine kinase

Table 27.1. Overview of tumour markers suitable for each cancer type

Cancer type	Marker
Gynaecological region	
• **Mammary gland**	– CA 15-3, CEA, TPA/S, steroid receptors
• **Ovary**	1
– non-mucinous	- CA 125
; mucinous	– CA 19-9, CA 72-4
germinal	AFP, Hcg
• **Cervix uteri**	
; epidermoid	; SCCA, CYFRA 21-1
adenocarcinomas	CEA

Gastrointestinal tract	**Oesophagus** – superior third – SCCA, CYFRA 21-1; inferior third – CA 72-4, CEA, CA 19-9 **Stomach** – CA 72-4, CEA **Colon** – CEA, CA 19-9 **Liver** - hepatocellular - AFP, CEA; cholangiocellular - CA 19-9 **Pancreas** – CA 19-9, CEA
Lungs	small-cell cancer - NSE; non-small-cell cancer - CYFRA 21-1, CEA, SCC
Gynaecological regio	• **Mammary gland** – CA 15-3, CEA, TPA/S, steroid receptors • **Ovary** – non-mucinous - CA 125; mucinous – CA 19-9, CA 72-4; germinal – AFP, hCG • **Cervix uteri** – epidermoid - SCCA, CYFRA 21-1; adenocarcinomas – CEA • **Corpus uteri** – CA 125, CEA • **Vulva** – SCCA • Chorion (choriocarcinoma, mola hydatidosa) – hCG Urogenital region • **Kidneys** – TPA/S, CEA, NSE • **Urinary bladder** – TPA/S, CYFRA 21-1 • **Prostate** – PSA, fPSA, chromogranin • **Testes** – seminomas - hCG, AFP, NSE; non-seminomas – hCG, AFP Haematological malignancies • **Leukaemia** – thymidine kinase, ferritin, lactate dehydrogenase • **Lymphoma** – Hodgkin – β2-microglobulin, ferritin, lactate dehydrogenase; non-Hodgkin – thymidine kinase, β2-microglobulin, lactate dehydrogenase • **Multiple myeloma** – β2-microglobulin, paraproteins
Urogenital region	• **Kidneys** – TPA/S, CEA, NSE • **Urinary bladder** – TPA/S, CYFRA 21-1 • **Prostate** – PSA, fPSA, chromogranin • **Testes** – seminomas - hCG, AFP, NSE; non-seminomas – hCG, AFP

Haematological malignancies	• **Leukaemia** – thymidine kinase, ferritin, lactate dehydrogenase • **Lymphoma** – Hodgkin – β2-microglobulin, ferritin, lactate dehydrogenase; non-Hodgkin – thymidine kinase, β2-microglobulin, lactate dehydrogenase • **Multiple myeloma** – β2-microglobulin, paraproteins
Other	• **Head, neck** – SCCA, CYFRA 21-1 • **Carcinoid** – 5-hydroxy, 3-indoleacetic acid, NSE • **Thyroid gland** – differentiated carcinoma – thyroglobulin, CEA; medullary - CT, CEA (NSE); anaplastic TPA/S • **Melanoma** – S-100B, NSE, thymidine kinase • **CNS** – neuroblastomas – NSE; gliomas – CEA; astrocytomas – thymidine kinase

Gastrointestinal tract

Oesophagus – superior third – SCCA, CYFRA 21-1; inferior third – CA 72-4, CEA, CA 19-9

Stomach – CA 72-4, CEA

Colon – CEA, CA 19-9

Liver - hepatocellular - AFP, CEA; cholangiocellular - CA 19-9

Pancreas – CA 19-9, CEA

Lungs – small-cell cancer - NSE; non-small-cell cancer - CYFRA 21-1, CEA, SCC

Gynaecological region

- **Mammary gland** – CA 15-3, CEA, TPA/S, steroid receptors
- **Ovary** – non-mucinous - CA 125; mucinous – CA 19-9, CA 72-4; germinal – AFP, hCG
- **Cervix uteri** – epidermoid - SCCA, CYFRA 21-1; adenocarcinomas – CEA
- **Corpus uteri** – CA 125, CEA
- **Vulva** – SCCA
- Chorion (choriocarcinoma, mola hydatidosa) – hCG

Urogenital region

- **Kidneys** – TPA/S, CEA, NSE
- **Urinary bladder** – TPA/S, CYFRA 21-1
- **Prostate** – PSA, fPSA, chromogranin
- **Testes** – seminomas - hCG, AFP, NSE; non-seminomas – hCG, AFP

Haematological malignancies

- **Leukaemia** – thymidine kinase, ferritin, lactate dehydrogenase
- **Lymphoma** – Hodgkin – β2-microglobulin, ferritin, lactate dehydrogenase; non-Hodgkin – thymidine kinase, β2-microglobulin, lactate dehydrogenase
- **Multiple myeloma** – β2-microglobulin, paraproteins

Other

- **Head, neck** – SCCA, CYFRA 21-1

- **Carcinoid** – 5-hydroxy, 3-indoleacetic acid, NSE
- **Thyroid gland** – differentiated carcinoma – thyroglobulin, CEA; medullary - CT, CEA (NSE); anaplastic TPA/S
- **Melanoma** – S-100B, NSE, thymidine kinase
- **CNS** – neuroblastomas – NSE; gliomas – CEA; astrocytomas – thymidine kinase

27.2. Tests for Tumour Markers, Indication and Interpretation

The process of tumour marker testing includes the indication for the test, the pre-analytical phase, the test itself and the proper interpretation of results.

In terms of **pre-analytics**, the specific features of some tumour markers (see above) have to be considered in addition to generally accepted principles, in particular the effect of prostate agitation on the PSA, effect of haemolysis on NSE and the effect of secretions (saliva) on the SCC level.

Tumour markers are predominantly protein in nature and are determined by **immunoassays** with different types of labelling (radioimmunoassay, enzymoimmunoassay, fluorescence analysis, chemiluminiscence). The same method and diagnostic set from the same manufacturer should be used because results acquired using different techniques may vary; otherwise**,** re-baselining is required. In terms of analysis, problems may be caused by the cross reactivity of molecules with a similar structure, the hook effect caused by a high marker concentration or by interference from heterophilic antibodies and human anti-mouse antibodies (HAMA), which may be produced during biological therapy, for example.

As regards to the **indication and interpretation** of tumour markers, it is important to know that they are intended primarily for **monitoring** patients with cancer, i.e. the test is usually indicated by the oncologist. The dynamics of change is important. First, the individual level should be determined, i.e. the concentration in the stabilized condition following surgery, i.e. removal the tumour mass**.** The next step is repeated follow-up monitoring, frequent at the beginning, and later at intervals of about 3-6 months. A rise in 3 consecutive samples and a rise in the marker by more than 25 % are considered significant. Such increased marker level, although in the reference range, may provide an earlier warning of relapse than imaging methods**,** and suggests the indication of additional diagnostic and therapeutic procedures. Several markers should be monitored at the same time because this increases the chance that a relapse is detected (usually the 2 markers most suitable for the region are chosen).

Tumour markers are not directly suitable for **diagnostic purposes,** but may be a useful aid for the diagnostic process. A positive finding of tumour markers has a diagnostic value; a negative finding does not necessarily mean the absence of cancer. Histopathological examination complemented by tumour marker demonstration is always decisive for the diagnosis. An increase in a tumour marker does not always necessarily mean cancer, it may be caused by inflammation, benign lesion, trauma, use of effective therapy or liver or kidney injury in the case of markers eliminated by these organs.

Only blood in the faeces (occult blood) is tested for **screening**; other markers are only tested for specific groups (calcitonin in families with medullary thyroid cancer, CA 15-3 in BRCA mutations). The use of other markers (not even PSA) for screening is not recommended at present.

27.3. Tumour Marker Evaluation

Tumour markers are assessed in terms of their specificity (true negativity in healthy individuals) and sensitivity (true positivity in individuals with the diseases), or their positive and negative predictive value (true positivity in all positive and true negativity in all negative results). ROC (receiver operating characteristic) curves and the area under the curve are used for marker evaluation.

The importance of laboratory diagnostics in oncology is constantly increasing**.** Tumour marker testing plays an important role**,** especially in monitoring patients with tumour diseases**,** and can detect a relapse of the disease earlier than imaging methods or before the disease manifests itself clinically. Laboratory examinations also provide information about the nutritional status, presence of inflammation or failure of organs affected by the tumour process. Selected tests may also help in the differential diagnosis process in patients with suspected tumour disease. Additional parameters will probably be used in future; such tests could enable the effective targeted treatment of cancer.

CHAPTER 28

28. Cytochemical Examination of Cerebrospinal Fluid

Author: doc. MUDr. Milan Dastych, CSc.

Reviewer: MUDr. Ing. David Zeman, Ph.D.

28.1. Cerebrospinal Fluid Cytology

The Fuchs-Rosenthal chamber is used for the quantitative determination of cerebrospinal fluid elements. The number of red blood cells is determined in the native preparation, which is subsequently stained with fuchsine to highlight and count white blood cells. 0 - 3 elements per µl are considered normal. The chamber volume is 3.2 µl, so the total cell count found is divided by three.

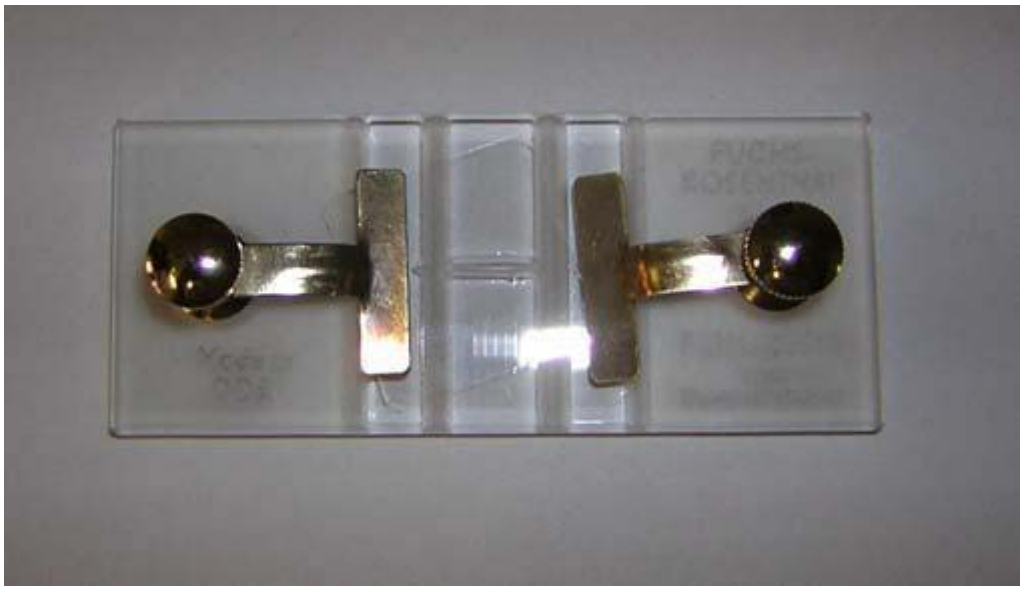


Figure 28.1. Fuchs-Rosenthal chamber

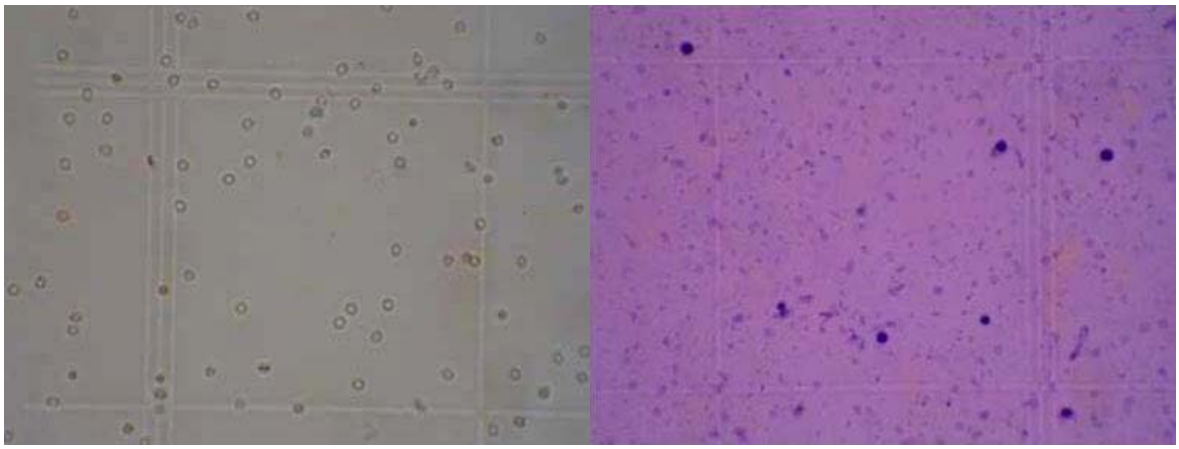


Figure 28.2. Unstained (native – left) and stained (right) liquor in the Fuchs-Rosenthal chamber

The morphological differentiation of white blood cells into mononuclear and polynuclear leukocytes in the Fuchs--Rosenthal chamber is not very reliable. A permanent cytological preparation must be made for a more accurate differentiation of elements. The basic staining method is Pappenheim's stain. A permanent cytological preparation is made

using cytosedimentation or careful cytocentrifugation methods.

The preparation is used for cytological diagnosis, and enables the evaluation of the counts of each cell type and the state of cell function.

Physiologically, only mononuclear elements are present in the liquor, with resting elements being predominant. The presence of red blood cells, polynuclear cells (except for an artificial admixture of blood or neonatal CSF), plasmocytes and phagocytes is pathological.

28.1.1. Type of Cells in the CSF

28.1.1.1. Lymphocyte

The resting lymphocyte is a cell of 8 - 10 µm, has a narrow rim of plasma and a compact nucleus filling almost the whole cell. These cells can be found in normal cerebrospinal fluid. Following an encounter with an antigen, lymphoid cells transform and change appearance. An activated (reactive) lymphocyte is a lymphoid cell of 11 - 18 µm with a distinctive basophilic cytoplasm and a large nucleus. The lineage of T-lymphocyte development ends with lymphoid cells; only B-lymphocytes develop into plasmocytes. T-lymphocytes are responsible for cytolysis and the activation of macrophages. Plasmocytes are the largest lymphoid cells, up to 21 µm in size with basophilic plasma. They have a round or oval, usually eccentric, nucleus with visible nucleoli. The perinuclear halo of the cytoplasm is typical.

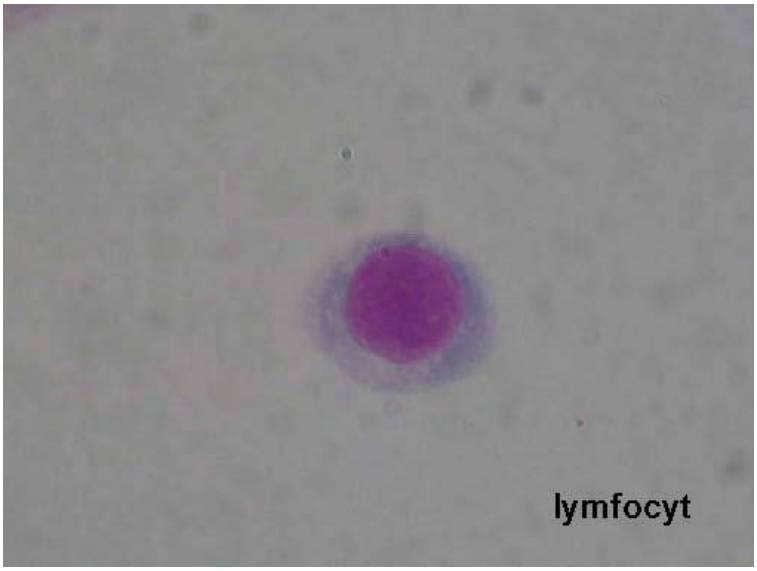


Figure 28.3. A slightly activated (reactive) lymphocyte

28.1.1.2. Plasmocyte

Plasmocytes are the ultimate stage of B-lymphocytes following immunological stimulation, and produce immunoglobulins. Their presence in the cerebrospinal fluid is pathological; most commonly they can be found in patients with viral inflammations.

28.1.1.3. Monocyte

Monocytes are cells 15 - 30 µm in size. The nucleus usually has an irregular shape; most often it is horseshoe or bean-shaped. Stained plasma is lighter than stained lymphocytes. These cells have very varied morphology. Activated monocytes differ from resting monocytes by their larger size and vacuolated cytoplasm. They are capable of phagocytosis of other cells or non-cellular elements. Activated monocytes with devoured material are called macrophages. Different types can be distinguished depending on the phagocytized material; such types include erythrophages, hemosiderophages, lipophages, etc. Monocytes are part of the normal cell count, but activated monocytes or macrophages suggest a pathological process.

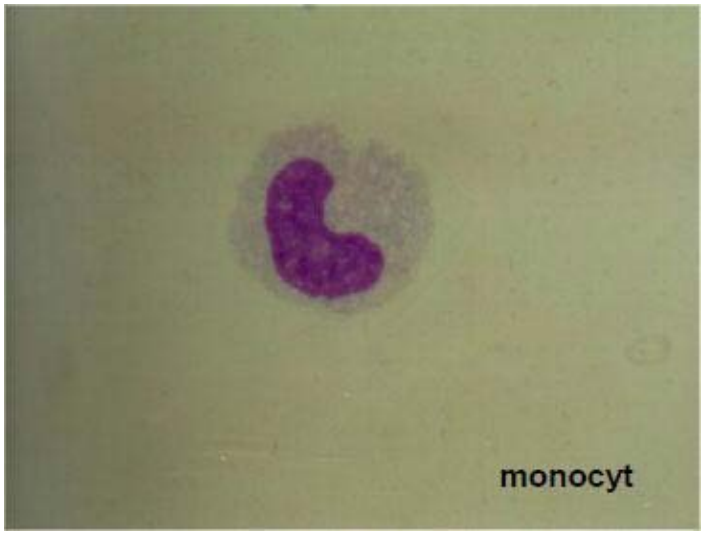


Figure 28.4. Resting monocyte

28.1.1.4. Neutrophilic polymorphonuclear cell

Neutrophilic polynuclear cells are 10 - 15 µm in size. Under normal circumstances they can be found in the CSF only in their mature, segmented form. The characteristic morphology with a split nucleus and granules in the cytoplasm corresponds to blood neutrophils. Of the granulocytic lineage of leukocytes, these are the type most frequently found in the CSF, and accompany acute CNS inflammations, bacterial inflammations in particular. Granulocytes are also capable of phagocytosis; they engulf bacteria, for example. However, they are incapable of engulfing other leukocytes or erythrocytes. Eosinophil granulocytes are somewhat larger than neutrophils. The nucleus is mostly split into two parts and the cytoplasm contains many reddish granules. They are rather rare in the CSF – their presence may suggest a chronic process, or they can be seen in parasitic diseases.

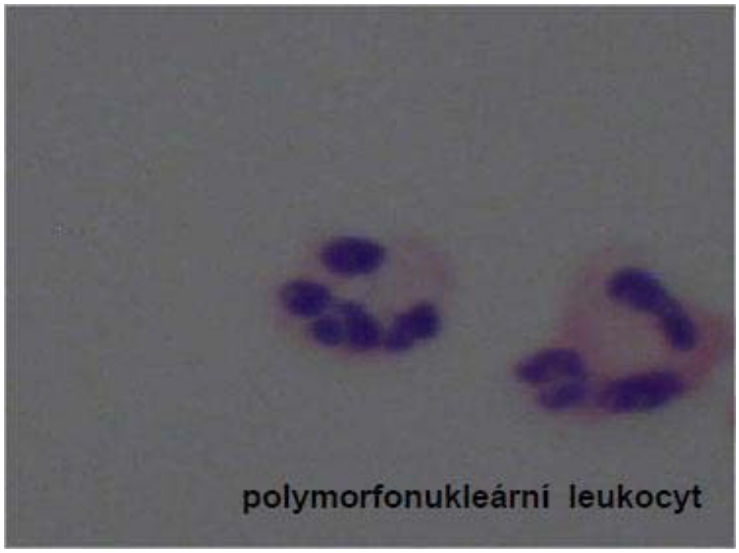


Figure 28.5. Neutrophilic polynuclear cell

28.1.1.5. Tumour cells

Tumour cells in the CSF originate much more often from extracranial tumours than intracranial tumours. Cytological CTF examination and the search for tumour cells are important especially in patients with metastatic tumours, leukaemia, lymphomas, primary brain tumours and meningeal irritation of unknown origin. The generally accepted criteria of cell malignancy include a substantial change in the nuclear-cytoplasmic ratio in favour of the nucleus, large nucleoli, increased mitosis, the considerable cell size, and tendency to cell clustering.

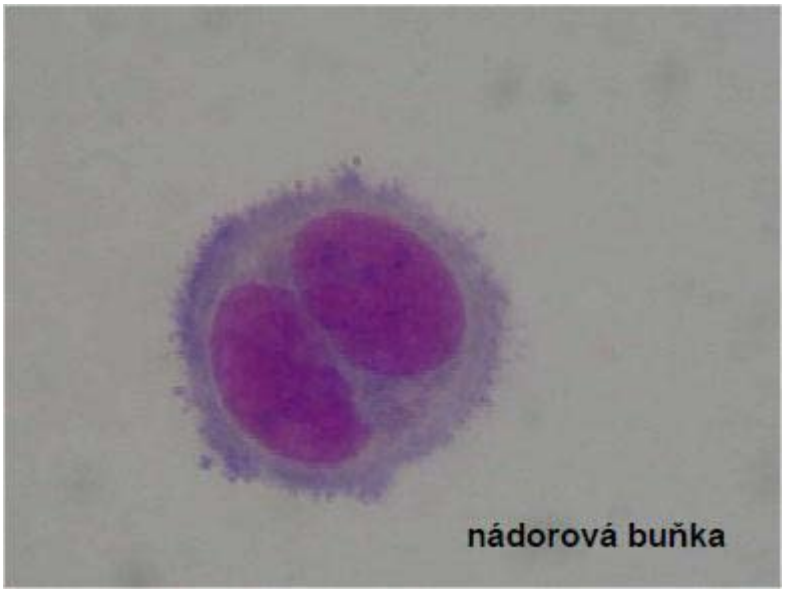


Figure 28.6. Tumor cell

28.2. Types of Cytological Findings in CSF

Depending on the prevalent cell lineage, pleocytosis or pathological oligocytosis is classified as granulocytic, lymphocytic, monocytic or tumorous.

Granulocytic pleocytosis with a prevalence of neutrophils occurs in purulent inflammations; eosinophils can often be found in parasitic and mycotic diseases or chronic pathological processes.

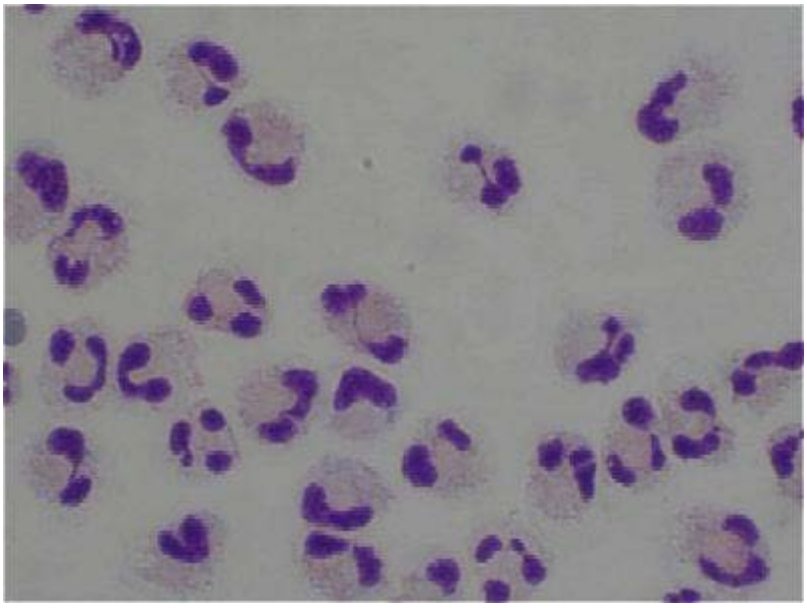


Figure 28.7. Granulocytic pleocytosis - neutrophils: segmented and banded

Granulocytic oligocytosis is common at the initial stage of non-purulent inflammations or brain ischaemia.

Lymphocytic pleocytosis occurs in non-purulent inflammations (viral infections and infections caused by Borrelia, Leptospira or tuberculosis bacilli).

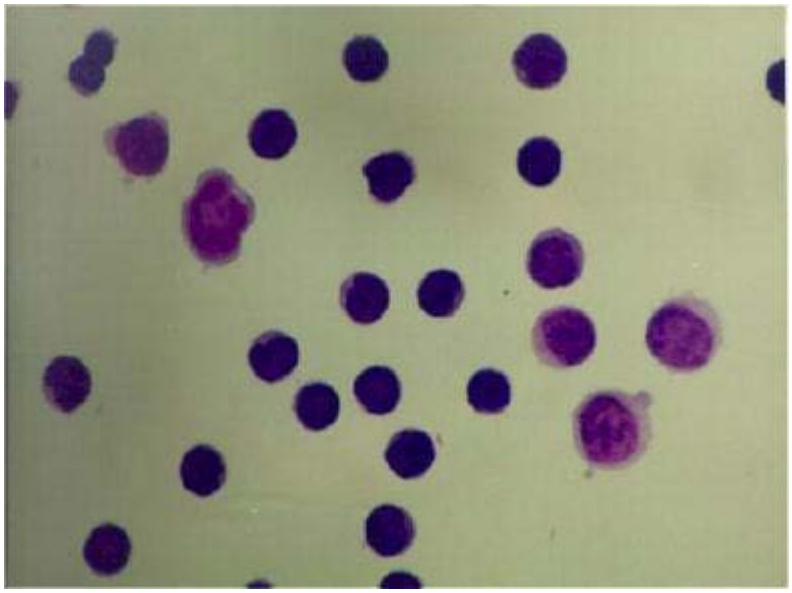
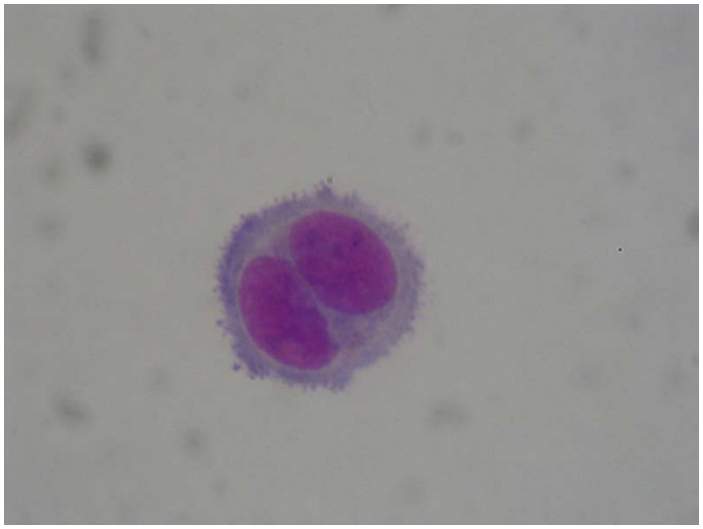


Figure 28.8. Lymphocytic pleocytosis: Larger cells with a lighter nucleus are activated lymphocytes

Lymphocytic oligocytosis often occurs in multiple sclerosis.

Monocytic pleocytosis or oligocytosis with detected activated monocytes is a non-specific finding, typical of non--infectious diseases such as compression syndrome or autoimmune disease. The finding is also characteristic of the final stage of inflammation.

Tumorous pleocytosis or oligocytosis is indicative of a malignant disease.



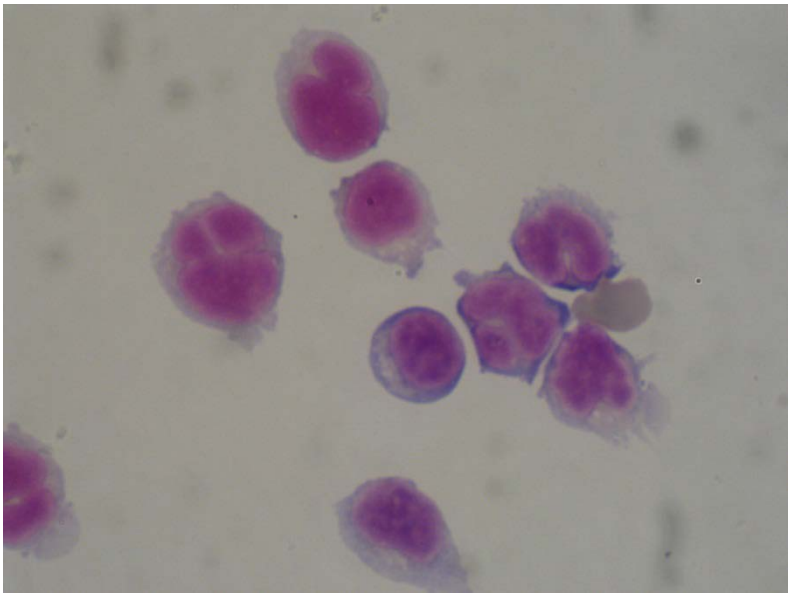


Figure 28.9. Tumor cells in CSF

A normal cytological finding in the CSF is oligocytosis (cell count under 3 - 5 per µl), lymphocytes, monocytes are present, activated cell form counts are under 10%.

28.3. Biochemical Examination of Cerebrospinal Fluid

Appearance of CSF

- Clear and colourless CSF (normal finding)
- Sanguinolent CSF is caused by the pathological or artificial admixture of blood
- Xanthochromic CSF (yellowish) may be a sign of an old bleeding into the CSF system or occurs in a severe blood-CSF barrier dysfunction
- Turbid CSF can be found in purulent meningitis

Total protein – the total protein level in its normal condition is about 200 times lower than the serum protein level. Under normal circumstances, 80% of the total protein is of serum origin; the remaining 20% is of brain origin. Physiological levels of the total protein from lumbar puncture are 0.15 - 0.40 g/l. The total protein in the ventricular system is markedly lower. Newborns have a higher CSF total protein level due to the immature blood-CSF barrier. Albumin is the most abundant component of the total protein. Some serum proteins such as prealbumin and transferrin are modified in the CSF space. Immunoglobulins transfer to the CSF from blood or are produced by intrathecal synthesis in CSF spaces. Inflammatory markers include haptoglobin, C-reactive protein, C3 and C4 complement components, antithrombin III and α1-antitrypsin. Orosomucoid is a tumour marker.

Increased total protein is clinically important since it occurs in inflammations (in bacterial inflammations there is a pronounced increase up to tens of grams), CSF circulation disorders (prolapsed intervertebral disc) or intrathecal synthesis of immunoglobulins following immune system activation.

Glucose is the basic source of energy for nerve tissue; the CSF glucose level depends on the serum glucose level, and is about 60% of the serum level. The CSF glucose/S glucose ratio is more significant, and is 0.6 in normal conditions. A reduction in CSF glucose or the CSF glucose/S glucose ratio is clinically significant; it is associated with bacterial inflammations, meningeal infiltration in tumour diseases and subarachnoid haemorrhage.

Lactate – the CSF lactate level does not depend on the plasma concentration; lactate practically does not pass the blood-CSF barrier. Physiological lactate levels are 1.2 – 2.1 mmol/l. Lactate is produced by glucose metabolism in the absence of oxygen. Elevated lactate levels are clinically significant. Increase occurs in inflammations, and a slight increase accompanies viral or tuberculous meningitis. A marked increase occurs in bacterial inflammations. The value for discriminating between types of meningitis is about 4 mmol/l. Lactate concentration is often increased in ischaemic vascular lesions.

Chlorides – physiological levels are 120 - 132 mmol/l; chloride testing is being abandoned. Chlorides are usually

decreased in pronounced blood-CSF barrier dysfunction, particularly in specific inflammations such as tuberculous meningitis.

Albumin is synthesized in the liver and enters the CSF only through the blood-CSF barrier. It is used for assessing the function of the barrier. To assess the dysfunction of the blood-CSF barrier function, the sensitivity of the CSF/serum albumin ratio is higher than the absolute CSF albumin concentration.

The albumin quotient Qalb=CSFalb/Salb is calculated from the CSF and serum albumin concentration. Albumin quotient values depend not only on barrier permeability in the narrower sense of the word, but also on CSF turnover (circulation). It is markedly higher in newborns. Normal Qalb values are under 7.4; more precisely, the albumin quotient depends on age according to the formula Qalb ≤ 4 + (age in years/15). A slightly damaged barrier can be found in chronic neuroinfections or multiple sclerosis. Medium-severe barrier impairments can be seen in viral infections, for example, and severe barrier impairment is typical of bacterial meningitis, Guillain-Barré polyradiculoneuritis, malignant meningeal infiltration or herpetic encephalitis.

Immunoglobulins – the source of immunoglobulins is either serum, which is where they come from if the barrier is damaged, or intrathecal synthesis in CSF spaces in CNS diseases connected with immune response.

28.4. Examination of Intrathecal Immunoglobulin Synthesis

Intrathecal antibody synthesis in the central nervous system comes from perivascular infiltrates of B-lymphocytes, which proliferate locally, ripen into plasmocytes and produce the relevant antibodies. Unlike the known and characteristic reaction in the serum, where the generation of antibodies changes from the IgM class to the IgG class at the subacute or chronic stage of the disease, such transition does not take place in intrathecal synthesis. IgG/IgM/IgA class antibodies are generated right at the beginning of the disease and the generation is relatively constant. An intrathecal antibody response need not always necessarily mean an acute disease. It may also be present in an inflammatory disease of the CNS manifested by an increased number of cells in the CSF, in residual intrathecal synthesis following a disease, or in a chronic inflammatory autoimmune process in the CNS. Today, modern diagnosis of multiple sclerosis from the CSF is primarily based on demonstrating the intrathecal synthesis of immunoglobulins.

28.4.1. Quantitative Demonstration of Intrathecal Immunoglobulin Synthesis

The CSF immunoglobulin concentration may be increased as a result of increased serum immunoglobulin concentration, blood-CSF barrier dysfunction or both these factors. Protein diffusion from the serum to the CSF is the dominant mechanism of transport through the blood-CSF barrier. According to the physical laws of diffusion, larger molecules such as IgM pass more slowly than smaller molecules such as IgG and albumin. An increase in certain serum protein concentration results in an increase in the same protein in the CSF, but the gradient expressed by the CSF to serum protein ratio remains constant. If the blood-CSF barrier is impaired, there is a pathological increase in the CSF protein concentration, which is why albumin and the relevant immunoglobulins have to be tested in the serum and CSF at the same time. Determination of the intrathecal synthesis of IgG, IgM and IgA is the most significant for clinical diagnosis. Reiber's formula, describing the relationship between the albumin quotient and relevant immunoglobulin quotient, is currently recommended for the demonstration of intrathecal immunoglobulin synthesis. The relationship is not linear but hyperbolic. The blue line represents the border between local immunoglobulin synthesis and passive immunoglobulin transfer. Values above this line mean intrathecal synthesis and the range is shown with dotted lines and expressed as a percentage. The vertical dashed line separates the normal and impaired blood-CSF barrier.

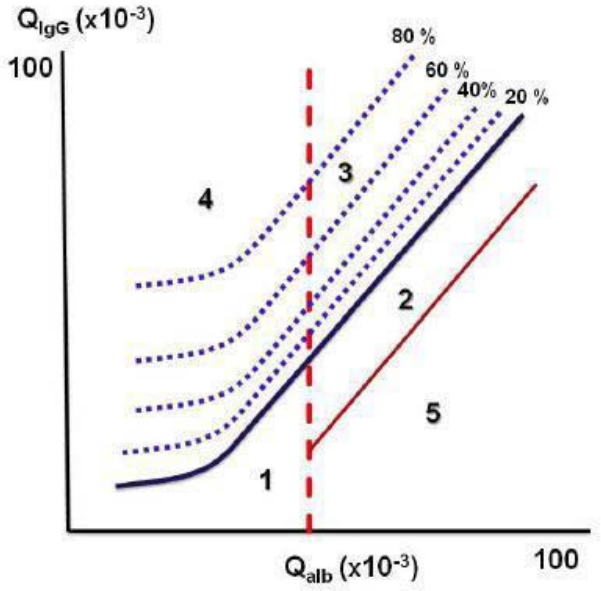


Figure 28.10. Reiber's diagram

28.4.2. Interpretation of pathological findings in relation to the blood-CSF barrier function and intrathecal synthesis:

- Zone 1 – normal finding
- Zone 2 – blood-CSF barrier dysfunction without intrathecal synthesis
- Zone 3 – blood-CSF barrier dysfunction with intrathecal immunoglobulin synthesis
- Zone 4 – intrathecal synthesis with blood-CSF barrier function preserved
- Zone 5 – analytical error zone

28.4.3. From the diagnostic point of view, the ability to distinguish between conditions is facilitated by presence of intrathecal synthesis of each immunoglobulin:

- Intrathecal Ig synthesis not demonstrated: early bacterial and viral meningitis, Guillain-Barré syndrome;
- Dominant intrathecal IgG-synthesis: multiple sclerosis, neurosyphilis, chronic HIV encephalitis;
- Dominant intrathecal IgA-synthesis: neurotuberculosis, brain abscess;
- Dominant intrathecal IgM-synthesis: neuroborreliosis.

28.4.4. 31.4.4 Qualitative Demonstration of Intrathecal Immunoglobulin Synthesis

The qualitative demonstration of intrathecal immunoglobulin synthesis consists in the detection of oligoclonal immunoglobulins by isoelectric focusing (IEF). Serum diluted to the concentration corresponding to the cerebrospinal fluid is tested by analogy at the same time. The demonstration of two or more oligoclonal bands in the CSF without their counterparts in the serum is indicative of intrathecal synthesis. The method of isoelectric focusing is used to divide compounds that are amphoteric in nature (having acid and alkaline groups) with the pH gradient as the borderline depending on the isoelectric point. Based on international consensus, IEF of the CSF and serum reveals five recognized patterns:

Type 1: Normal condition with no oligoclonal bands (OCBs) demonstrated

Type 2: OCBs in CSF but not in S (typical of multiple sclerosis)

Type 3: OCBs in both materials, but additional bands in CSF (neuroborreliosis, cerebrospinal sclerosis)

Type 4: Identical OCBs in S and CSF (paraneoplastic syndromes, SLE)

Type 5: Monoclonal bands in CSF and S (monoclonal gammopathy)

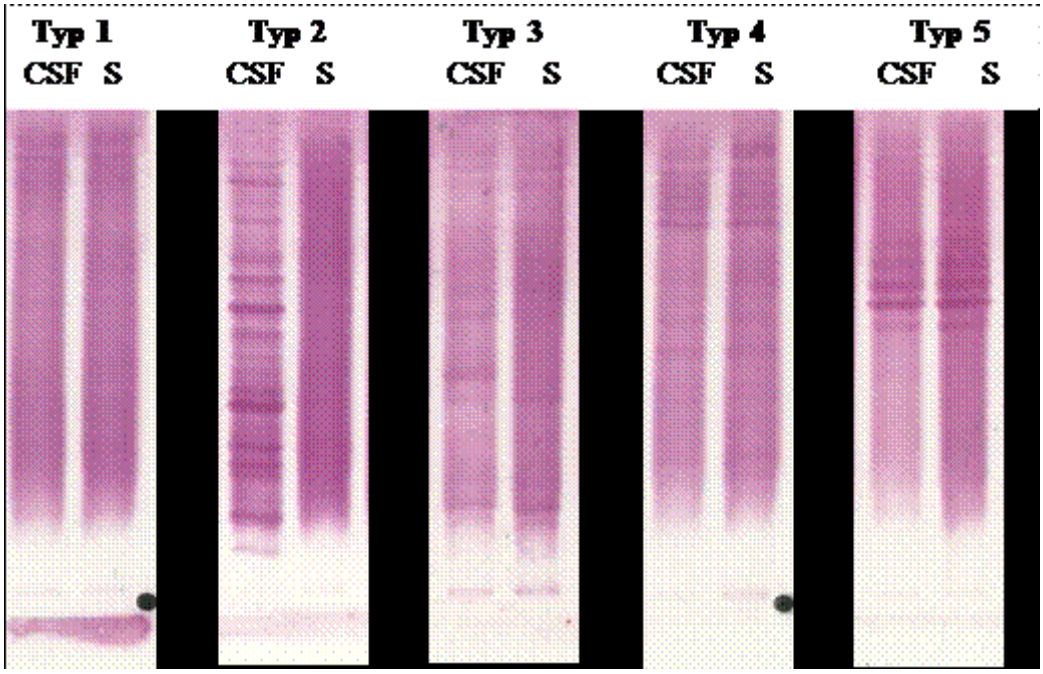


Figure 28.11. Isoelectric focusing (blood - liquor)

Basic IEF patterns

The determination of oligoclonal bands is positive in up to 98% of patients with a diagnosis of multiple sclerosis. Oligoclonal bands can also be found in infectious processes and autoimmune or systemic diseases. The introduction of the specific MRZ (morbilli, rubeola and varicella zoster) reaction has brought important results for the diagnosis of multiple sclerosis based on CSF analysis, since it enables some chronic CNS diseases to be detected already at the time of first clinical signs. The MRZ reaction is an intrathecal synthesis of specific IgG antibodies against neurotropic measles, rubella and herpes zoster viruses. The presence of the MRZ reaction is explained by the modern theory of an immunological network, where every immune reaction induced by a specific antigen affects the entire immunological network. In addition to specific antibodies against the causative agent, each immune reaction produces an increased number of many other antibodies and autoantibodies, which is called a polyspecific immune response. Concentrations of each specific antibody in the CSF and serum are measured and their ratio relates to the ratio of "total" antibodies of the same class as the antibody index (AI). Values higher than 1.4 are pathological and indicate the intrathecal synthesis of specific antibodies. The MRZ reaction is specific to chronic autoimmune inflammatory diseases of the nervous system, multiple sclerosis in particular.

28.5. Spectrophotometry of Cerebrospinal Fluid

CSF spectrophotometry is used if an intermeningeal haemorrhage is suspected. It is useful at early stages when no alterations can be found yet in the cytological analysis. Spectrophotometry is 10 times more sensitive than the human eye; a positive finding can be detected even in a seemingly colourless CSF. Spectrophotometry records visible light absorption (380 – 700 nm) and the presence of oxyhaemoglobin and bilirubin is detected.

The detection of oxyhaemoglobin may suggest a fresh haemorrhage; oxyhaemoglobin appears 4 - 8 hours after a haemorrhage has started. The absorbance maximum is at 415 nm; 2 minor peaks are at 540 nm and 580 nm. The maximum for bilirubin is reached 3 days after a haemorrhage has started, persists for 3 weeks, and the absorbance maximum is at 450 nm.

28.6. Microbiological Examination and Pathogenic Agent Detection

Basic procedures include culture and sensitivity testing, and the microscopic and bacteriological examination of smears. Specific antibodies are examined in the serum and CSF, and intrathecal antibody synthesis is demonstrated. Polymerase chain reaction (PCR) is a specific and very sensitive method for the detection of DNA pathogenic agents (most often used to detect Borrelia and herpetic viruses).

28.7. Prospects for New Advances in CSF Examination

Attention is currently turning to new opportunities in the diagnosis of degenerative and metabolic diseases of the central nervous system, as well as prion infections. The research so far has tended towards the experimental. For example, an increased concentration of the 14 - 3-3 protein, which is currently considered a general CSF marker for tissue destruction, suggests in the CSF a diagnosis of Creutzfeldt-Jakob disease. The CSF-based diagnosis of Alzheimer's disease relies on the determination of β-amyloid, τ-protein and phospho-τ-protein. The β-amyloid protein is a component of amyloid plaques accumulating in the brain. The τ-protein can be found in the CNS cytoskeleton. Decreased β-amyloid and increased τ-protein values are found in Alzheimer's disease. The τ/ β-amyloid index is usually increased. Another parameter that may help in the diagnosis of neurodegenerative diseases is cystatin C. Cystatin C is a cysteine protease inhibitor; it is quite stable in the circulatory system and is known in clinical practice as an indicator of estimated glomerular filtration. Increased cystatin C levels in the CSF are found in patients with neurodegenerative diseases of the CNS. Structural S-100 protein belongs to the family of calcium-binding proteins. Increased S100 (S100B) levels are a sign of the structural involvement of the CNS tissue; the S100 (S100B) protein can also be measured in the serum to estimate the prognosis and to monitor the patient's status without lumbar puncture. β2 microglobulin is a protein present in all body fluids. An increase in β2 microglobulin concentration in the CSF can be found in conditions generally associated with the activation and multiplication of lymphocytic and macrophagic elements. The presence of neuronal antibodies has been examined in patients with multiple sclerosis – antibodies against myelin sheath molecules have been found in the CSF. In particular, it is the myelin basic protein (MBP) from the group of structural proteins which forms the basis of myelin. Antibodies against MBP reflect the level of myelin destruction.

28.8. Determination of Liquorrhoea

It is sometimes important to determine whether the secretion from a patient's nose following head trauma is just an inflammatory mucosal discharge or cerebrospinal fluid. The latter case presents a serious condition with CSF passage into the nasal cavity, which endangers the patient by the transfer of bacterial flora to the meninges. CSF can be detected by determining the CSF-specific parameter. A beta-trace protein test is available – beta-trace protein is an enzyme synthesized in choroid plexus cells. CSF beta-trace protein concentrations are 20 - 30 times higher than serum concentrations. β2 transferrin is another parameter. Sialic acid residues cleave off from serum transferrin in CSF spaces (by neuraminidase in the brain) to form asialotransferrin, which can be detected by immunofixation electrophoresis in the β2-globulin region.

CHAPTER 29

29. Inherited Metabolic Diseases – Laboratory Diagnostics

Author: MUDr. Ivan Šebesta CSc., Doc. MUDr. Viktor Kožich CSc.

Reviewer: Prof. MUDr. Richard Průša, CSc.

29.1. IMD Characteristics

Inherited metabolic diseases (IMDs, originally also referred to as “inborn errors of metabolism”) comprise a large group of more than 800 diseases which have a common characteristic – genetic defect in the production of proteins, most often enzymes. IMDs are rare diseases (i.e. diseases with an incidence under 1:2,000 in the population), and comprise about 10 - 15% of all known nosological entities.

29.2. Incidence

Incidence can only be determined for diseases tested in every individual born, i.e. only if neonatal screening of general population is conducted. In the current IMD screening system in the CR, the expected incidence for the ten screened IMDs (see below) is about 1:4,000.The incidence of defects not included in the screening programme cannot be reliably determined due to the unknown proportion of patients who escaped correct diagnosis because the suspected IMD was not taken into account in the differential diagnosis, and special tests were not performed. About one hundred new patients with IMDs have been diagnosed annually by selective screening in the CR during the last decade (see Figure 1). On the basis of information from world (nebo jen literature) literature, the estimated cumulative incidence of all known IMDs is at least 1:500 of children born. It should be emphasized that the actual incidence is probably much higher, which means that every physician of any specialization meets patients with IMDs in their practice. With the ever widening offer of more and more advanced laboratory tests, it has transpired over the last few years that the number of patients with the first clinical manifestation of a metabolic disease in adult age is growing. These diseases may include for example, disorders of mitochondrial energy metabolism or metabolic urolithiases (cystinuria, inherited xanthinuria, 2,8-dihydroxyadenine lithiasis), kidney injury (familial juvenile hyperuricemic nephropathy and others). IMDs are by no means, therefore, only the preserve of paediatricians.

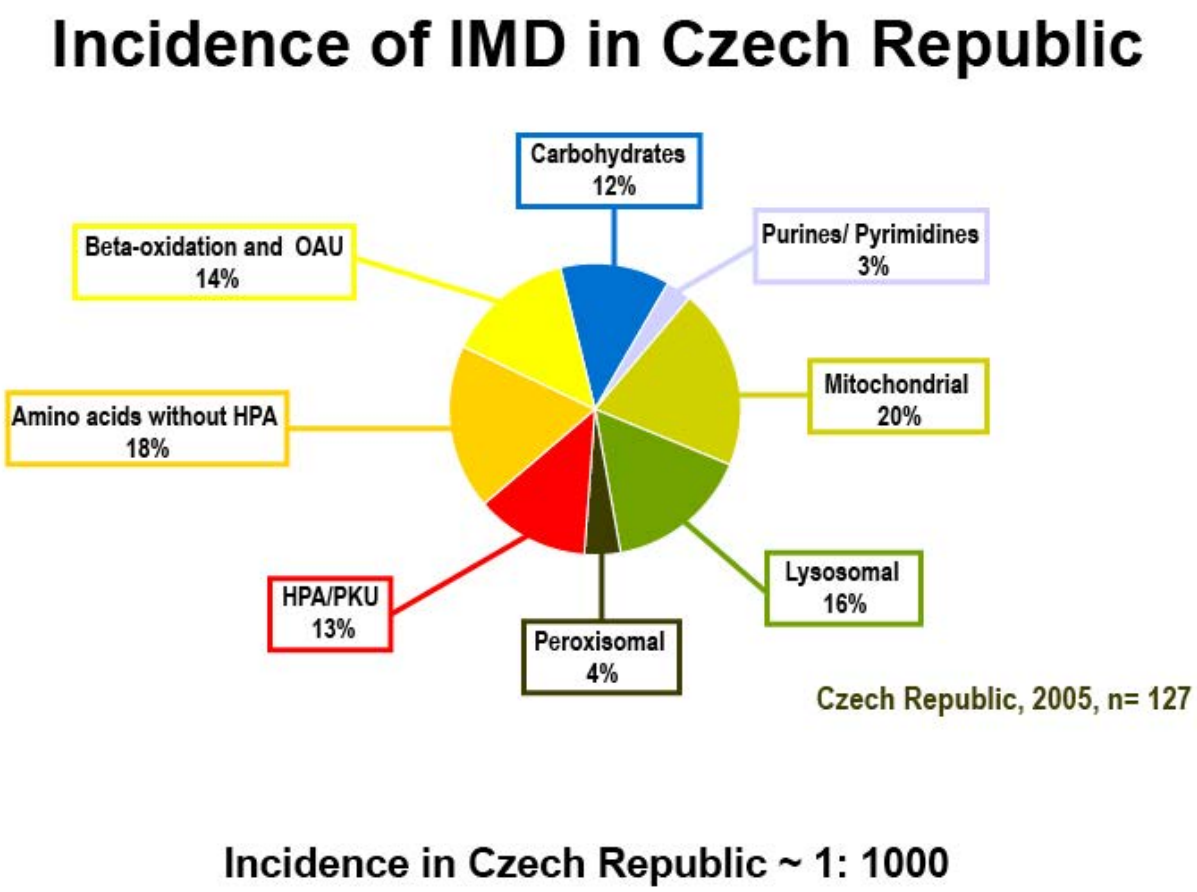


Figure 29.1. Incidence of IMD in Czech Republic

29.3. Pathophysiology

The underlying cause of this group of diseases is nuclear or mitochondrial gene mutations leading to a reduced production and/or changed function of the gene product (most often enzymes, transport or regulatory proteins, receptors, non-enzymatic blood proteins). These changes at the protein level subsequently cause substrate accumulation and insufficient generation of products in the relevant metabolic pathways. The combination of these mechanisms affects various organs, which leads to clinical signs explicable by several typical basic mechanisms.

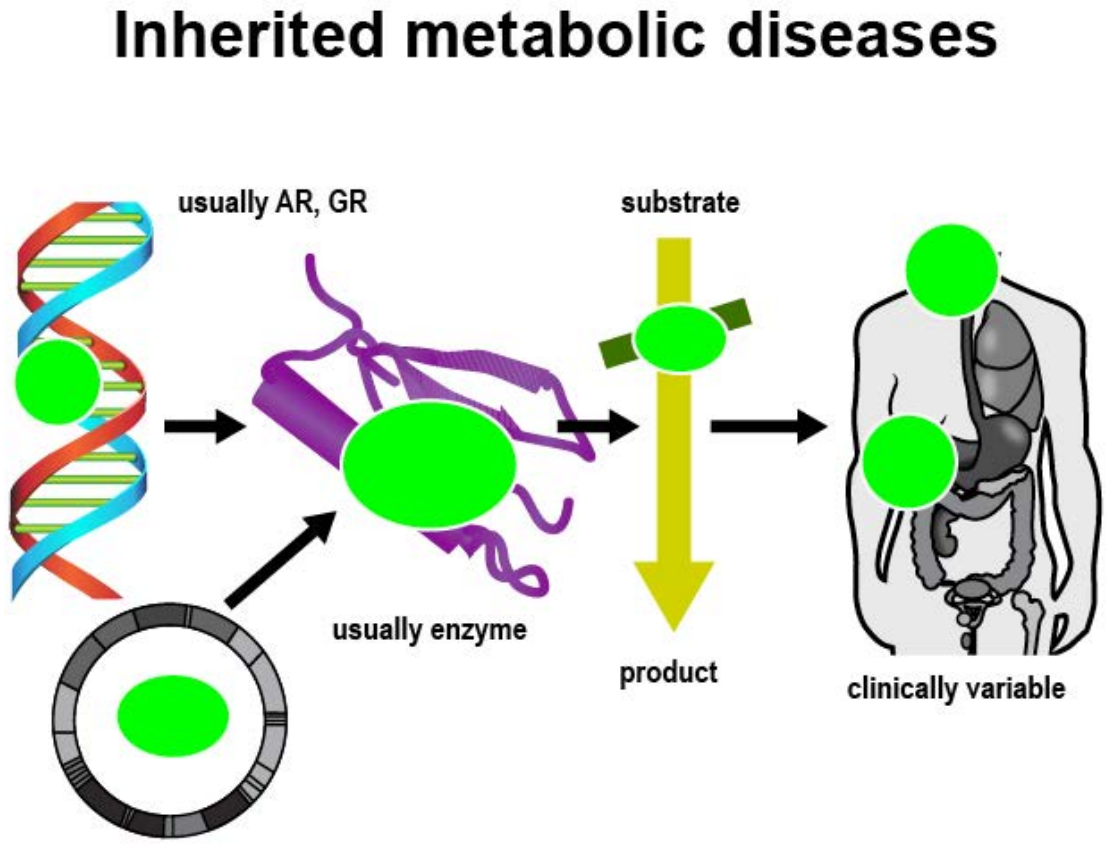


Figure 29.2. Pathophysiology of IMD

Accumulation of substrates and absence of products

The reduced or zero activity of some enzymes leads to a metabolic block with physiological or atypical metabolite accumulation in tissues or body fluids. Substrates of enzymatic reactions accumulate in front of the block creating an absence behind it (analogous changes can be found in the case of the impaired transport of a substance between two compartments separated by membrane). This considerable accumulation of substrate is not to be underestimated (the relevant metabolite concentrations are often several orders higher than concentrations in a healthy individual) – see the schematic illustration in Figure 2. This feature of the considerable accumulation of a certain metabolite (sometimes an atypical metabolite not found in a healthy body) is then used in the diagnosis. Most special biochemistry assays focus on a rapid demonstration of the accumulating metabolites. The predominating pathogenetic mechanism of some IMDs is accumulation of substances (e.g. hyperammonemic coma in patients with urea cycle disorders, attack of metabolic acidosis and unconsciousness in patients with organic acidurias), while symptoms of other diseases occur mainly as a result of the lack of a product (e.g. hypoglycaemic coma in patients with fatty acid beta-oxidation disorders). Another factor forming the clinical presentation of the disease is the physical and chemical properties of substances in the metabolic pathway involved.

Small molecule diseases

Diseases involving small molecules are caused by disorders in the metabolism of substances supplied externally from food (e.g. carbohydrates, amino acids, fatty acids or purines). Most often they manifest themselves as acute problems in the neonatal period or by repeated attacks later in life. Attacks due to the accumulation of highly toxic small molecules typically manifest themselves as slowly progressing disturbances of consciousness, usually ending in coma, and are caused either by an increased intake of amino acids from food, or, much more often, by an increased supply of amino acids in catabolic conditions such as febrile infection. An increased supply of proteins or carbohydrates such as fructose and galactose from food may also be a factor inciting IMDs with acute signs similar to those of sepsis, intoxication or meningoencephalitis. Some patients with organic acid accumulation may have repeated attacks of acetonaemic vomiting and/or metabolic acidosis in their history. Many errors of small molecule metabolism may have subacute to chronic progression. Inherited metabolic disorders of small molecule also include the insufficient generation of intermediate products of energy metabolism (e.g. the release of glucose from glycogen in hepatic forms of glycogenosis or the insufficient generation of ketone bodies in fatty acid beta-oxidation disorders) with the development of clinical

signs of hypoglycaemia following shorter or longer fasting periods. Situations involving higher expenditure of energy by the body are particularly clinically dangerous, especially during fever or increased muscular activity.

Complex molecule diseases

Complex molecule diseases are caused by the impaired generation, elimination or transport of endogenously synthesized complex macromolecules such as mucopolysaccharides or complex glycosphingolipids, or by their decreased synthesis with subsequent impaired production of cell membranes and organelles, such as plasmalogens in peroxisomal diseases. These genetic defects often manifest themselves by abnormalities of the cell membranes and organelles, lysosomes and peroxisomes in particular. When non-degraded complex molecules slowly accumulate, the course of the disease is very often chronic, with onset after an asymptomatic period of several months to decades. Some patients may be monitored for a neurodegenerative disease of unclear aetiology, which may also be similar to a haemato-oncological or other tumour disease. An important role in the pathophysiology of some IMDs is played by the impaired synthesis of complex molecules and cell membranes; these diseases may already manifest themselves after birth as congenital developmental defects, and may imitate chromosomal aberrations. Unlike disorders involving small molecules, these diseases have a progressive course independent of exogenous stimuli and are usually not affected by catabolism or fever. Patients with complex molecule diseases quite often have structural organ involvement with craniofacial dysmorphism, organomegaly or CNS involvement documented by MR examination. Table 1 shows the typical properties of the two basic types of metabolites – small and complex molecules, and also the characteristic signs of each disorder.

	Small Molecules	Complex Molecules
Typical relative molecular weight	< 1500	> 1500
Character of substances	Small, simple molecules with reactive functional groups	Complex molecules or macromolecules, usually containing carbohydrates and lipids
Examples of substances	Amino acids, ammonia, carboxylic acids, simple lipids, monosaccharides including glucose and their phosphoesters, pyrimidines and purines including ATP and others	Glycosaminoglycans (mucopolysaccharides), glycoproteins, complex glycolipids, plasmalogens, glycogen and others
Acute manifestation of disease	Common in highly toxic metabolites	Usually not present
Origin of metabolite	Predominantly exogenous	Predominantly endogenous
Dependence of symptoms on food supply and/or catabolism	Common in highly toxic metabolites	Usually not present
Chronic toxicity and/or chronic progression of disease	May exist in slightly toxic metabolites	Typical symptom
Site of primary disorder within cell	Cytosol, mitochondrion	Membranes, other organelles (lysosomes, peroxisomes)
Site of symptoms within body	Local symptoms in site of enzyme deficiency as well as systemic symptoms in distant organs	Typically only in tissues and organs with enzyme deficiency
Presence of dysmorphism, developmental defects	Rare	Frequent sign
CNS involvement	Qualitative disturbances of consciousness, more or less stationary psychomotor retardation, epilepsy	Typically without qualitative disturbances of consciousness, progressing psychomotor retardation, common abnormalities found by MRI of the brain or electrophysiological examinations
Organomegaly	Relatively frequent isolated hepatomegaly with hepatopathy	Relatively frequent hepatosplenomegaly without hepatopathy, cardiomyopathy
Abnormalities in routine laboratory examinations	Common	Rare

Treatment	Diet and/or pharmacological doses of vitamins, usually effective	Diet and vitamins are ineffective; enzyme replacement therapy and transplantations are possible in some diseases

Table 29.1. IMDs with abnormal metabolism of small and complex molecules

29.4. Classification of IMDs and Characteristics of Basic Groups

The classification of IMDs reflects the growing knowledge of these diseases. These diseases are usually classified by the type of metabolite (e.g. errors of amino acid or carbohydrate metabolism) or the typical diagnostic method (e.g. organic aciduria or hyperammonaemia), or the organelle involved (e.g. peroxisomal, mitochondrial and lysosomal diseases).

Urea cycle disorders

This group of several diseases features a lower capability of the body to transform ammonia into urea; the main pathogenetic mechanism is the accumulation of ammonia leading to brain oedema and coma. These diseases with a peracute and intermittent course are difficult to treat with diet and the administration of substances detoxifying ammonia.

Other disorders of amino acid metabolism

These are a group of several dozen small molecule diseases with variable clinical presentation, depending on the character of the accumulated metabolite, usually with a chronic course. Most of these diseases can be treated quite well with the relevant amino acid-controlled diet or administration of vitamin cofactors.

Organic acidurias

This roughly defined group of several dozen diseases featuring a demonstrably elevated urinary concentration of carboxylic acids coming mostly from the carbon skeleton of amino acids. Typically these are small molecule diseases with an accumulation of highly toxic substances and encephalopathy, and usually also metabolic acidosis; they have a peracute course in the neonatal period and decompensation attacks later in life.

Fatty acid beta-oxidation disorders

This group of more than 10 diseases is caused by the impaired transport of carnitine derivatives of fatty acids or their mitochondrial beta-oxidation; these are small molecule diseases for which the dominating mechanism is the insufficient generation of ketone bodies following prolonged fasting with subsequent hypoglycaemia or intolerance of muscular effort. Some diseases may also include toxic manifestations of some accumulated unprocessed fatty acids presented as myopathy or cardiomyopathy.

Disorders of gluconeogenesis

This group of several disorders of carbohydrate and glucogenic amino acid metabolism manifest themselves by fasting hypoglycaemia.

Disorders of monosaccharide metabolism

Disorders of galactose and fructose processing manifest themselves as acute small-molecule IMDs, particularly by hepatopathy and generalized tubulopathy.

Liver glycogenoses

This is a group of 7 types of glycogen storage diseases (GSDs) with impaired glycogen degradation, gluconeogenesis and glycogen synthesis.¨

Muscle glycogenoses

This is a group of 9 types of GSDs with symptoms of myopathy and/or cardiomyopathy.

Disorders of purine and pyrimidine metabolism

A group of more than 15 chronic small molecule diseases with predominant neurological and psychiatric symptoms; some diseases feature the accumulation of poorly soluble urates.

Lysosomal diseases – mucopolysaccharidoses and glycoproteinoses

These diseases usually have a chronic course. They are caused by the accumulation of complex glycoconjugates in lysosomes and are one of the most common groups of IMDs. Clinically they manifest themselves as a multi-system disease with connective tissue involvement, hepatosplenomegaly, typical facial dysmorphism (gargoylism) and different levels of CNS involvement. The course of some mucopolysaccharidoses can be slowed down by enzyme replacement therapy; haematopoietic stem cell transplantation may be considered for some patients.

Lysosomal diseases – lipidoses and other diseases

These diseases are due to the lysosomal accumulation of complex glycolipids and constitute one of the most common groups of IMDs. They usually clinically manifest themselves as a multi-system disease with a chronic course. Some lipidoses feature parenchymatous organ involvement (spleen and liver in Gaucher's disease and kidneys in Fabry's disease), while in others different levels of central and peripheral nervous system involvement are dominant – from psychiatric signs to epilepsy. Some patients with lipidosis may be treated successfully with enzyme replacement therapy or substrate reduction therapy.

Peroxisomal diseases

This group of IMDs is caused by a disorder of peroxisome biogenesis with generalized impairment of peroxisomal functions (complex molecule diseases such as Zellweger syndrome) or isolated impairment of some peroxisomal function (e.g. X-linked adrenoleukodystrophy).

Disorders of mitochondrial energy metabolism

This clinically and biochemically heterogeneous group of diseases is caused by impairment of the respiratory chain enzyme function. Mitopathies are a prototype of diseases caused by deficient production of small ATP molecules. The pathogenesis also involves increased production of free radicals and reactive forms of oxygen.

Disorders of glycosylation (CDG syndromes)

Carbohydrate deficient glycoprotein (CDG) syndromes are a heterogeneous group of multi-system large molecule diseases with impaired protein glycosylation and many somatic alterations found during physical examination.

29.5. Clinical Symptoms and Indications for IMD Examination

The clinical symptoms of IMDs are extremely varied and non-specific. First clinical signs may manifest themselves at any age (from the neonatal period through childhood to adult age, but may also be manifested prenatally by *hydrops fetalis*) – Figure 32.3.

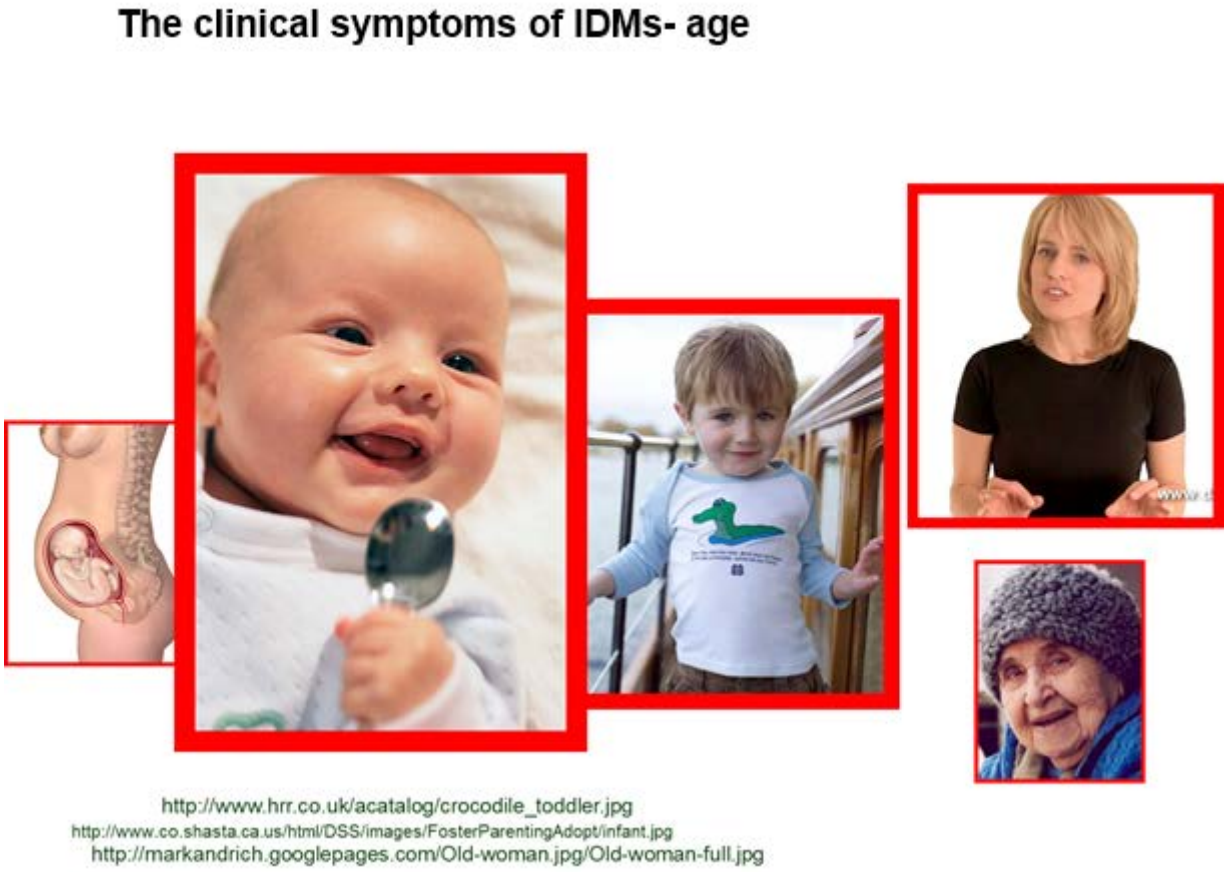


Figure 29.3. The clinical symptoms of IDMs- age

Inherited metabolic diseases may manifest themselves by the abnormal function and impaired structure of almost all organs and tissues, and may do so in a great variety of combinations – Figure 32.4.

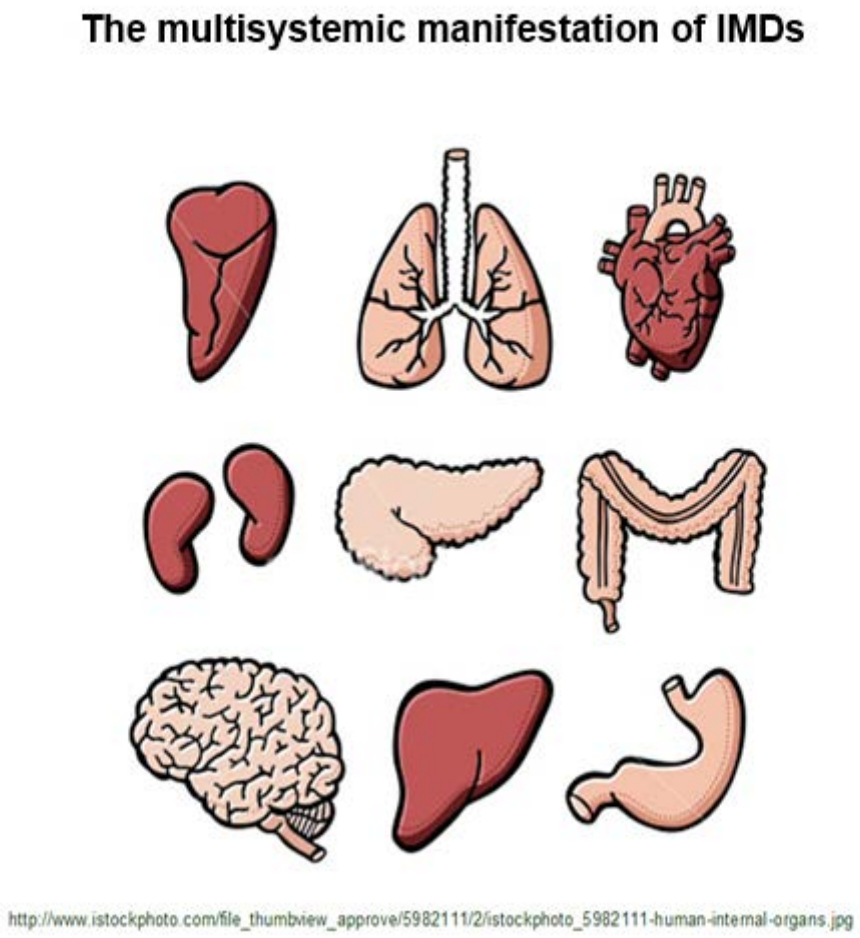


Figure 29.4. The multisystemic manifestation of IMDs

This is the reason that clinical symptoms can be seen not only by paediatricians and internists, but also by physicians of many other medical specializations (neurology, nephrology, haematology, ophthalmology, dermatology, gynaecology, psychiatry and others). The course of the IMD can be stationary, progressing or acute (with peracute or intermittent character). It is absolutely clear from the above that any suspicion of IMD may only be raised after a thorough assessment of the history, clinical condition and routine laboratory and imaging examinations. At present there are no simple rules that could help the attending physician indicate specialized laboratory examinations. Table 2 shows a non-exhaustive list of examples of some of the most common clinical situations, typically requiring differential diagnosis specific to IMDs. In addition, Table 3 shows laboratory findings that may be used as an evidence of presence of the IMD in the patient.

Relation between signs and diet – aggravation following protein, fructose or galactose intake
Manifestation of signs during increased output of energy – especially during fasting and increased muscular activity
Relation between signs and febrile infection, when protein catabolism increases
Multi-system signs
Evidence of unusual metabolite accumulation Abnormal colour of urine Abnormal odour of urine and sweat Insoluble substances in urine – crystalluria, urolithiasis
Selected deviations found in routine laboratory examinations – see Table 30.4

Table 29.2. Some auxiliary signs arousing suspicion of IMD

Urine examination
Abnormal odour may suggest the presence of small volatile molecules: Sweaty feet (isovaleric aciduria) Caramel/maple syrup (leucinosis) Stewed cabbage (hypermethioninemia) Fishy smell (trimethylaminuria) Mousy smell (phenylketonuria)
Abnormal colour of urine or nappies: Red to orange (urates – disorders of purine metabolism) Brown to black when left standing for a period of time (alkaptonuria) Red to brown (myoglobinuria) Blue (Hartnup disease) Green (4-OH-butyric aciduria)
Crystalluria: the presence of crystals listed below may suggest an IMD Cystine 2,8-dihydroxyadenine Xanthine Urates Oxalates
Urine pH: Elevated pH in absence of urinary tract infection (renal tubular acidosis)
Ketone bodies: Increased excretion of ketone bodies with concurrent metabolic acidosis, particularly in newborns (organic aciduria) Absence of or mild ketonuria with severe hypoglycaemia (fatty acid beta-oxidation disorders)
Uric acid: Increased daily excretion (disorders of purine metabolism) Decreased daily excretion (xanthine oxidase deficiency)
Myoglobin: Exercise-induced myoglobinuria (disorders of fatty acid beta-oxidation and muscle glycogenosis)
Blood tests
Glucose level: Fasting hypoglycaemia with ketonuria (liver glycogenoses, disorder of gluconeogenesis, mitochondrial diseases) Hypoketotic hypoglycaemia from prolonged fasting and catabolic condition (disorders of fatty acid beta-oxidation) Hypoglycaemia after ingestion of fructose/saccharose/sorbitol (intolerance of fructose)
Ammonia: Hyperammonaemia (urea cycle disorders, some organic acidurias, disorder of ATP synthesis)

Acid-base balance tests (Astrup): Metabolic acidosis (accumulation of organic acids – organic aciduria, lactate acidosis in mitochondrial disorders, loss of bicarbonate in generalized tubulopathy) Respiratory alkalosis (hyperventilation in hyperammonaemia due to urea cycle disorders)
Aminotransferases/bilirubin: Hepatopathy is a non-specific sign of liver parenchyma involvement in many small-molecule IMD types.
Creatinine: Decreased concentrations may occur in disorders of creatine biosynthesis Increased concentrations occur in IMDs affecting renal functions
Cholesterol: Decreased concentrations can be found in Smith-Lemli-Opitz syndrome and mevalonate kinase deficiency Increased concentrations can be found in many IMDs
Triacylglycerols: Pseudotriacylglycerolaemia (without any corresponding finding in lipoprotein electrophoresis) is indicative of glycerol kinase deficiency Increased triacylglycerol concentrations are a secondary sign in many IMDs
Alkaline phosphatase Decrease (hypophosphatasia) Increase (secondary in hepatopathy or tubulopathy accompanying some IMDs)
Uric acid: Decreased concentration (isolated or combined xanthine oxidase deficiency, hereditary renal hypouricemia) Increased concentration (disorders of purine metabolism or renal transport)
Creatine kinase and myoglobin: Increased concentrations (mitochondrial disease, type II glycogenosis) Exercise-induced transitory increase (some disorders of fatty acid beta-oxidation, muscle glycogenosis and some disorders of purine metabolism)
Lactate: Increased concentration (provided pre-analytical phase conditions are met) in absence of hypoxia may suggest mitochondrial diseases and some glycogenoses
Homocysteine: Hypohomocysteinaemia occurs in sulphite oxidase deficiency Pronounced hyperhomocysteinaemia in remethylation and transsulfuration forms of homocystinuria
Blood count: Macrocytic anaemia (disorders of homocysteine, folic acid and cobalamin metabolism, orotic aciduria) Thrombocytopenia (Gaucher's disease)

Haemocoagulation: Prolonged prothrombin time (disorders of glycosylation – CDG syndromes) Severe hypoproductive coagulopathy (some IMDs affecting the liver)
Cerebrospinal fluid examination
Glycorrhachia: Hypoglycorrhachia (disorder of glucose transport)
Total protein: Pronounced hyperproteinorrhachia (some lysosomal diseases such as lipidoses)
Lactate: Increased concentration (disorders of mitochondrial energy metabolism)

Table 29.3. Evidence for IMD from routine laboratory examinations

29.6. Diagnosis of IMDs

Since clinical symptoms are very varied and non-specific, it is possible to diagnose IMDs only using special laboratory examinations carried out by a specialized laboratory of biochemistry and genetics. There are two fundamentally different approaches to effective diagnosis:

a) Neonatal screening of the entire population of asymptomatic newborn infants;

b) Selective screening of a small selected group of symptomatic patients with suspected IMD.

29.6.1. Neonatal screening

Neonatal screening in the CR is performed in every newborn child born 48 - 72 hours after birth. A dried blood spot on a neonatal screening card is tested by tandem mass spectrometry. *(Instructions for blood collection, laboratory operations and aftercare are specified by "Metodický návod k zajištění celoplošného novorozeneckého laboratorního screeningu a následné péče" – see Věstník MZ ČR 2009, Issue 6).* This method is used in screening for 13 different IMDs out of about the 800 known.

- Disturbances of amino acid metabolism
  - Phenylketonuria
  - Glutaric aciduria, Type I (glutaryl-CoA dehydrogenase deficiency)
  - Isovaleryl-CoA dehydrogenase deficiency (Isovaleric acidemia)
  - Maple syrup urine disease
- Disorders of fatty acid oxidation
  - Carnitine uptake/transporter defects
    - Carnitine-acylcarnitine translocase deficiency
    - Carnitine palmitoyl transferase I deficiency (CPT I)
    - Carnitine palmitoyl transferase II deficiency (CPT II)
  - Very long chain acyl-CoA dehydrogenase deficiency (VLCADD)
  - Long chain L-3 hydroxyacyl-CoA dehydrogenase deficiency (LCHADD)
  - Medium chain acyl-CoA dehydrogenase deficiency (MCADD)

Despite the fact that most patients suffering from some of the IMDs being screened are detected, it should also be noted that mild forms of the disease are not necessarily detectable by the screening methods used several days after birth. Therefore, a potential IMD also has to be considered in patients with negative screening results if clinical signs of the disease begin to be manifested at a later time.

Selective screening for IMDs

Neonatal screening detects only a small proportion of patients with IMDs. Most patients are diagnosed worldwide in the selective screening system. Two factors play a role in the effective use of screening. The first is the availability of screening methods. Foreign experience suggests that it is useful to combine the methods in specialized centres where both screening tests and follow-up targeted tests for enzymes and molecules are available. The second factor is the effective indication of specialized examinations. The most critical step in selective screening is the expression of the suspicion of an IMD by the first-contact physician or specialist as part of their opinion based on differential diagnosis. This group of diseases is often considered only after other causes of the disease have been eliminated, which means the correct diagnosis of IMD is established late. An indication for selective screening is given by two fundamental aspects:

a) Positive family history

b) Clinical signs (including available laboratory examinations)

The problem is that some metabolic disorders often have non-specific clinical presentation and a relatively low incidence in the population. Therefore, it is not surprising that present clinical knowledge about metabolic diseases often lags behind knowledge of biochemistry and genetics.

For these reasons it is also quite common that staff from a specialized facility (biochemistry and genetics laboratories for IMDs) provide the attending physician with a metabolic test interpretation, including a complete consultation and suggestions for further examinations and treatment. New defects from this group of diseases are also studied and discovered in specialized laboratories.

Given the great clinical variability of IMDs, there are no simple criteria, which, if met, would clearly point to a suspicion of these diseases. Nevertheless, there are some clinical situations where suspected IMD should be part of the early phases of standard differential diagnosis! Some selected common situations are listed in Tables 2 and 3, and more detailed information about IMD symptoms can be found in the literature (see Recommended Reading). As regards family history, it is important to note that it does not always help establish the indication. Such cases include a first child with IMD with autosomal recessive inheritance.

The diagnosis of IMD is usually established gradually, in several stages involving specialized examinations of blood and urine metabolites, enzyme activities in blood cells or fibroblast cultures, or gene mutation tests. The laboratory diagnosis process takes place at the level of metabolites, enzymes/proteins, tissues/cells and DNA.

Metabolite assays include semiquantitative, quantitative and profile tests.

Enzyme screening tests can be indicated without prior agreement.

Targeted confirmatory enzyme and molecular genetics tests are made only by prior agreement with a specialized laboratory.

Tissue examination is also made by prior agreement. The most commonly collected specimens include blood (blood smear), bone marrow (bone marrow smear), skin (also for fibroblast culture), liver or muscle, and, less commonly, myocardium, conjunctiva or other tissues. Biopsy as an invasive intervention should be made only after a thorough differential diagnosis analysis, complete non-invasive examinations and following consultation with the laboratory where the biopsy specimen is going to be examined. This is the only way of avoiding a repeated biopsy.

Indications for biopsy/autopsy is skin biopsy/autopsy for fibroblast culture (collection and transport in the culture medium or, in emergency, sterile saline), receipt of material for enzyme assay (if a peripheral blood cell assay is not possible or reliable), histological, histochemical and electron microscopy examination (if required for the diagnosis), molecular genetics tests (mitochondrial disorders, nuclear genes, or if a peripheral blood cell assay is not possible or reliable), chorion biopsy or amniotic fluid collection for prenatal IMD diagnosis; also, post-mortem collection of tissue specimens (autopsy) is required if IMD is suspected in the case of sudden, unexpected or unavoidable death. The basic condition for adequate examination of a tissue specimen is **proper specimen processing after collection and proper specimen transport to the laboratory**.

29.7. Options for IMD Treatment and Prevention

As present knowledge develops, the number of successfully treatable hereditary metabolic disorders is growing. Some form of therapeutic action is nowadays possible in about one third of patients with IMDs, and treatments for many other IMDs are under research. This relatively high amount of diseases without available causal therapy is due to

the fact that the pathogenesis of these defects is not known, despite the knowledge of clinical symptoms and causes at the level of molecular genetics. Haemodialysis or haemodiafiltration is used to eliminate toxic products in acute phases of IMD. Dietotherapy with a reduced supply of proteins supplemented with a mixture of selected essential amino acids in dietary preparations is used for the long-term maintenance of metabolic compensation of amino acid metabolism. A frequent diet with low fat content is recommended for disorders of fatty acid beta-oxidation, while a low-sugar diet is used for disorders of pyruvate dehydrogenase metabolism. The administration of high doses of specific vitamins or pharmacological chaperones is used to stabilize mutant enzymes in the treatment for some IMDs. A proportion of patients with lysosomal diseases responds favourably to parenteral administration of recombinant enzyme, whereas transplantation of haematopoietic stem cells or entire organs such as the liver, kidneys or heart is used for treatment of other lysosomal disorders. The effectiveness of treatment is different for different patients and diseases; effective treatment with minimum side effects enables the patient to live a normal life (e.g. administration of vitamins in vitamin-sensitive IMDs).

Treatment for small-molecule IMDs in many patients includes dietary restrictions of substances that the patient cannot process, or an increased supply of the substance that the patient cannot produce in his/her body. Some patients may favourably respond to the administration of vitamins acting as cofactors for enzymatic reactions. Elimination treatment by haemodialysis or haemodiafiltration is a life-saving method in cases of acute attacks of the disease. As the cause of IMDs is genetic, primary prevention in the afflicted individual is not possible. Secondary prevention consisting in early diagnosis of the relevant disease type and timely treatment (if available) is very important. Neonatal screening programmes are a prime example of successful secondary prevention. Except for mitochondrial inherited diseases, the overwhelming majority of IMDs can be diagnosed in at-risk families by prenatal examination; the development of preimplantation diagnosis can be expected in the near future.

CHAPTER 30

30. Laboratory Test for Urolithiasis

Author: RNDr. Jiří Zadina, CSc.

Reviewer: doc. MUDr. Vladimír Soška, CSc.

30.1. Characteristics of Urinary Concrements

Urolithiasis is a disease characterized by the formation of concrements in the kidneys and the urinary tract. The formation of concrements is largely affected by biochemical and physiological factors (insufficient fluid intake for an extended period, food composition), genetic factors - especially in young individuals, and many cases of urolithiasis are affected by urinary tract infection or some diseases (hyperparathyroidism, cancer).

The process of lithogenesis results from the interaction of many factors:

- High concentration of lithogenous substances in the urine: hypercalciuria, hyperoxaluria, hyperuricosuria, hyperphosphaturia, cystinuria
- Presence of organic components forming nuclei of crystallization
- Urine pH
- Concentration of lithogenesis inhibitors: primarily magnesium and citrate

Substances found in urinary concrements are listed in Table 1. Since concrements are crystalline in nature, mineralogical nomenclature is used.

Compound:	Mineralogical Name:
Uric acid	Uricite
Uric acid $2H_2O$	
Ammonium urate	
Sodium urate H_2O	
Calcium oxalate H_2O	Whewellite
Calcium oxalate $2H_2O$	Weddellite
Calcium phosphate (pure)	Whitlockite
Calcium phosphate (apatite structure)	Apatite
“Carbonate apatite”	Dahllite
Calcium hydrogen phosphate $2H_2O$	Brushite
Magnesium ammonium phosphate $6H_2O$	Struvite
Protein	
Xanthine	
Cystine	
2.8-dihydroxyadenine	
Artefact/counterfeit of unknown origin	

Table 30.1. Components and mineralogical names of urinary concrements

Urinary concrements mostly occur in the form of a mixture. Two-component concrements are prevalent (about 50%); single-component (about 30%) and three-component concrements (about 20%) are less common.

Urinary concrements are classified into several groups:

- Calcium oxalate concrements are the most common type of concrements (about 60% in the Czech popu-

lation).They consist of a mixture of whewellite and weddellite (40%), pure whewellite (15%) or a mixture of oxalate and uric acid (5%). The causes of their formation include hypercalciuria, hyperoxaluria, hyperuricaemia, hyperuricosuria, hypocitraturia and primary hyperparathyroidism.

- Calcium phosphate concrements (about 10%) usually occur in a combination of apatite with oxalates. Pure apatite or brushite concrements occur sporadically. Possible causes of formation of these stones include hypercalciuria, hyperphosphaturia, primary hyperparathyroidism or renal tubular acidosis.
- Urate concrements (about 15%) consist primarily of uric acid (anhydride and dihydrate); other urates are rare. Urate concrements are the result of hyperuricaemia and hyperuricosuria, low diuresis and acidic urine pH.
- Infectious concrements (about 15%) consist of struvite, dahllite and sporadically ammonium urate. They occur in chronic urinary tract infections caused by urea-degrading bacteria, which is related to urine alkalinization.
- Cystine concrements (less than 1%) occur in homozygous patients with cystinuria. Cystine crystallizes primarily in acidic urine.
- The incidence of each type of lithiasis differs according to geographical region. As in the CR, calcium oxalate concrements dominate in Germany, while calcium phosphate concrements are most widespread in France, Sweden and the US.

30.2. Laboratory Diagnostics of Urolithiasis

Biochemical assays focus on the search for risk factors in the formation of concrements, and are used to monitor pharmacotherapy and dietary therapy and to establish the risk of relapse. The following analytes are tested in the serum: Ca, Mg, inorganic phosphate and uric acid. The first step of urine examination is the chemical assay of urine sediment and urine. The quantitative assay for urinary wastes includes the following analytes: Ca, Mg, inorganic phosphate, uric acid, creatinine, oxalate, and sometimes citrate. Urine samples for the analysis of Ca, phosphates, oxalates and citrates should be acidified by adding HCl to avoid precipitation of salts, which would lead to falsely lower or negative findings. Alkalinization of urine specimens for a uric acid assay is required for the same reason. Additional special tests for hyperparathyroidism, renal tubular acidosis or a challenge test for hypercalciuria differentiation are made depending on the type of disorder detected.

30.3. Analysis of Urinary Concrements

Knowledge of the composition of the concrement is important for finding the metabolic background of urolithiasis. Previously used methods of chemical analysis have been replaced by more precise physical methods such as infrared spectroscopy, X-ray diffraction and polarized light microscopy. Examination under a polarizing microscope is based on the different optical properties of different crystalline substances in polarized light. The method of infrared spectroscopy is used to indentify compounds based on spectra, assessed quantitatively using a computer program by comparing them with a library of known urinary concrement spectra. A very small amount of concrement (30 mg) is sufficient for the assay. The accuracy of the assay is subject to an error of about 10%; the 10 - 15% range of analytical error is clinically unimportant. Each concrement layer should be analyzed in justified cases to clarify the genesis of the concrement.

30.4. Case Reports

30.4.1. Case Report 1

A 61-year-old female incontinent patient in the long-term care of a hospital for terminal illnesses with a history of ethylic encephalopathy, polyneuropathy and hepatomegaly. A gynaecological examination indicated due to a discharge showed that the discharge has a primary urological aetiology. One morning, the patient found in her bed an object the size and shape of a small chicken egg, which was sent to the department of biochemistry for urinary concrement examination.

Result of examination:

Urinary concrement of grey to white colour, smooth surface, oval shape, size 47 x 36 x 29 mm, very hard. Overall composition: Struvite 100%.

Evaluation:

It is urinary concrement of infectious origin, located in the urinary bladder. The slackened sphincter and urethra permitted the spontaneous passage and expulsion of the large-sized object. Considering the history and immobility, the patient did not experience any problems.

30.4.2. Case Report 2

A 57-year-old male patient, first cystine concrement found in the right-hand kidney at the age of 23, the stone was removed surgically by pyelolithotomy. Later a relapse of the concrement in the left, but the patient did not have surgery nor metaphylaxis of the lithiasis. Currently, a complicated “casting” lithiasis in the left. The patient underwent repeated percutaneous extractions of the concrement. Removed concrements were analyzed in the department of biochemistry.

Result of examination:

1. Analysis: A large amount of similar pieces of light brown colour, slightly crystalline surface, irregular shape, sizes from 10 x 8 x 3 to 5 x 4 x 2 mm, very hard. Overall composition: dahllite 80%, whitlockite 20%.

2. Concrement: A large amount of pieces of light brown colour, crystalline surface in places, irregular shape, sizes from 10 x 7 x 7 to 3 x 3 x 3 mm, medium hard. Individual crystals of cystine are clearly visible under the polarizing microscope. Overall composition: Dahllite 40%, whitlockite 30%, cystine 30%.

Evaluation:

Cystine metabolism-dependent lithiasis. Since the appropriate diet and hydration were not followed, a relapse occurred because cystine crystal residues served as nuclei for crystallization of other components. Infection also took part in the formation of the concrement.

The region with cystine crystals was not removed during the first surgery, so cystine could not have been detected.

30.4.3. Case Report 3

A 63-year-old female patient, severely obese. The patient had problems with urolithiasis for many years, the first concrement was diagnosed in the left kidney 15 years ago. Underwent repeated extracorporeal shock wave lithotripsy. The patient is under long-term care of the metabolic unit due to urolithiasis, is administered metaphylaxis, has enough exercise but overeats.

Result of examination:

The evaluation included 5 fragments corresponding to the originally oval concrement, sized about 14 x 9 x 8 mm with visible layers. Individual layers were analyzed by means of polarizing microscope:

a) Nucleus: 5 x 4 x 3 mm, grey to white, composition: uricite (maybe with whewellite admixture - unable to distinguish).

b) Middle layer: about 2 – 3 mm, dark brown to black, orange after crushing, very hard, composition: whewellite.

c) Envelope: about 1 mm, orange, low hardness, composition: uricite

Overall composition: Whewellite 50%, uricite 50% (expresses average proportions in the entire concrement).

Evaluation:

Oxalate concrement formed by urine oversaturation with lithogenic substances as a result of their excess supply in the food, which is confirmed by results of supplementary biochemical tests:

Serum – glucose level 7.0 mmol/l, other analytes normal.

Urine – diuresis 1840 ml/day, pH 5.1, hypercalciuria 9.35 mmol/day (normal under 7.2), hypernatriuria 364 mmol/day (normal under 220), elevated chloride output 381 mmol/day (normal under 270), slight hyperoxaluria 546 umol/d (normal under 500), corrected creatinine clearance

1.86 ml/s/1.73 m2 is within the normal range.

A detailed analysis of each layer may help clarify the formation and course of concrement depositing; this case probably reflects changes in the dietary habits of the patient.

Test questions: (correct answers)

1. The name whewellite is used for a urinary concrement composed of:

- Calcium oxalate$2H_2O$
- Calcium hydrogen phosphate$2H_2O$
- *Calcium oxalate H_2O*
- Magnesium ammonium phosphate$6H_2O$

2. Which urine pH is prevalent when phosphate concrements are formed:

- pH 5
- *pH 6*
- *pH 7*
- pH 8

3. Urinary tract infection leads to the formation of infectious concrements; these include:

- Struvite
- Brushite
- Carbonate apatite
- Weddellite

CHAPTER 31

31. Laboratory Examinations during Pregnancy

Author: Ing. Drahomíra Springer, Ph.D.

Reviewer: prof. MUDr. Jaroslav Racek, DrSc.

31.1. Introduction

Pregnancy is connected to a change in some biochemical processes to adjust to the development of the foetus. The maternal organism also undergoes many hormonal changes. Some changes are physiological, others may be a sign of problems in the maternal organism or the developing foetus.

Over the last few years, prenatal diagnosis has become a standard aspect of care given to pregnant women. This is important not only from the individual's point of view, but also from society's perspective, considering the economic and social impacts. It is in the interest of parents and medical professionals to be well informed about possibilities of various preventive programmes, to be aware of what modern medicine is able to offer, and also to know at least the basic principles of preventive care. Prenatal diagnosis includes a series of examinations made in order to detect potential developmental defects of the still unborn child. This diagnosis is based on interdisciplinary cooperation of specialists from medical genetics, gynaecology and obstetrics, ultrasonic diagnosis, clinical biochemistry and other branches of medicine.

31.2. Laboratory Examinations during Pregnancy

31.2.1. Confirmation of Pregnancy

Human chorionic gonadotropin (hCG) is a hormone produced by a pregnant woman's placenta immediately after conception, supporting the production of oestrogens and progesterone in the corpus luteum. The hCG level rises on day 8 to 10 after conception. Its quantity is determined from the serum and is used to confirm pregnancy. A qualitative analysis is performed using urine – the pregnancy test indicates levels over 20 IU/l. Absolute values may vary individually during pregnancy; a rising trend is always important (hCG doubles every two to three days at the beginning of pregnancy). The trend can suggest a failing pregnancy, a drop may be indicative of a missed abortion. Repeated hCG blood tests are important for diagnosing ectopic pregnancy; the rise in hCG is usually less steep. If the increase in hCG concentration is less than 60% over 48 hours, it is an ectopic pregnancy or death *in utero*. HCG tests are also useful in diagnosing trophoblastic diseases, or as a germinal tumour marker; an elevated hCG level may also be caused by myomas and ovarian cysts.

31.2.2. Examinations during Pregnancy

A pregnant woman is admitted to counselling care organized by the attending gynaecologist. Examinations are performed based on the state of health of the pregnant woman and the length of pregnancy, while the extent of examinations must comply with the professionally recommended minimum. The Czech Gynaecological and Obstetrical Society of the Czech Medical Society of J. E. Purkyně (ČLS JEP) has defined the basic frequency and extent of care for women during the physiological course of pregnancy. Based on the medical history and current clinical findings, pregnancy is rated by the degree of risk for the pregnant woman or developing foetus, and the frequency and types of examinations are adjusted accordingly. A family history of diabetes, congenital defects in the family, the outcome of a previous pregnancy, high blood pressure and similar factors may suggest potential future risks for the pregnancy, which then have to be monitored. Some examinations are made at every visit to the pregnancy counsellor – blood pressure, signs of foetus vitality, urine chemistry, etc. Other examinations are referred to as irregular. They are made in a specific week of pregnancy and include blood group, rhesus-D status or serological assays for specific antibodies. Table 1 shows

the frequency and types of examinations.

The screening for **gestational diabetes mellitus** (GDM), associated with perinatal short-term and long-term complications, is made by oral glucose tolerance test (oGTT) in all healthy pregnant women according to an established schedule between week 24 and 28 of pregnancy. According to the Recommendations of the Czech Diabetes Society (ČDS) and the Czech Society of Clinical Biochemistry (ČSKB), a load by 75 g of glucose is used and the plasma glucose level is assessed before and two hours after the challenge. Gestational DM is diagnosed from laboratory results if the plasma glucose level (FPG) is repeatedly ≥ 5.6 mmol/l or ≥ 7.7 mmol/l two hours after the load.

Week of Pregnancy	Recommended Laboratory Tests
11 - 13	Blood group, rhesus-D status, blood count, screening of irregular anti-erythrocyte antibodies, HBsAg, HIV, syphilis, fasting glycaemia. Also advisable are fT4, TSH and antiTPO tests.
11 - 13	DS screening in 1st trimester (PAPP-A, free β hCG, NT)
16 - 18	DS screening in 2nd trimester (AFP, hCG, uE3)
24	Gestational DM screening
28	Screening of irregular anti-erythrocyte antibodies in rhesus-D negative women, blood count, syphilis antibodies
36 - 38	Vaginal infection detection

Table 31.1. Overview of laboratory tests in specific weeks of pregnancy

Irregular examinations also include an assessment of the risk of congenital chromosomal aberration (CHA) of the foetus.

31.2.3. Diagnosing Gestational Diabetes

The screening for **gestational diabetes mellitus** (GDM), associated with perinatal short-term and long-term complications, is made by oral glucose tolerance test (oGTT) in all healthy pregnant women according to an established schedule between week 24 and 28 of pregnancy. According to the Recommendations of the Czech Diabetes Society (ČDS) and the Czech Society of Clinical Biochemistry (ČSKB), a load of 75 g of glucose is used and the plasma glucose level is assessed before and two hours after the load. Gestational DM is diagnosed from laboratory results if FPG is repeatedly ≥ 5.6 mmol/l or ≥ 7.7 mmol/l two hours after the load.

31.2.4. Changes in Haematological Parameters

Water retention increases during pregnancy. Qualitative and quantitative metabolic changes take place in the body as a result of the rapidly growing foetus and placenta as well as their increasing nutritional requirements. Metabolic changes are influenced by a change in the production and secretion of some maternal hormones and the production of new hormones in the foetoplacental unit, and, last but not least, also by the transplacental transfer of substances. O_2 consumption and the basal metabolic rate increase during pregnancy, by up to 15–20% in the 3rd trimester.

The volume of plasma increases with the size of the foetus – the greater the foetus, the greater the plasma volume. The volume also increases depending on the woman's height, parity and multiple pregnancies. As the plasma volume doubles compared to the red blood cell mass volume, the concentration of haemoglobin and haematocrit decreases and dilutional anaemia develops.

The absorption of Fe from food increases from the normal 5–10% up to 30% in the 2nd trimester, and up to 40% in the 3rd trimester of pregnancy. The average daily loss of Fe also decreases to roughly one half. Fe is transported by active transport through the placenta in one direction only – from the mother to the foetus, also from mothers with an Fe deficiency.

The overall white blood cell count increases up to 12.0×10^9/l during pregnancy; values of up to 16.0×10^9/l have been found in the 3rd trimester. The white blood cell count is up to $25–30 \times 10^9$/l during delivery. The lymphocyte and monocyte count is relatively stable; the polymorphonuclear cell count increases. The basophil granulocyte count decreases slightly. The eosinophil count increases in some pregnant women, or decreases or stays unchanged in others; the cause is unknown.

The erythrocyte sedimentation rate (FW test) increases to 44 to 114 mm/h during pregnancy, 78 mm/h on average.

The high ESR is due to increased plasma globulin and fibrinogen levels. ESR is not used for diagnosis during pregnancy.

31.2.5. Changes in Biochemical Parameters

Enlarged kidneys and a dilated excretory system are typical of pregnancy. Renal functions change during pregnancy, mainly due to the increased production and secretion of some maternal and placental hormones (ACTH, ADH, aldosterone, cortisol, hPL) and due to the increased volume of plasma. While glomerular filtration markedly increases during pregnancy, the amount of urine does not increase. In this sense, the excretory system is more efficient during pregnancy than in non-pregnant women. Plasma urea, creatinine, uric acid, glucose and amino acid clearance grows and their levels decrease during pregnancy. In the 1st trimester, creatinine concentration drops from 73 to 65 µmol/l, then to 51 µmol/l in the 2nd trimester, and is roughly 47 µmol/l at the end of pregnancy. Plasma urea concentration decreases as a result of a lower elimination of proteins and elevated clearance. Respecting these physiological changes is important for timely detection of renal insufficiency during pregnancy. If the creatinine level at an advanced stage of pregnancy is over 75 µmol/l and urea is over 5.4 µmol/l, supplemental tests of renal functions are necessary.

No morphological changes of the liver occur during normal pregnancy, only function changes. Alkaline phosphatase activity and the production of cholesterol, globulins and fibrinogen increase. Non-protein nitrogen and uric acid levels decrease. Concentrations of bilirubin, transaminases and bile acids remain unchanged. The overall fat level in the plasma increases from 6 g/l to 10 g/l during pregnancy depending on body weight gain. In the 2nd half of pregnancy, the triacylglycerol level triples, and phospholipid and cholesterol levels increase by 50%.

The protein amount in urine changes only slightly during pregnancy. Daily protein loss in a healthy, non-pregnant woman is 200–300 mg; if a pregnant woman's daily loss increases to more than 500 mg, the cause must be sought, since proteinuria may be the first sign of a pathological process.

Serum Electrolytes

Na^+ – concentration decreases by 2–3 mmol/l (130–148 mmol/l)

K^+ – concentration decreases by 0.2 - 0.3 mmol/l (3.8–5.1 mmol/l)

Ca^+ – concentration decreases by 10% (2.25–2.75 mmol/l)

Mg^+ – concentration decreases by 10–20% (0.75–1.25 mmol/l)

Cl^- and phosphorus concentrations do not change markedly (98–106 and 0.65–1.62 mmol/l)

Normal ranges found in non-pregnant women are indicated in brackets.

Thyroid Gland

Increased renal clearance of iodine during pregnancy causes a decrease in iodine reserve in the body despite its sufficient supply from food. To compensate this change, the thyroid gland enlarges and the iodine uptake increases. This ensures sufficient synthesis of thyroid hormones, or euthyroid state of the body, respectively. The high oestrogen level stimulates thyroid binding globulin (TBG) production, which is why the interpretation of clinical tests dependent on the TBG level or on the thyroid's ability to take up iodine must be corrected during pregnancy.

In addition, the high hCG level causes a decrease to the TSH level, so especially in the 1st trimester it is advisable to use specific reference intervals which are roughly 0.05 – 3.5 mU/l) for TSH.

FT4 levels under 10 pmol/l are serious because they may have a negative effect on the psychomotor development of the child. Assays for anti-TPO antibodies may reveal women at risk of postpartum thyroiditis (about 15% in the Czech population), which develops in half of positive women and may endanger not only the women herself but also any subsequent pregnancy.

Trends in the level of other hormones essential for pregnancy have been shown in Figure 34.1.

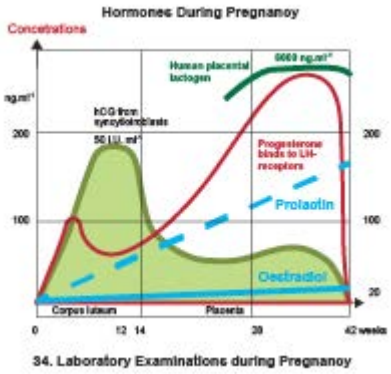


Figure 31.1. Trends in the level and concentration of hormones during pregnancy

Picture source: http://www.mfi.ku.dk/ppaulev/chapter29/Chapter%2029.htm

31.2.6. Conditions Threatening Pregnant Women

31.2.6.1. Pre-eclampsia

Pre-eclampsia is a serious condition occurring during pregnancy. Incidence is reported roughly 4–8% of pregnancies. This condition features increased blood pressure (limit is 140/90) and protein i**n urine** (0.5 g/24 hrs). **Oedemas** are not a typical symptom, as assumed before. It is caused by **maternal endothelial dysfunction**. This means that blood coagulant and anticoagulant factors are not balanced in the woman, and vascular cells produce more factors that induce vasoconstriction, the **narrowing of blood vessels**. Another factor is a **poor growth of the placenta**. Pre-eclampsia is manifested as **hypertension**. The first and mostly the only symptoms are headache, fatigue and nausea.

Laboratory Assays

Proteinuria – usually semiquantitative as + to +++, with values over 300 mg/day.

A rise in LDH, AST, ALT and bilirubin is found in the **maternal serum**. Elevated ALT levels may occur sporadically in pre-eclampsia without the HELLP syndrome. A full manifestation of the HELLP syndrome can be expected at a later phase in about one third of these cases.

Uric acid – a **rise over 320 µmol/l** is a marker of pre-eclampsia.

Malondialdehyde (MDA) – product of lipid peroxidation, a new marker of pre-eclampsia.

Supplementary serum assays: creatinine, urea, total protein

Blood count – abnormal, drop in platelet count

A rise in haemoglobin and haematocrit.

Changes in haemocoagulation – aPTT (activated partial thromboplastin time), fibrinogen and prothrombin time are usually normal; FDPs (fibrin degradation products) increase, and thrombin time becomes longer.

31.2.6.2. Markers of Pre-Eclampsia

PlGF and sFlt-1

Findings have shown that angiogenic growth factors such as PlGF (Placenta Growth Factor) and sFlt-1 (soluble fms--like tyrosine kinase-1, or VEGFR1) play a substantial role in the development of pre-eclampsia. PlGF is responsible for the normal function of the placenta, and therefore for a healthy pregnancy, while sFlt-1 affects the end of pregnancy in the last weeks of gestation. The level of these proteins circulating in the body changes in women affected by pre--eclampsia. The plasma sFlt-1 level increases both before diagnosis and at the time of clinical manifestations of pre--eclampsia, while the PlGF level is lower compared to normal pregnancy. The combined sFlt-1/PlGF ratio subsequently makes it possible to distinguish normal pregnancy from pregnancy with pre-eclampsia.

ADAM12

ADAM12 could be a valuable marker of complications during pregnancy. ADAM (A Disintegrin And Metaloprotease) proteins are multidomain glycoproteins with adhesive and proteolytic activities. High concentrations of ADAM12-S are found in the placenta, can be found in the serum of pregnant women from the first trimester with a growing trend. ADAM12 was decreased in the first trimester in all women who subsequently developed pre-eclampsia. This marker is one of the promising markers of pre-eclampsia. ADAM 12 is also likely to affect leptin metabolism and may be a mediator of the entire pre-eclampsia process. In addition to predicting pre-eclampsia, ADAM12 also speaks of chromosome abnormalities.

31.2.6.3. Complications of Pre-Eclampsia

Complications of pre-eclampsia include eclampsia, DIC and the HELLP syndrome.

Eclampsia is a life-threatening condition which usually follows pre-eclampsia. It is a paroxysmal disease manifested by **tonic-clonic seizures** occurring without any other pathology of the brain. Eclampsia is one of indications for termination of pregnancy irrespective of foetal age and maturity.

Disseminated intravascular coagulopathy (DIC) is a life-threatening condition. It is a breakdown of haemostasis with a production of **thrombi in small veins** followed by depletion of platelets, coagulation factors and uncontrollable bleeding. This condition can be cured if treated in time.

The **HELLP syndrome** is a serious complication of pregnancy occurring most commonly in connection with severe pre-eclampsia. This syndrome develops on a basis similar to that of pre-eclampsia and may or may not follow pre--eclampsia. HELLP usually occurs in the second or third trimester of pregnancy. Major symptoms are epigastric pain with nausea, vomiting, headache, hypertension, and protein in the urine. Typical laboratory findings are used for diagnosis: haemolysis (H), elevated liver enzymes (EL) and low platelet count (LP). Treatment for the syndrome is similar to that for pre-eclampsia, and supplementary treatment depends on additional symptoms with the aim of stabilizing the mother's general status, and adjusting the haemodynamic condition and impaired haemocoagulation.

Women with chronic renal or hepatic disease are exposed to this risk along with DM, pregnant women with autoimmune diseases or hypertension. Attention has been paid in recent years to carriers of thrombophilic mutations. The HELLP syndrome incidence in the Czech Republic, where prenatal care is at a high level, is estimated at around 4–5 cases out of 1000 births.

31.3. Conditions and Extent of Screening

Irregular examinations also include an assessment of the risk of chromosomal aberration (CHA) of the foetus. An essential requirement for chromosomal aberration examinations is to understand the difference between screening and the diagnostic test. Screening examinations are always a statistical model and are intended to distinguish that group of examined individuals which has an increased risk of CHA of the foetus. A diagnostic test will then finally determine which foetus has which developmental defect. Therefore, a positive screening result does not definitely mean the presence of a CHA. Informed consent is required for the screening test; the pregnant woman should accept the possibility of a false positive or false negative test result. Ultrasonic examinations are made by certified specialists on the appropriate instruments. Blood collections are scheduled depending on the gestational age found by ultrasonic examination, and biochemical markers are defined by a laboratory able to demonstrate external quality control of the markers tested. Appropriate and validated software is used to assess the risk. The screening centre is required to collaborate with medical genetics and foetal medicine facilities.

Recommendations for laboratory screening for congenital developmental defects in the first and second trimesters of pregnancy were prepared in collaboration with the professional societies for genetics, biochemistry, immunoanalysis and gynaecology in 2002. These recommendations are regularly updated (last update 2010) and are available (not only for laboratories) on the website of ČSKB ČLS JEP.

The first and second trimester screens use biochemical methods to calculate the risk of foetal harm by trisomy 21 (Down syndrome). For each screening procedure the risk in the population, false positivity (FP) and recovery (DR) have been clearly defined. The resulting biochemical marker levels also make it possible to estimate the risk of foetal harm by other chromosomal aberrations (Trisomy 18, known as Edwards syndrome, or Trisomy 13, known as Patau syndrome), neural tube defects, neural tube defect (NTD), or metabolic disorders (cholesterol metabolism disorder, known as

the Smith-Lemli-Opitz syndrome). Ultrasound methods are then used for the essential examination of nuchal translucency (NT) in the 1st trimester, and a detailed examination of foetus morphology is made in the 20th week of pregnancy.

31.4. Most Common Developmental Defects

The most common defects that can be detected using prenatal biochemical screening combined with specialized ultrasound examination are chromosomal or NTD of the foetus.

31.4.1. Chromosomal Defects

Chromosomal aberrations are caused by the absence or excess of a part of the chromosome (or the whole chromosome or even multiple chromosomes). Humans have a genetic set of 46 chromosomes in every cell. They consist of 44 autosomal (somatic) chromosomes and 2 sex chromosomes, X and Y, determining a person's sex. Women have two X chromosomes (46,XX), men have one X and one Y chromosome (46,XY). An error in the chromosome count always means serious problems for the individual. Most of these diseases are not familial and their occurrence is purely accidental.

31.4.1.1. Somatic Chromosome Aberrations

Down syndrome

One of the most common and best known chromosomal aberrations is Down syndrome (previously referred to as mongolism). Down syndrome is named after J. L. H. Down, who was first to describe it in 1866. It is a characteristic physical and mental disorder caused by chromosome 21 trisomy. The karyotype of such child is 47, XX, +21 or 47, XY, +21.

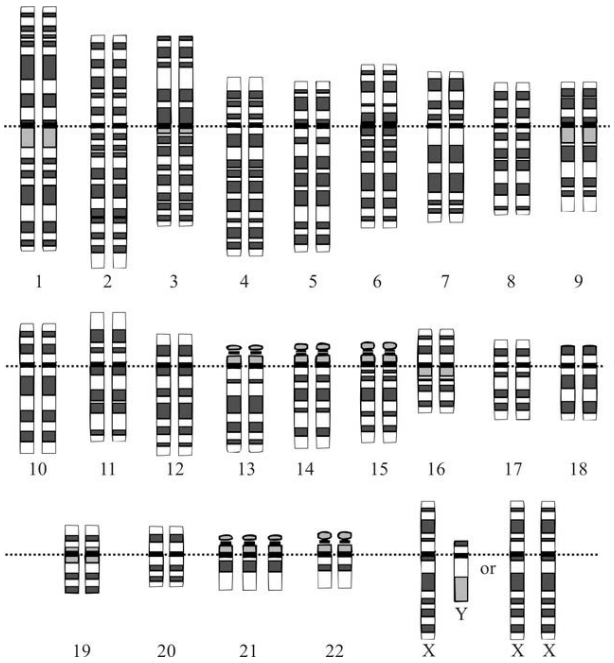


Figure 31.2. Karyotype – Down syndrome

Down Syndrome Karyotype" by Courtesy: National Human Genome Research Institute - Human Genome Project. Licensed under Public domain via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:Down_Syndrome_Karyotype.png#mediaviewer/File:Down_Syndrome_Karyotype.png

The incidence of Down syndrome in the Czech population is 1:700 - 800 live born children. Pregnancy is usually without complications, the children are born at full term, with normal birth weight. Individuals with Down syndrome have a stunted, squat stature (final height 144 - 155 cm), IQ 25 - 50 with a wide variability of skills. They are socially adaptable, love music, and are usually very friendly. They have typical facial features, short fingers and simian palmar creases. One third of the children have a heart disease, thyroid disorders and some other congenital developmental defects. Any woman may have a child with Down syndrome, but the risk increases with the mother's age.

Age at birth	Risk of DS at full term (1:n)
20	1500
25	1400
29	1000
35	380
40	110
43	50
45	30

Table 31.2. The risk of development of a foetus with Down syndrome depending on maternal age

Edwards syndrome

Edwards syndrome is caused by an extra 18th chromosome. It is a severe malformation syndrome. Pregnancy is usually complicated by the stunted growth of the foetus, and intrauterine death is common. The prognosis for life is poor; most children die during the neonatal period, only about 12% survive infancy.

P**atau syndrome**

A severe malformation syndrome caused by an extra 13th chromosome. Pregnancy often ends in spontaneous miscarriage or premature birth. More than 90% of children die within 1 year of birth.

Other chromosomal aberrations of somatic chromosomes are rarer. In every case, only symptoms are treated, because an already-existing aberration is untreatable.

31.4.1.2. Sex Chromosome Aberrations

Sex chromosome aberrations, mainly affecting sex organ development, are quite common but are not necessarily phenotypically distinctive.

Klinefelter syndrome

Klinefelter syndrome afflicts males with an extra X chromosome (47,XXY). The incidence is 1 boy in 670 male newborns. The onset of puberty is delayed, common hypogenitalism, practically always aspermia (sterility), gynaecomastia and female fat distribution, long lower extremities are typical.

Turner syndrome

A woman suffering from Turner syndrome has only one X chromosome (45,X). The incidence is 1 girl in 2000 female newborns. The newborn is born at full term, children have a triangular face with drooping eyelids, malocclusion, and puberty does not develop. Maximum height at adult age is 105 - 155 cm, IQ is usually normal.

31.4.2. Foetal Abnormalities Not Skin Covered

Neural tube defect (*spina bifida*, cleft spine) is caused by an incompletely closed spinal canal at the early phases of embryonic development. This defect may occur anywhere in the spine; if the head part is afflicted, **anencephalus** will develop (almost complete absence of brain and skull). There are two types of abdominal wall defects: **Omphalocele**, in which abdominal organs of the embryo remain outside the abdomen in a sac in the place of umbilical cord attachment, or **gastroschisis**, in which a cleft forms in the existing abdominal wall, through which abdominal organs may protrude.

31.5. Screening for Congenital Defects

31.5.1. Non-Invasive Methods

Examinations using non-invasive methods include blood sample tests using immunoanalytical methods or ultrasonic examinations. The following basic parameters are specified for each screening system: **sensitivity** (rate of afflicted pregnancies identified) and **false positivity** (percentage of healthy foetuses falsely identified as positive). The objective is to perform screening with maximum attainable sensitivity (> 90%) and minimum false positivity (3 - 4%).

31.5.1.1. Immunoanalytical Methods

Blood for screened markers can be taken anytime during the day, because the assayed substances have minimum diurnal variability. The requirement of fasting blood collection is justified only if radioimmunoanalysis is used, because serum chylosis may affect the antigen-antibody complex precipitation. The test is made in the serum or plasma by immunochemical methods, the reproducibility of which is an essential requirement for the screening. Luminiscence, radioimmunoanalytical or fluorescence methods are used; enzymeimmuno-analysis is not recommended due to its insufficient accuracy.

First Trimester Screening

First trimester screening has been gradually developing over the last 10 years. The most common method is referred to as the **combined test**. The test consists in a joint evaluation of the results from a biochemical test (PAPP-A, free beta-subunits of hCG) and ultrasound examination. The examination is made from the 11th to 13th week of pregnancy. In addition to the benefits of earlier response, this method also has higher sensitivity (about 85%) with a false positivity of about 4 - 5%. First trimester screening can be recommended to women who wish to obtain an early result and do not wish to undergo the second trimester test. The advantage of the combined test is its high reliability combined with minimum risk for the pregnant woman, accurate determination of gestational age, individual evaluation for multiple pregnancy, birth date determination and early detection of severe foetal malformations. A lower PAPP-A level warns of potential hazard for the foetus – a risk of chromosomal abnormality or abortion.

Biochemical markers for first trimester screening

PAPP-A

PAPP-A belongs to the group of Pregnancy Associated Plasma Proteins detected by immunochemical methods in the serum of pregnant women. There are 4 types of these proteins in total:

- **A** – Zn-dependent metalloproteinase able to degrade IGFBPs which play a role in the regulation of local proliferative reactions. IGFBP degradation activates IGF.
- **B** – The function is so far unknown.
- **C** – SP1 protein, previously used for screening, has enabled more precise determination of gestational age.
- **D** – human placental lactogen (hPL)

The highest PAPP-A concentration is at the end of pregnancy, and decreases immediately after birth. The biological half-life is 3 - 4 days. The PAPP-A level is reduced in most chromosomal aberrations, but only in the first trimester in Down syndrome. Very low PAPP-A levels are found in at-risk pregnancies ending in spontaneous abortion.

Free β-subunit of hCG

The measurement of the free β-subunit of chorionic gonadotropin is used for first trimester screening. It is a problematic marker as regards the pre-analytical phase. To guarantee correct results, the material has to be cooled and quickly transported to the laboratory, where serum should be separated immediately. Serum and plasma stability is not longer than 6 hours at +20 to 25 °C or 1 day at +4 to 8 °C. If frozen to -20 °C, free β hCG is stable for 1 year.

Second Trimester Screening

It was discovered in the second half of the 1980s that mothers carrying a foetus with Down syndrome have different levels of some substances in their blood. This applies primarily to alpha-1-foetoprotein (**AFP**), which is reduced by about 25% compared with the usual mean value, and human chorionic gonadotropin (**hCG**), which is almost twice as high.

Screening for Down syndrome in mothers in the 2nd trimester of pregnancy has been carried out in the Czech Republic since the early 1990s; the test is available to every pregnant woman and is covered by health insurance. The organization of the test itself is very simple and consists in blood sampling in the 2nd trimester of pregnancy (preferably in week 16 - 17) and in the test for two or three biochemical parameters: AFP, hCG and/or free oestriol (**uE3**). This test, sometimes also referred to as the **double** or **triple test**, is evaluated using sophisticated software, which also considers the maternal age, weight and the ultrasound-determined gestational age to calculate the individual risk for each pregnant woman. The result is probability expressed as 1 : X, where there is a limit (usually 1:250 - 300), from which results are regarded as positive. A standard part of the test is the check made by a geneticist on all results (including negative

ones). The sensitivity of this method of screening is about 65-70% and false positivity is about 6 - 7%. Every pregnant woman with a positive screening result is offered an invasive diagnostic procedure, amniotic fluid or chorionic villus sampling, to determine the karyotype of the foetus from the collected material culture.

From a modern perspective, with new developments in ultrasonic and genetic diagnosis, separately performed 2nd trimester screening seems outdated, but it is important to preserve it. The first reason is that a certain percentage of women come to the gynaecologist too late to make an early examination. Another reason is connected with the determination of AFP, which is the only biochemical marker of the neural tube defect. This defect is usually detected by ultrasound examination in the 20th week of pregnancy, but AFP measured in the 16th week warns of the risk of this defect one month earlier. A significant and perhaps the main reason for preserving the second trimester screen is that it is possible to include it in the integrated test.

Biochemical markers for second trimester screening

AFP (α-1-foetoprotein)

AFP is an oncofoetal glycoprotein produced in the yolk sac and in the foetal liver. AFP transfers to the maternal serum through the placenta and is an important indicator of the physiological course of pregnancy. The transport function of AFP is important. AFP concentration grows as pregnancy progresses. The highest AFP concentration in the maternal serum is around the 30th week of pregnancy, and the level drops dramatically after birth.

hCG (human chorionic gonadotropin)

During early pregnancy, hCG stimulates progesterone synthesis in the corpus luteum. It is a glycoprotein comprising two different subunits, α and β. The α-subunit is the same as for LH, FSH and TSH, so the β-subunit is considered specific for hCG. A steep increase is typical of hCG concentration at the beginning of pregnancy; hCG starts to fall after the 10th week.

uE3 (unconjugated oestriol)

Unconjugated oestriol is an oestrogenic hormone produced by the placenta. Its concentration expresses the interaction of the foetus and the placenta. uE3 is an unstable structure which readily disintegrates, so measured uE3 concentrations may be biased by inadequate transport or storage. Since the uE3 assay also presents other problems, such as the effect of lipaemia or the different affinity of antibodies depending on the diagnostic set manufacturer, many laboratories have stopped using uE3 assays. The low oestriol level itself may often change the screening result into positive, without this being caused by any congenital defect of the foetus. Nevertheless, a free oestriol assay may detect the risk of some rarer defects: Smith-Lemli-Opitz syndrome, where uE3 is very low; all of the three screening markers are low in Edwards syndrome. Free oestriol also warns of impaired placenta with a risk of complications at the end of pregnancy.

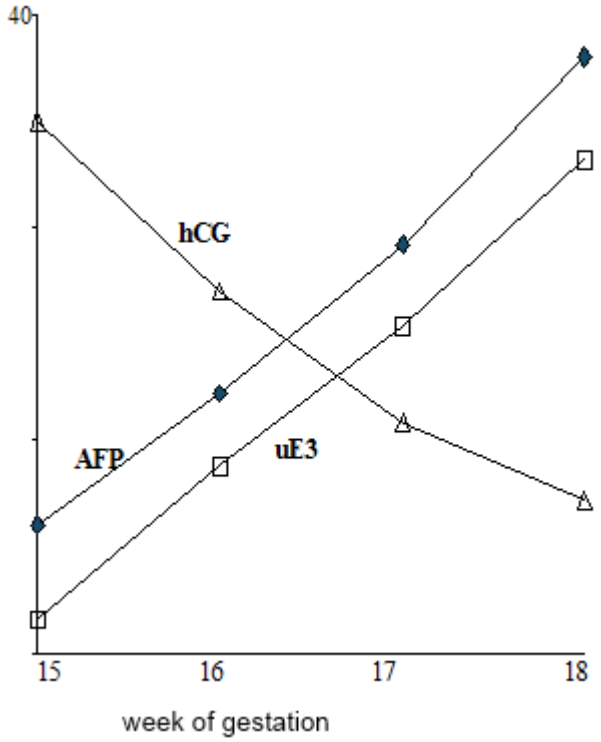


Figure 31.3. Changes in marker levels in the 2nd trimester: the dynamics in pregnant women with DS are delayed, hCG is higher, and AFP and uE3 are lower than in the healthy population.

Inhibin

Dimeric inhibin A is the fourth marker used worldwide for Down syndrome screening in the 2nd trimester, which is then called the quadruple test. Inhibin is a glycoprotein of placental origin similar to hCG. Maternal serum inhibin levels are relatively stable from the 15th to 18th week of pregnancy. Inhibin levels in foetuses afflicted by Down syndrome are about twice as high as healthy foetuses. Inhibin tests are not yet common in the Czech Republic.

Multiples of the Median

Since the all biochemical screening is based on a statistical assessment of the measured parameters, it is necessary to have a base set with which to compare the results. Every laboratory must have examined at least 50 women in each tested week of pregnancy (15th to 18th week) and calculate medians for each parameter from the measured values. The median is the central value in a set and is used for screening because the arithmetic mean includes extreme outliers that do exist but may affect the average. Results therefore are not given as absolute value but as MoMs (multiples of the median). MoM is the number by which to multiply the relevant median to obtain the value measured in the serum of the pregnant woman. If the median for the 16th week of pregnancy is, for example, 30kU/l and the measured value is 60 kU/l, the MoM is 2.0. The calculations are made by specialized evaluating software. The use of the MoM makes it possible to compare different methods, and external quality control usually requires not only the absolute value but also the MoM. An increased risk of Down syndrome is indicated by MoMs of hCG higher than 2.5, MoMs of AFP lower than 0.5, MoMs of free β-hCG higher than 2.0, and MoMs of PAPP-A lower than 0.4.

Integrated Test, Sequential Test, Serum Integrated Test

Another option in assessing the risk of Down syndrome in a foetus is to assess the combined results of the first and second trimester tests. The standard **integrated test** is described as a statistical evaluation of all data, i.e. the first trimester biochemical (PAPP-A, free β hCG) and ultrasound (NT) screening following the second trimester biochemical (AFP, hCG, uE3) screening. This approach has advantages and disadvantages. The most common drawback is the stress for the pregnant woman due to the delayed information about the result. For this reason the **sequential test** is a suitable modification of the integrated test, because if a high risk is detected, the woman is already notified of the positive result after the first trimester part of the test. A low PAPP-A or a high NT present a positive result. The advantage of an early result is not thereby lost. Of course, the pregnant woman is then offered immediate invasive examination, or she may wait until the 2nd trimester for the overall test result.

If the first trimester combined test result is negative, the woman undergoes biochemical screening in the second trimester. Integration of the results provides a screening system having not only a sensitivity of far above 90%, but also a very low false positivity of around 1 - 2%.

Since proper NT measurement requires not only a high-quality ultrasonic scanner, but also a trained specialist certified by the Fetal Medicine Foundation (FMF) of London, high-quality nuchal translucency measurement is not available to all women at present. The value of this parameter has a very heavy weight in the computing algorithm, and may considerably shift the resulting risk. There is also the option of a **serum integrated test**, which in the first trimester only includes the biochemical assay for PAPP-A, or sometimes also free beta hCG. This system can also be usefully applied in places where samples are brought collectively to the laboratory and the woman does not have to travel anywhere. The only problem remains the lower stability of free β hCG, which requires strict compliance with all pre-analytical phase requirements. This test has not only low false positivity (5%) but it also enables an NT screen to be added for that group of pregnant women who have PAPP-A under 0.35 MoM (only 2 - 3% of women). If the risk is high (over 1:30 – about 1 woman out of 200), chorionic villus sampling (CVS) with subsequent karyotyping can be made.

The results of current extensive international studies and the practice of some centres in the Czech Republic show that the sequential form of the integrated test may be considered a screening system which meets professional requirements on the one hand, and, on the other hand, gives pregnant women a higher level of certainty when their pregnancy is screened for congenital developmental defects.

31.5.1.2. Ultrasound Diagnosis

Imaging methods such as ultrasound are non-invasive and may warn of many diseases or defects.

As a minimum, the ultrasonic part of examination includes measurement of the nuchal translucency (NT) of the foetus. Some systems also include a screen for nasal bone (NB); the nasal bone is not visible in a foetus with Down syndrome. It is a disputable marker – if something cannot be seen it does not mean it is nonexistent. The sonographer has to be certified by the FMF and audited on a regular basis to be able to measure NT. Nuchal translucency (NT) is measured from the 11th to 13th week of pregnancy. The ultrasound measures accumulated fluid in the foetal neck subcutis. The presence of large amounts of fluid in this space is often connected with a heart defect or chromosomal aberration. About 70% of foetuses with Down syndrome have a significant oedema on the neck caused by accumulated fluid. The examination can usually be made through the abdominal wall. In addition to nuchal translucency, the foetus is thoroughly examined by ultrasound. The number of foetuses is determined, and if multiple pregnancy is found, a chorionicity scan is made and the gestational age confirmed or corrected based on the CRL (crown-rump length), and the morphology of the foetus is checked. The best staffed and equipped facilities are able to measure additional ultrasound markers that may increase the sensitivity of the result. Ultrasound screening for trisomy 21 markers such as nasal bone, tricuspid regurgitation, ductus venosus flow or other markers requires separate certification and regular auditing to renew licences for each marker (NT, NB, TR, DV). Sonographers may use only those markers for screening for which they hold a valid licence. At the end of the 1st trimester, ultrasound may rule out most serious defects of foetal morphology (the parameters include skull shape, width of brain ventricles, face of the foetus, shape of the chest and torso, basic heart anatomy, size of the urinary bladder, structure of the upper and lower extremities, neural tube and anterior abdominal wall defects). In practice, however, it is hard to achieve such extent in normal screening.

Prenatal diagnosis of normal pregnancy ends with the ultrasound examination in the 20th week of gestation.

	False Positivity at 85% Recovery	Recovery at 5% False Positivity
First trimester combined test PAPP-A, free β hCG + NT	3.8 – 6.8%	85%
Second trimester AFP, hCG + uE3	9.3 – 14%	69%
Integrated test PAPP-A (free β hCG) + AFP, hCG (uE3)	0.8 – 1.2%	94%
Serum integrated test PAPP-A (free β hCG) + AFP, hCG (uE3)	2.7 – 5.2%	85%

Table 31.3. Comparison of false positivity and recovery of each test (source: SURUSS and FASTER studies)

31.5.2. Invasive Methods

If the result of screening for congenital defects is positive, the pregnant woman is informed about the option to use invasive examination techniques. For the time being, a sample of the foetal genetic material for the test (either a cytogenetic or molecular genetics test), which is often essential to confirming or disproving the suspicion of chromosomal developmental defects of the foetus, has to be taken using standard invasive methods. Commonly used methods include amniocentesis (amniotic fluid sampling), chorionic villus sampling (CVS) and cordocentesis (umbilical cord blood sampling). Invasive prenatal diagnosis is made only in indicated cases under ultrasonic control. The problem of all invasive methods, a problem that should be communicated to the pregnant woman in advance, consists in the risk of complications which may lead to spontaneous abortion. The risk of abortion due to amniocentesis or chorion sampling is about 0.5 – 1%.

31.5.2.1. Amniocentesis (AMC)

Amniocentesis is amniotic fluid sampling for chromosomal and biochemical analysis and is usually made between week 15 and 17 of gestation. It is the most common method of invasive prenatal diagnosis in the Czech Republic. Previously, this test was automatically offered to women over 35 because of the steeply rising risk of genetic defects in that population (Table 2). The currently preferred opinion is that age is only one of the markers to assess the risk of foetal defect. Amniotic fluid is usually sampled through the abdominal wall of the pregnant woman using a very thin needle inserted into the amniotic sac with the aid of ultrasound guidance. Approximately 20 ml of amniotic fluid is extracted. The overall risk of complications in connection with the procedure is estimated at 0.5 - 1.0%. The results of the karyotyping are available within 2 weeks of sampling. The result of the PCR or FISH analysis is known within 48 hours, which considerably reduces the stress of the pregnant woman. However, these tests detect only the most common chromosomal defects; the result of complete karyotyping must always be known.

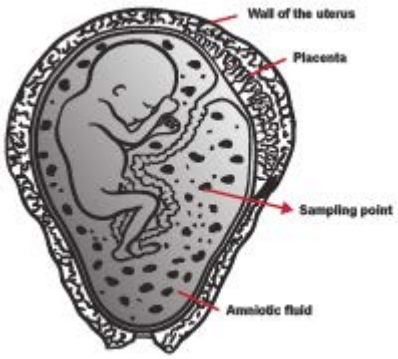


Figure 31.4. Method of amniotic fluid sampling

31.5.2.2. Chorionic Villus Sampling (CVS)

Chorionic villus sampling is usually made between week 11 and 13 of gestation, often following a positive first trimester screening result. Chorionic villi are part of the placenta. Biopsy can be made through the abdominal wall or vagina. The risk of abortion in connection with this procedure is similar to the risk of amniocentesis – below 1%.

31.6. Cytogenetic Methods

Foetal cells obtained by amniocentesis or chorionic villus sampling are cultivated to detect all deviations in the number and structure of chromosomes. The test is used for cystic fibrosis, muscular dystrophy, haemophilia and other genetic defects in the family history. This test also exactly determines the child's sex.

31.6.1. New Methods for Genetic Analysis

Foetal cells are most often tested by fluorescence *in situ* hybridization (**FISH**). This technique of molecular cytogenetics is suitable to test for aneuploidy of the embryo, to determine the sex of the embryo, or determine whether the parents are carriers of structural aberrations. Fluorescence *in situ* hybridization is based on the principle of the hybridization of short, fluorescently labelled DNA sequences (probe) with relevant sections of the target DNA sequence in the tested embryonic cell. The result of this hybridization is read under the fluorescence microscope. This method gives an overview of the number of copies of the selected DNA section representing certain chromosome in the nucleus of the examined embryonic cell.

A molecular genetics method known as polymerase chain reaction (PCR) is used to detect monogenic diseases. PCR was developed in the 1980s and today it represents one of the most commonly used methods of molecular genetics--based diagnosis.

Information about the possibility of capturing the DNA of the foetus freely circulating in maternal blood was published in 1997. Testing foetal DNA fragments in the maternal blood is safe for the foetus, and the method can detect a wide range of genetic diseases. The discovery of foetal DNA circulating in the maternal blood has surely been a major contribution to prenatal diagnosis. However, the clinical use of this discovery has been confronted with numerous obstacles, many of which have yet to be overcome.

The first requirement was to prove that maternal blood contains the entire genome of the foetus. The blood contains only fragments of foetal DNA, which are surrounded by multiple amounts of maternal DNA fragments. These fragments are separated by comparing them with paternal DNA. This way it is possible to create a map of the foetal genome coming from the father. It is more difficult to separate maternal DNA, which represents up to 90% of fragments in the plasma, from foetal DNA coming from the mother. Nevertheless, there is a difference between the average size of maternal and foetal DNA fragments. The solution to this problem has opened the way to using this technology for non-invasive diagnosis of genetic defects of the foetus.

Thi s area is still undergoing research, and new, partial discoveries continue to be made. Although this method is quite expensive, as the technology develops it will surely be cheaper and thereby become more available for a wide spectrum of diagnosis.

31.7. Conclusion

All the congenital developmental defect screening algorithms mentioned above are available in the Czech Republic, although not equally in all regions. This screening is a typical example of interdisciplinary collaboration, in which biochemistry and genetics laboratories participate with each other alongside gynaecologists. Only by their cooperation and willingness to support one another, without favouring any one component, is it possible to provide rapid and high--quality examination of the pregnant woman. The possibilities of developmental defect risk assessment are becoming more and more sophisticated and accurate, but with the increasing requirement for the high quality of each part of the screening process, they often become less available to a certain part of the pregnant population. As the awareness of pregnant women about this issue is growing, physicians as well as other medical professionals should also be aware of the options available to the pregnant woman. The economic aspect also plays a role because invasive procedures (CVS, amniocentesis) are then followed by other laboratory examinations which are very expensive, and if used pointlessly based on poor screening, they are ineffective and uneconomical.

Societal attitude towards induced abortions has changed quite dramatically in the 20th century. It has been influenced by the tradition, culture, religion and history of each country. One can hardly claim that the official position set out in the legislation is supported by all citizens of the country. A person's attitude towards induced abortion is highly individual.

The Czech Republic is one of the countries with liberal laws, where induced abortion is available on the written request of a pregnant woman whose pregnancy is not longer than 12 weeks, unless medical reasons prevent the intervention. If genetic reasons suggest that pregnancy be artificially interrupted, it is allowed until the end of the twenty fourth week of pregnancy.

In 2010, the Parliamentary Assembly of the Council of Europe adopted Resolution 1763, which contains, among others, the following principles:

1. No person, hospital or institution shall be coerced, held liable or discriminated against in any manner because of a refusal to perform, accommodate, assist or submit to an abortion, the performance of a human miscarriage, or euthanasia or any act which could cause the death of a human foetus or embryo, for any reason.

2. The Parliamentary Assembly emphasizes the need to affirm the right of conscientious objection together with the responsibility of the state to ensure that patients are able to access lawful medical care in a timely manner.

31.8. Case Reports

31.8.1. Case Report 1

First trimester screening

Age at term: 31 years

Ultrasound normal, measured CRL of 53.8 mm corresponds to gestational age of 12+1

- NT 1.7 mm 1.2 MoM
- free β hCG 91.8 ng/ml 2.13 MoM
- PAPP-A 1.53 mIU/l 0.58 MoM
- Risk of DS estimated based on age 1:823
- Risk of DS (NT+age+biochemistry) **1:919**

The conclusion of the first trimester screening was negative, although hCG was slightly elevated and PAPP-A was lower than expected for the given week of pregnancy.

Second trimester screening

- AFP: 29.4 ng/mL 1.02 MoM
- Total hCG: 146.67 kIU/L 3.93 MoM
- Gestational age at sampling by ultrasound 16+0
- Risk of DS calculated from markers for second trimester: **1:70**

The conclusion of the second trimester screening was positive. The hCG level was almost 4 times higher than the median for the given week of pregnancy.

When the first and second trimester screening results were combined, **the integrated risk of DS was 1:190.**

Conclusion of the integrated test: positive screening

Amniotic fluid was sampled and the material thus obtained confirmed trisomy 21 using the PCR technique. The culture confirmed the presence of Down syndrome in the tested foetus. The pregnancy was aborted at the request of the pregnant woman in the 20th week.

31.8.2. Case Report 2

First trimester screening

- Age at term: 35 years
- Ultrasound normal, measured CRL of 56 mm corresponds to gestational age of 12+3
- NT 2.1 mm 1.6 MoM
- free β hCG 7.6 ng/ml 0.2 MoM
- PAPP-A 0.1 mIU/l 0.15 MoM
- Risk of DS estimated based on age 1:286
- Risk of DS (NT+age+biochemistry) **1:14**
- Risk of T18 and T13 **1:19**

The conclusion of the first trimester screening was positive; the risk of congenital developmental defect was very high for other trisomies as well. Foetal karyotyping was recommended. Chorionic villus sampling was made for this purpose in the 12th week. The culture confirmed a normal finding, 46 XX. The pregnancy nonetheless ended by spontaneous abortion in the 16th week. This was 4 weeks after the CVS, so the abortion could not have been related to the CVS. It is advisable to monitor all pregnancies with low PAPP-A levels, which may indicate an approaching abortion.

31.8.3. Case Report 3

Combined test – first trimester test

- Age at term: 29 years

Ultrasound normal, gestational age 11+4 based on CRL of 55.0 mm

- NT 1.7 mm 1.2 MoM
- free β hCG 18.0 ng/ml 0.327 MoM
- PAPP-A 0.35 mIU/l 0.231 MoM
- Risk of DS based on age 1:987
- Risk of DS (NT+age+biochemistry) **1 : 2 213**
- Risk of DS (age+biochemistry) **1 : 550**
- **Risk of T13 and T18 1 : 150**

Conclusion: Given the low measured nuchal translucency, the screening was evaluated as **negative** despite the higher risk of congenital defects T13 and T18. No other examination was offered, although the biochemical results are alarming. The pregnancy ended with the birth of a child with Down syndrome in week 38 of pregnancy. The low free β hCG level affected the calculation of risk. Not all foetuses with trisomy 21 have an elevated NT value, as shown by Case Report 1.

CHAPTER 32

32. Specificities of Laboratory Examination during Childhood

Author: prof. MUDr. Průša Richard CSc.

Reviewer: doc. MUDr. Vladimír Soška, CSc.

32.1. Metabolic Differences

In terms of biochemistry and metabolism, children are a unique group; they are not miniature adults. For example, newborn and premature infants have immature hepatic, renal and pulmonary functions. This also has a significant effect on the pharmacokinetics of administered drugs. In most cases, reference limits for infants differ markedly from older children and adults in biochemical and haematological examinations. Gestational age (mature/immature) and birth weight (low/very low) should also be taken into account in newborns. This also applies to the reference range for creatinine.

Postnatal Age	Unit	Gestational Age 28 Weeks	Gestational Age 32 Weeks	Gestational Age 36 Weeks	Gestational Age 40 Weeks
		S-creatinine	S-creatinine	S-creatinine	S-creatinine
2 days	umol/l	40-220	27-175	23-143	18-118
7 days	umol/l	23-145	19-119	16-98	13-81
14 days	umol/l	18-118	15-97	12-80	10-66
21 days	umol/l	16-104	14-86	11-71	9-57
28 days	umol/l	15-95	12-78	10-64	9-53

Table 32.1. Serum creatinine reference values for newborns based on the age and gestational age of the newborn

In the neonatal period, the oxygen consumption is 180–280 ml/min/m2 per kg of weight, while a 10–year-old child has 120–220 ml/min/m2 per kg of weight. The already high heart rate and a high cardiac output are limiting factors. Easy oxygen reserve depletion in newborns and infants is the cause of the fast development of cardiorespiratory insufficiency as a result of a higher oxygen requirement in the case of hypothermia, infection, stress, pain, and pulmonary or cardiac disease. Monitoring the water and ion balance is of paramount importance in children. Small children are easily subject to dehydration due to their relatively large skin surface (surface/weight ratio). The skin of immature and premature newborns is highly permeable and the risk of dehydration is huge. In addition, as a result of immature tubular renal function, this group exhibits very high salt losses and the inability of the kidneys to compensate haemoconcentration due to insensible losses. Children under 1 year of age easily develop metabolic acidosis as a result of insufficiently effective urine acidification and a low renal threshold for bicarbonate re-absorption. Hypoglycaemia is often found in children. Children have a low reserve of glycogen and immature capacity for gluconeogenesis. Stress episodes quickly lead to hypoglycaemia. Physiological neonatal hyperbilirubinaemia and the differential diagnosis of pathological hyperbilirubinaemia in children is a regular part of paediatric laboratory diagnosis.

Another area is inherited metabolic diseases, which require specialized laboratories for their diagnosis. Neonatal screening for hypothyroidism, congenital adrenal hyperplasia, phenylketonuria and cystic fibrosis is based on biochemical assays on a dried blood spot (TSH, 17-hydroxyprogesterone, phenylalanine and the phenylalanine to tyrosine ratio, and immunoreactive trypsinogen assays). If cystic fibrosis is suspected, basic tests in a laboratory of paediatric biochemistry also include chloride tests as part of the sweat test by pilocarpine iontophoresis.

Paediatric laboratory diagnosis is distinguished mostly by differences given by the growth and development of infants, the collection of paediatric biological specimens, reference limits for children, and the selection of instruments and methods to process very small amounts of biological samples (5 - 10 ul/test). Special requirements are imposed on the monitoring of drug levels due to the different pharmacokinetics of drugs administered to children.

32.2. Collection of Biological Material from Children

Collection of blood from puncture. Fingertip or heel puncture is less traumatizing for younger children and newborns. Another possible, although less advisable method is sampling from the big toe or earlobe. Lateral or medial parts of the sole or heel are preferred sites in newborns, and fingertips are preferred in infants and younger children. Calcaneal osteomyelitis is a potential complication of the heel puncture. The risk of calcaneal damage by the puncture is greater in premature infants as the distance between the skin surface and periosteum grows as the infant gains weight. But calcaneal perichondrium injury will not occur if the puncture is not deeper than 2.4 mm. Since the distance from the skin surface to the perichondrium is lower on the posterior curvature of the heel, using this site for blood sampling is not recommended.

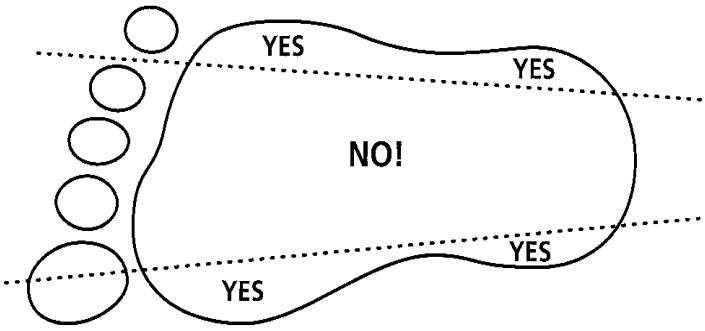


Figure 32.1. Convenient and inconvenient sites for heel puncture

Procedure for taking blood samples from newborns and infants for diagnostic purposes:

It is very important to heat the skin before puncturing, because the blood obtained is a mixture of arterial, venous and capillary blood, which also contains interstitial and intracellular fluid (this applies mainly to samples for blood gas analysis). The easiest way is using a cloth, such as a towel, soaked in hot water (38-42 °C). It has been proved that heating the skin at the site of puncture may increase blood flow up to seven times. The increase in blood flow is primarily arterial. To prevent nosocomial infections, the person taking the sample must remove all jewellery from their hands, wash their hands and forearms up to the elbows and use a new pair of latex gloves. After heating the skin, the site should be cleaned with a 75% aqueous isopropanol solution. Then the site should be dried with sterile gauze to prevent haemolysis by residual alcohol. Betadine should not be used at all because it may lead to an increase in potassium ion, phosphate or uric acid concentration in the sample. A short, sterile lancet or an automatic device making a puncture not deeper than 2.4 mm should be used. After selecting the site of puncture and puncturing, the first drop of blood should be wiped off because it often contains a lot of tissue fluid. Once the blood sample has been taken from the heel or fingertip, the extremity should be lifted above heart level and the site of puncture compressed with sterile gauze until bleeding stops. Using adhesive bandages (patches) is not recommended because the infant's skin is highly sensitive. In addition, there is a risk of the patch being pulled off and subsequent aspiration. The infant's heel should be grasped firmly with a gentle grip, putting the index finger on the arch of the foot, and the thumb under the site of puncture on the ankle. The puncture should be made smoothly, carefully and vertically to the surface of the skin. The thumb pressure should be relieved when blood drops are forming and flowing into the prepared collecting container. Strong massaging of the puncture site should be avoided due to the risk of haemolysis or interstitial fluid getting into the sample, which could lead to an increase in potassium ion and magnesium level in the sample. In comparison with venous blood collection, the levels of some analytes are different in this method; for example, the glucose level is 10% higher than venous blood glucose.

Venipuncture. This technique is used in older children, because they have larger and more visible veins and better tolerate the sight of the needle. In general, *v. cephalica* is most commonly used to obtain blood. The tourniquet should be applied midway between the elbow and the shoulder of the child to compress the vein, not the artery. The needle should be inserted in the axis of the vein and form an angle of 15° with the skin. The tourniquet should be removed as soon as the blood starts flowing into the syringe. After removing the needle, the puncture site should be compressed with cotton gauze until the bleeding stops (about 3 minutes), and covered with a patch. Creams and patches (EMLA 5%) are available to control pain during blood collection from paediatric patients. The only disadvantage is that the onset of puncture site anaesthesia is about one hour after application. The cream and the patch contain lidocaine (2.5%) and

prilocaine (2.5%).

The difficulty of venipuncture has lead to the use of non-invasive transcutaneous methods (e.g. pulse oximetry, transcutaneous bilirubinometer).

Volume of sample. Although many modern analyzers use small reaction volumes and the analyzers have low sample consumption, the size of samples remains the primary issue in the paediatric laboratory. This is because the number of premature and immature infants is growing (particularly the number of children with a birth weight under 1000 g has grown). The volume of plasma is relatively constant during lifetime, 4-5% of body weight. Although the total blood volume in a healthy newborn is about 85 ml/kg, it may be much less in a premature infant.

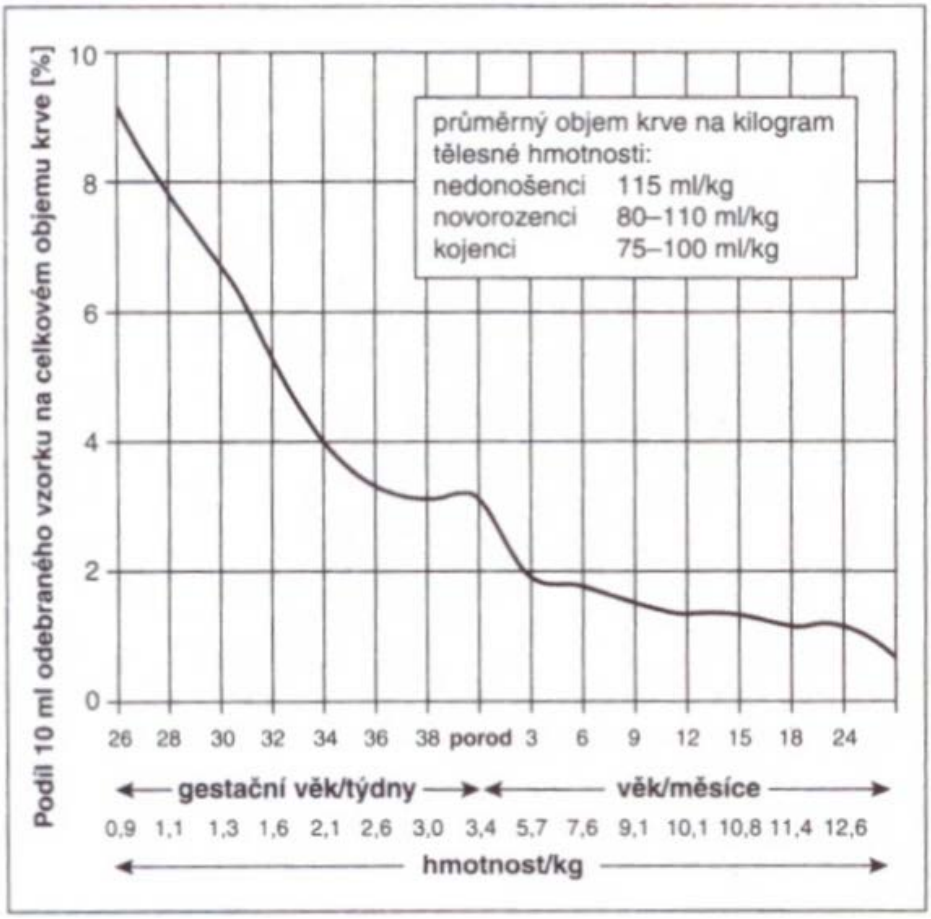


Figure 32.2. Proportion of a 10-ml blood sample in the total blood volume depending on age

1 ml of blood of a newborn weighing 1 kg corresponds to about 70 ml of an adult's blood. This results in the need for transfusion if the volume of blood samples taken in 3 days exceeds more than 10% of the total blood volume. For example, transfusion is required if only 8 ml of blood is taken from a 750-g child. Good planning of required laboratory tests for premature infants and all children in general will avoid excessive blood losses – nosocomial anaemia. It is necessary to monitor haematocrit and haemoglobin, and/or cover the losses if needed. The volume of blood taken for each sample from paediatric patients under 2.5–3.5 kg must not exceed 2.5 ml; the total amount for the whole time of hospitalization (shorter than 1 month) should not exceed 23 ml. Recommended maximum amounts of blood samples relative to the weight of the child (single samples and total amount of samples throughout hospitalization) are shown in Table 35.2.

Patient Weight (kg)	Maximum per Sample (ml)	Maximum per Month (ml)
0.5-0.9	1	8
0.9-1.8	1.5	12
1.8-3.0	2	17
3-4	2.5	23
4-5	3.5	30
5-7	5	40

7-9	10	60
9-11	10	70
11-14	10	80
14-16	10	100
16-18	10	130
18-21	20	140
21-23	20	160
23-25	20	180
25-27	20	200
27-30	25	220
30-32	30	240
32-34	30	250
34-36	30	270
36-39	30	290
39-41	30	310
41-43	30	330
43-46	30	350

Table 32.2. Maximum allowed amount of blood sample taken in a single collection and total amount throughout hospitalization (Note: Recommendation of the Institute of Clinical Biochemistry and Pathochemistry at the Faculty Hospital Motol, Prague)

Sample evaporation may be the main problem of small volumes. The thickening of an uncovered serum sample is already noticeable after the first hour. The values of a 0.1-ml sample may change by up to 20% at room temperature in two hours; the changes are smaller in larger (5 ml) samples (Figures 3 and 4).

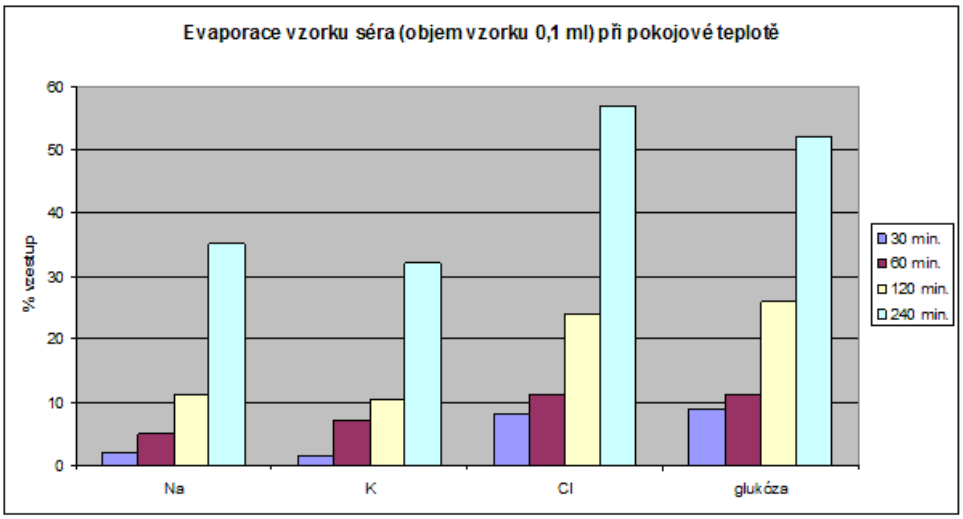


Figure 32.3. Serum sample evaporation (sample volume 0.1 ml) at room temperature

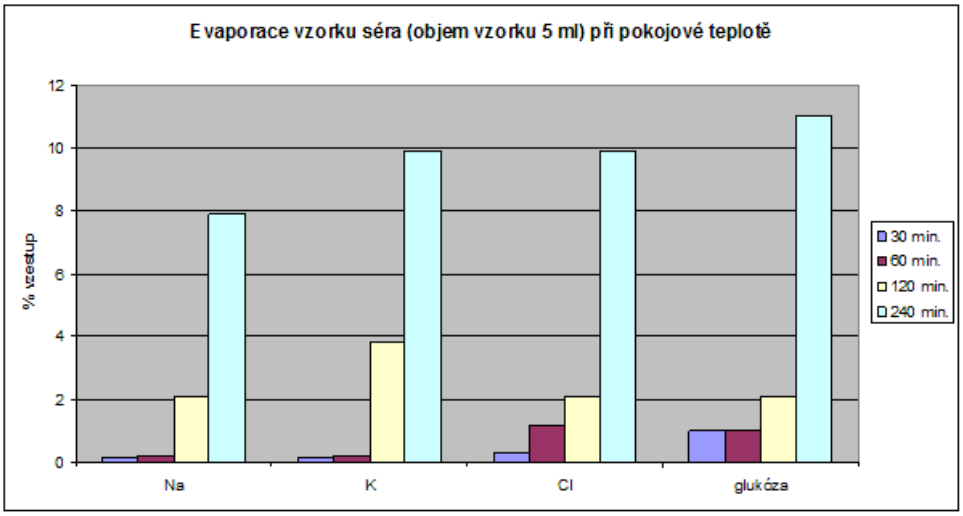


Figure 32.4. Serum sample evaporation (sample volume 5 ml) at room temperature

Considerable thickening by evaporation may occur in summer months. Small paediatric test tubes are sometimes difficult to label. Despite their high price, Microtainer tubes are preferred for serum or plasma collection from newborns. These test tubes have a serum separator and contain 600 µl of blood. They can be easily labelled. Serum is separated from cells after simple, short centrifuging. In addition, the serum separator does not affect levels of commonly assayed analytes. Plasma can be obtained using Microtainer tubes with lithium heparin salt. If greater amounts of blood are required for certain analyses such as hormone tests, 2.5-ml Vacutainer test tubes may be used for sampling. They are not used often because the veins of small children easily collapse. Although most laboratories use serum to measure analytes in clinical biochemistry, plasma is recommended. The reason is that the TAT (turnaround time) is shorter because waiting for coagulation is unnecessary, and also because haemolysis less frequent. In most routine tests, there is no difference in the results obtained from serum and plasma. Plasma should be used to measure the concentration of potassium ions, especially in patients with a high platelet or white blood cell count. It is important to make sure that the anticoagulation agent used does not directly affect the analyte being measured.

Urine collection. Proper urine collection is one of the most difficult tasks in paediatric practice. Since 24-hour collected urine volumes may be repeatedly inaccurate, it is good to know the estimated volume of urine of children of different age categories and so have an idea of the corresponding volume.

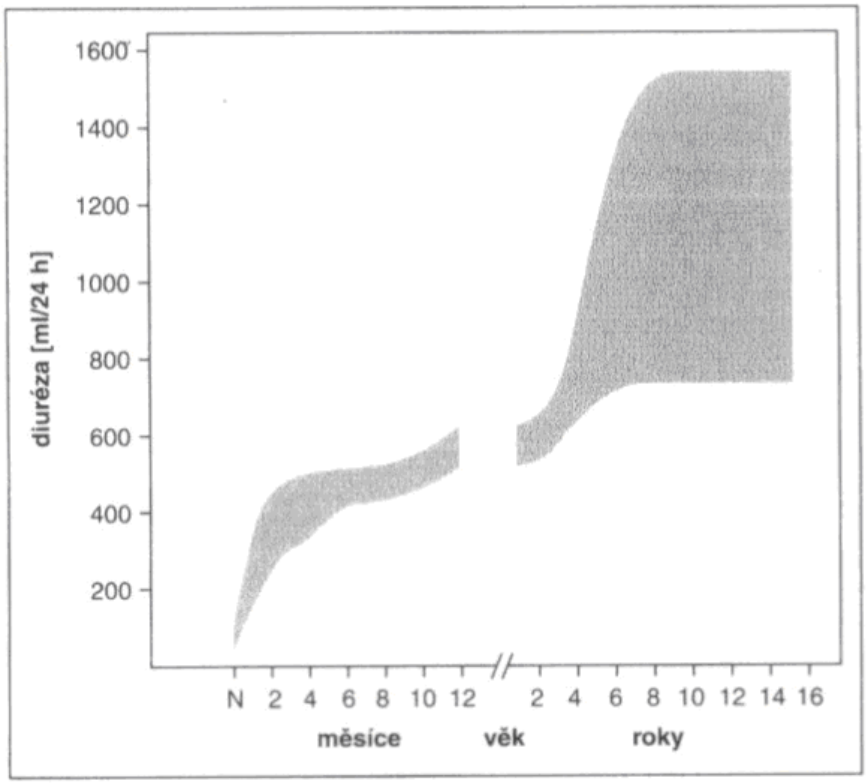


Figure 32.5. Daily urine volume relative to age

The figure shows expected volumes of urine in a healthy paediatric population. Samples for urine analysis must be refrigerated.

Stool collection. The patient and the parents should be informed that the stool sample must not be contaminated by urine.

32.3. Reference Range

Correct interpretation of laboratory data depends on the availability of suitable reference ranges (normal values) for the relevant child's age and degree of development. It is very difficult to obtain a sufficiently large sample of the population of healthy children in age groups from 1 day to 18 years, who could give samples, and obtain informed consent from their parents. This is why Hoffman's method is used. Normal reference limits are set for the paediatric inpatient population. Chauvenet's or Dixon's criteria and percentage cumulative frequency versus laboratory values (or versus logarithm of laboratory values, if the data distribution is not normal) are used to construct the reference limits (2.5^{th} and 97.5^{th} percentile); see the chapter Reference Limits by Bartoš. The advantage of this method is that, among other things, it respects the fact that inpatients commonly have higher thyroxine levels and lower cholesterol levels. Of course, one should remember that the reference limits depend on the analytical method used for the relevant biochemical test. Detailed paediatric reference values for clinical biochemical and haematological tests for difference age groups and genders for about 240 analytes are listed in publications by Soldin et al. or Meites et al. (SOLDIN, SJ., BRUGNARA, C., HICKS, JM. Pediatric reference ranges. Washington: AACC Press, 2003, 248 p., MEITES, S. Pediatric clinical chemistry. Reference (normal) values. Washington: AACC Press, 1989, 330 p.). Laboratory values of cord blood serum are often close to maternal serum values and may also change quickly. For example, cord blood T4 levels range between 85 and 225 nmol/l and increase to 142-278 nmol/l in 1-3 days of age. Serum proteins, immunoglobulins in particular, are also subject to significant changes throughout the development of a child from the postnatal period to adulthood.

1–5 days	16 days – 3 months	4 months – 1 year	1–6 years	6–15 years
5.7-12.7	2.3-6.6	2.2-11.2	5.5-14.7	6.7-15.0

Table 32.3. Example of dependence of reference limits on age. Reference limits for serum immunoglobulin G (g/l) by age groups

One analyte which shows marked changes with age is alkaline phosphatase. As newborns and children grow, the alkaline phosphatase level rises to levels deemed pathological in adults. These reference values are also different for each gender of the same age category.

Age	pH	pO2
0 min	7.11-7.35	1.07-3.2
5 min	7.09-7.30	4.41-10.0
10 min	7.21-7.38	4.13-11.3
30 min	7.26-7.49	7.2-10.7
60 min	7.29-7.45	7.2-12.7
2 days	7.36-7.45	11.1-14.4

Table 32.4. Example of dependence of reference limits on age. Reference limits for blood pH and pO2 (kPa) by age of the newborn

Normal blood gas values and pH are distinctively different in newborns and older children (pH 7.18–7.51 versus 7.35–7.44, respectively). The same holds true for sodium and potassium cations and calcium. For example, ionized calcium values in newborns under 1 month of age are 1.0–1.5 mmol/l, while the values are 1.18–1.32 mmol/l in adults. Serum reference values of enzymes such as amylase, GMT and AST are subject to a significant change with age. The most marked changes in the neonatal period can be seen in uric acid, bilirubin, renal clearance and fractional excretion of substances. The sexual maturation of children brings about phenomenal changes in the reference values of hormones, and sexual differentiation affects reference limits of many other analytes (prepubertal and postpubertal reference limits).

CHAPTER 33

33. Basics of Toxicology in Clinical Laboratory

Authors: Ing. Václav Senft

Reviewer: MUDr. Michaela Králíková

33.1. Introduction

Toxicology is a very broad domain, which derives knowledge from many basic sciences and scientific disciplines (chemistry, physics, biology, mathematics, statistics, medicine, pharmaceutics, ecology and others) and is applied to many areas of human activities (health care, hygiene, justice, legislation, forensic science, farming, public health, environment, working environment, waste management and others).

Clinical laboratory staff also come into contact with toxicological issues, albeit not all of them, only with some selected areas. Some laboratory principles and procedures are common to biochemistry and toxicology, however some methods, concepts and evaluations are different (e.g. the upper reference limit in biochemistry vs. maximum permissible level in toxicology, etc.).

This text contains the basic information, definitions and terms of toxicology placing special emphasis on areas stemming from the practice in clinical laboratories.

33.2. Toxicology

33.2.1. Definitions, History

The history of toxicology is very long. From time immemorial, people have found naturally occurring toxic substances (hemlock, datura, arsenic, etc.) and used them for hunting and battles in addition to involuntary events (poisoning). Toxicology (the name is derived from the Greek word “toxikón” meaning poison in which arrowheads used to be soaked) is a scientific branch concerned with the study of adverse effects of substances on living organisms.

We continue to come across toxicology today. Nowadays the substances are not only naturally occurring, but, above all, refer to artificially made products and preparations used by people for specific purposes on the one hand, however on the other hand we are encountering problems with them.

Examples:

- Mother, whose little child accidentally drank Diava (furniture polish), is surely very interested as to whether the product is toxic and what the first aid procedure is.
- We all worry about food becoming contaminated with heavy metals, herbicides and other hazardous substances introduced into the environment.

Although such substances may bring on physical effects such as electromagnetic radiation or crushed glass from a general point of view, attention is typically only paid to chemicals.

33.2.2. Subdisciplines of Toxicology

Toxicology is a very wide field, from which many subdisciplines have evolved focusing on specific areas.

Subdiscipline	***Short Annotation***
Environmental toxicology	Everyone is interested in toxicity in the home and external environment.

Industrial toxicology	Many people work in environments where toxic risks are present.
Clinical toxicology	Healthcare facilities take care of intoxicated patients (random accidents, suicides, abused substances).
Forensic toxicology	The criminal acts of poisoners are not just restricted to detective stories.
Military toxicology	Armies throughout the world pursue the matter of toxic warfare agents, their effects, deployment and protection against them.
Pharmaceutical toxicology	The application of almost every drug presents a risk of toxicity, either in the form of adverse effects or overdosing.
Experimental toxicology	Many experiments study the toxic properties of substances.
Predictive toxicology	Effort to predict toxicity without animal experiments using computer programs.

Table 33.1. Examples

33.3. Toxicological Indications in Clinical Laboratories

Various indications occur depending on the organization of laboratory disciplines in a specific medical facility. In smaller laboratories it tends to usually only involve the simple immunochemical monitoring of drug abuse, large consolidated (integrated) facilities also deal with special research and other requirements.

Subdiscipline	Requirements From	Reasons for Indication (Examples)
Clinical toxicology	Anaesthetic/resuscitation clinic (ARC), emergency, trauma surgery, internal medicine, psychiatry	Accidental and intentional intoxications, toxicological screening for unconsciousness of unknown aetiology, checking for drug overdose, checking for abstinence among drug addicts, etc.
Industrial toxicology	Occupational medicine, hygiene, practitioners and company doctors	Occupational exposure control, compliance with limits, detection of exposure effects, occupational diseases.
Environmental toxicology	ARC, emergency, hygiene	Accidental intoxications in the home and external environments, most often children.
Forensic toxicology	ARC, emergency, psychiatry, forensic medicine	Demonstration of substances of abuse (alcohol, drugs) e.g. after road accidents.

Table 33.2. Examples

33.3.1. Causes of Intoxications

For the purpose of illustration, the causes of intoxication can be divided into the following groups:

1. “Voluntary unknowing”. This group includes drug addicts (junkies) and sportspeople using some form of doping. These people are not aware of being intoxicated, and use drugs or doping agents voluntarily, regardless of the consequences.
2. “Voluntary knowing”. This group definitely includes suicide attempts.
3. “Involuntary”. This group includes various accidents, i.e. accidental poisonings, for example due to a mistake or unexpected contamination or concentration of substances. Accidents occur not only in the private sphere but also at work. To some extent, involuntary victims of crime also rank among this group.

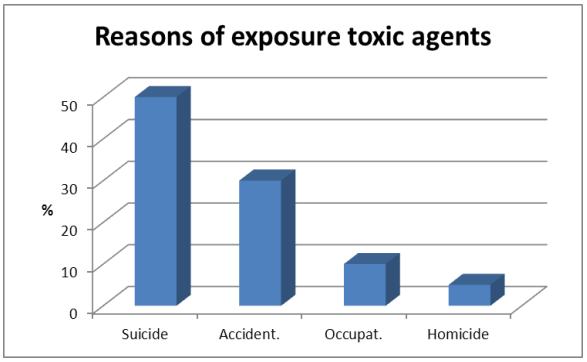


Figure 33.1. Shows the relative incidence of groups 2 and 3, i.e. cases reported in various statistics. It is clear that the incidence of group 1, i.e. the number of people abusing drugs or sportspeople using doping, is not exactly recorded, so only rough estimates are available.

33.4. Poisons

33.4.1. Definition

"Poison is a substance that causes negative effects to living organisms already in a small amount."

Almost all substances may have negative effects, the only difference is the dose required for the onset of toxicity.

For example, distilled water is not considered a poison because one would have to drink large amounts in order to be poisoned.

Physical actions carried out by substances (boiling water, ground glass, etc.) are usually not considered poisons either.

33.4.2. Sources and Forms of Toxic Substances

Humans have encountered **natural poisons** (hemlock, mushrooms, snake toxins, vegetable alkaloids, etc.) from time immemorial. Nowadays there is a huge amount of **artificial substances** (medicines, herbicides, pesticides, insecticides, warfare agents, detergents, solvents, heavy metal compounds, synthetic drugs, etc.), which have emerged over the past one hundred years. They exist in various forms (inorganic, organic, solids, liquids, gases, aerosols, etc.), either separately or when mixed with other substances.

Various **forms of toxic substances** exhibit different levels of toxicity.

Nickel Compound	***Form***	***Toxicity***
Ni	Lustrous metal	Nickel allergy, nickel dermatitis
NiO	Black powder insoluble in water	Possibly carcinogenic
$NiSO_4$, $Ni(NO_3)_2$	Green crystals soluble in water	Nephrotoxicity
$Ni(CO)_4$	Gaseous compound	Proven to be carcinogenic

Table 33.3. Examples. Nickel compounds.

Dust Particles	***Example of Profession***	***Consequence of Inhalation***
Biological origin (vegetable dust, animal fur, etc.)	Agricultural workers	Allergy, pulmonary involvement (farmer's lung, bronchial asthma, etc.)
Contain asbestos particles	Workers in asbestos mines, brake lining manufacture, construction, etc.	Asbestosis, increased risk of lung and pleura tumour
Contains silica particles	Quarrymen and coal miners	Silicosis, coal worker's pneumoconiosis

Table 33.4. Examples. It is important to identify the dust type in cases of frequent exposure to dust inhalation.

Xenobiotics such as organophosphates exist, which have adverse effects in any dose, or the adverse effect of which is proportional to the dose. There are also some xenobiotics, which are beneficial to the body when administered within a certain optimal dose range, however their insufficiency is bad, and, conversely, overdosing has toxic effects. A typical example is selenium, a useful antioxidant on the one hand, yet a highly toxic poison on the other.

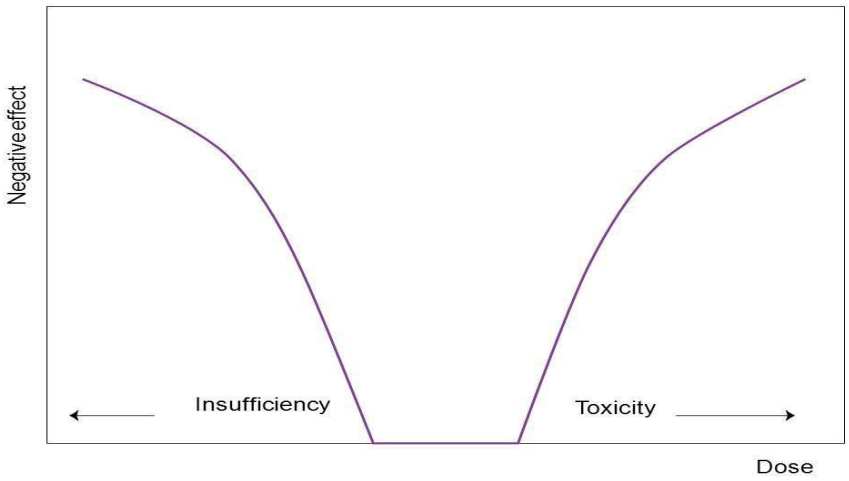


Figure 33.2. Shows a diagram revealing the dependency of negative effects on the dose of selenium.

Sources of toxic substances are present in the external, working and home environments.

External environment	Water, air, soil, food, plants, or animals may become contaminated with toxic substances. Spread in a large space, meaning concentrations are lower.
Home environment	Paints, detergents, disinfectants, impregnating agents (formaldehyde), drugs, plants, etc. Enclosed space, long-term contact.
Working environment	Different forms of technology use organic solvents, insecticides, additives, chemical compounds and other substances. High concentrations if the workplace and workers are not equipped with protective equipment (exhausts, PPE).
Accidents	Mistakes, accidental leakages during breakdowns, combustion products of fires, etc. Concentrations may be extremely high.

Table 33.5. Examples

33.4.3. Rapidity and Intensity of Toxic Substance Effects

Rapidity and intensity of toxic substance effects are influenced by many factors. Table 6 shows the most important factors.

Timing of exposure	Acute (shorter than 24 hrs), chronic (longer than 3 months).
Mode of exposure/route of poisoning	Ingestion, inhalation, transdermal, sniffing, intravenous, intramuscular – injections, etc.
Mode of poison effect	Local (caustic agents), general (carcinogens, mutagens).
Site of action	Hepatotoxic agents are activated upon entering the liver; nephrotoxic agents into the kidneys; CNS depressants into the brain, while irritating agents are already activated once they come into contact with the skin or lungs, and haemotoxic as soon as they enter the bloodstream.
Mode of biotransformation (metabolism)	Usually a process where lipophilic substances change into hydrophylic ones, which are easier to eliminate from the body, mostly in two phases: Phase I: Metabolic transformation includes oxidation, reduction and hydrolysis. Phase II: e.g. conjugation with glucuronic acid.
Route of eliminating poison from the body	Either directly as a poison or in the form of various metabolites, in urine, stools, exhaled air, bile, sweat, hair, breast milk, etc.
Dose of poison	The effect is generally directly proportional to the dose of poison.

Form of poison	Physical and chemical properties such as solubility are important. Refer to Table 3 for examples of different forms of nickel compounds.
Enhancement or elimination of the effect of poison	A big proportion of poisonings is not caused by one poison but rather a mixture of interacting substances.
Poison habit	The basis of drug addiction consists of the habit and need to increase doses to obtain the same effect.
Resistance (susceptibility)	Children are generally more susceptible to the effects of poison than adults. Women are usually more sensitive than men.
Health condition	If an individual is compromised by poison as well as another disease, the adverse effects of poisoning are more serious. For example, impaired kidney function will reduce the capability to eliminate poison from the body.

Table 33.6. The most important factors

33.4.4. Legislation

The Czech Republic is a member of the European Union, therefore, the following regulations of the European Parliament, Council and Commission apply to the area of chemical substances:

Regulation (EC) No. **1272/2008** of the European Parliament and of the Council entered into force in 2009. The regulation stipulates rules for the classification of substances and mixtures as hazardous, and rules for labelling and packaging hazardous substances and mixtures. European documents take precedence over basic Czech legislative documents concerning toxic substances.

Act 350/2011 Coll., Section 122, Act on chemical substances and chemical preparations and on the modification to some laws. For example, the document can be found on: http://aplikace.mvcr.cz/sbirka-zakonu/start.aspx

Decree 402/2011 Coll., Section 140 and annexes thereto, A decree to execute some provisions of the act on chemical substances and chemical preparations and on modifying several laws concerning the classification, packaging and labelling of hazardous chemical substances and chemical preparations. *Almost 700 pages of text.*

You can locate the document on: http://aplikace.mvcr.cz/sbirka-zakonu/start.aspx

The category of substances classified as **T+**, i.e. **highly toxic substances**, is important.

For example, these substances include sodium azide, a common part of commercially supplied sets for many routine biochemical tests. A frequently asked question is whether the sets should also be considered as highly toxic T+ substances. It is specified in Annex 2 of Physical act, where is writen that if the concentration of highly toxic substance in the product is under 2.5%, the product is not considered to be highly toxic.

Note: Normal sodium azide concentration as a preservative in laboratory sets is 1%, therefore commercial sets are not considered to be highly toxic products.

33.4.5. Important Terms Related to Poisons

Toxic Dose	The lowest amount of poison capable of causing signs of poisoning.
Lethal Dose	Commonly referred to as LD50, i.e. the dose that causes 50% of afflicted individuals to die. The dose expresses the individual sensitivity of an organism to the poison. These doses differ greatly; e.g. morphine LD50 is 900 mg/kg, botulin only 0.00001 mg/kg.
Therapeutic Dose	Single and daily doses are used for drugs, and express what doses are therapeutically effective yet still toxicologically tolerable.
Therapeutic Width	The difference between therapeutic and toxic doses. Safer drugs have a large therapeutic width (atropine); drugs with a low therapeutic width (digoxin, lithium) are less safe.
Biological Half-Life	The time required for the body to eliminate one half of the total amount of poison. The problem is that metabolites produced from the poison are sometimes much more toxic than the original poison. Examples: Alcohol – acetaldehyde, ethylene glycol – oxalic acid

Xenobiotics	In general terms, foreign substances that do not normally exist or exist in insignificant amounts in an organism, and are introduced voluntarily (drugs) or involuntarily (contaminants, mistakes, etc.) to the organism. Some of these substances are biologically inactive, but most of them are biologically active and may have positive or negative effects on the organism.
Biogenic Substances	Foreign substances useful for the organism, the negative effects of which may not only be due to overdosing but also due to insufficiency (e.g. selenium, iodine and other trace elements).
Hazard Symbols	Hazardous properties in substances can be identified by pictograms or letters on packages and catalogues: E EXPLOSIVE O OXIDIZING F+ EXTREMELY FLAMMABLE F HIGHLY FLAMMABLE T TOXIC T+ HIGHLY TOXIC Xn HARMFUL C CORROSIVE Xi IRRITANT N DANGEROUS FOR THE ENVIRONMENT See Annex 4 to Decree 402/2011 Coll. For example, the document can be found on: *http://aplikace.mvcr.cz/sbirka-zakonu/start.aspx*
Risk Phrases (R-Phrases)	Standard phrases identifying the specific risk of a substance, which have to be shown on the packaging of chemicals. There are 68 basic R-phrases in total, which may appear in combinations. For example, R26/28 indicates that the substance is higly toxic by inhalation and if swallowed. Note: The identification of R-phrases is currently being changed to **H-phrases** (Hazard phrases). See Annex 5 to Decree 402/2011 Coll. For example, the document can be found on: *http://aplikace.mvcr.cz/sbirka-zakonu/start.aspx*
S-Phrases (Safety Phrases)	*Standard instructions for safe handling of chemicals. There are 64 basic S-phrases in total, which may also be used in combinations. Example: S24/25 indicates that skin and eye contact with the substance should be avoided, i.e. that protective gloves and goggles should be worn.* Note: The identification of S-phrases is currently being changed to **P-phrases** (Prevention phrases). See Annex 6 to Decree 402/2011 Coll. For example, the document can be found on: *http://aplikace.mvcr.cz/sbirka-zakonu/start.aspx*
MAC	Maximum concentration of a toxic substance allowed. Usually specified for the concentration of a hazardous substance in the air, water, food, etc. Example: The MAC for air toluene in the workplace is 200 mg/ m^3.

Maximum Permissible Level	The highest concentration, which is still permissible, e.g. during professional handling of a toxic substance, in terms of the onset of negative effects of the poison. The value is affected by many factors (economics, health policy, toxicity level of the substance). Example: MPL for professional lead exposure is 400 µg/l, but only 100 µg/l for professionally non-exposed children.

Table 33.7. Examples

Labelling of Chemicals

Chemicals must be properly labelled and hazards identified.

Examples of warning hazard symbols according to Regulation (ES) No. 1272/2008 of the European Parliament and of the Council:

Figure 36.3. Caustic to skin – Flammable – Carcinogenic – Highly toxic

Figure 36.4. Toxic – Toxic to aquatic life – Oxidiser – Gas under pressure

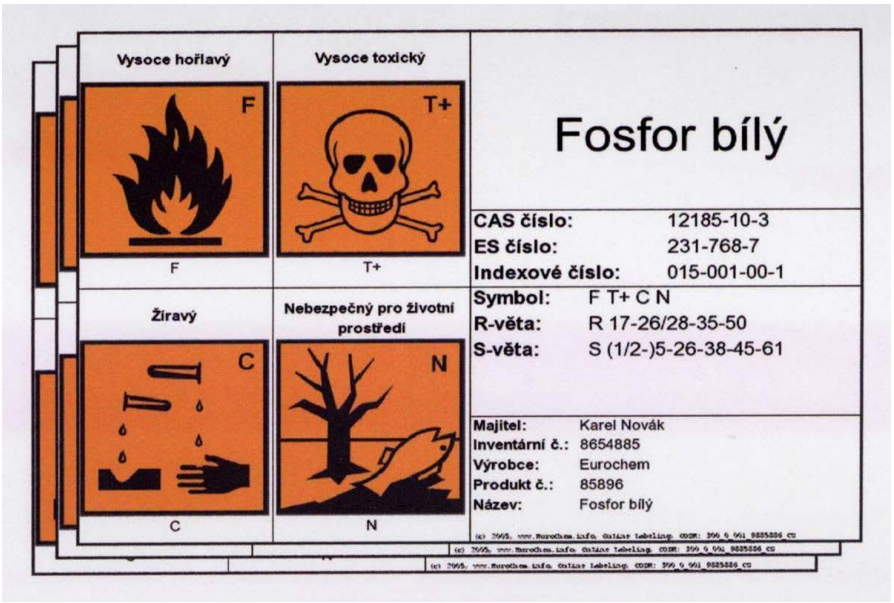


Figure 33.3. Example of a label on the chemical bottle with hazard symbols and R and S-phrases:

33.4.6. Toxicokinetics – "What the Body Does with the Hazardous Substance"

The processes which take place when the body is exposed to toxic substances can be outlined as follows:

Exposure → Absorption → Distribution (+ Excretion) → Biotransformation (metabolism) → Molecular target → Toxic action (negative effect induction)

Exposure

Exposure may be single, repeated or chronic. There are different routes of entry in the body depending on the situation and character of the hazardous substance (swallowing poison, inhalation – lungs, absorption through the skin).

Absorption of poison into the body

The hazardous substance has to penetrate through different barriers (membranes) in the intestine or lungs, for example, before it enters the blood stream. The rate of absorption is affected by the physical and chemical properties of poison, solubility, concentration and the mode of exposure.

Distribution of poison in the body

Following penetration into the blood stream, the poison is distributed, i.e. transported to various organs and tissues.

Biotransformation (metabolism) of poison

The objective of biotransformation (metabolism) can be briefly characterized as a process where lipophilic substances transform into more hydrophilic substances, which can be eliminated from the body more easily.

The ***first phase*** includes oxidations, reductions, additions or removals of functional groups (-OH, -COOH, $-NH_2$, etc.)

The ***second phase*** consists of conjugation with different substances such as glucuronic acid.

The course of biotransformation is affected by state of health, liver function (major biotransformation organ in the body), age and other factors.

Elimination (excretion) of poison

Biotransformation eliminates most substances from the body. There are many ways of excreting poison from the body. The major role in excretion is played by the kidneys (note that poison may be kept in the body due to renal impairment). A common way is the lungs (exhalation), some amounts are excreted in bile from the liver, some hazardous substances deposit in the hair (interesting), and poison may also be excreted in breast milk (important).

Cumulative poisons, biological half-life

The term **biological half-life** defines the time required for the body to reduce the toxic substance level to half. Substances have very different biological half-lives. For example the biological half-life of toluene is 2 - 3 hours, while the biological half-life of cadmium is 20 years.

The term **cumulative poison** refers to a toxic substance with a long biological half-life, which slowly excretes from the body. When the body is next exposed, the hazardous substance from the previous exposure still persists in the body, and poison slowly accumulates.

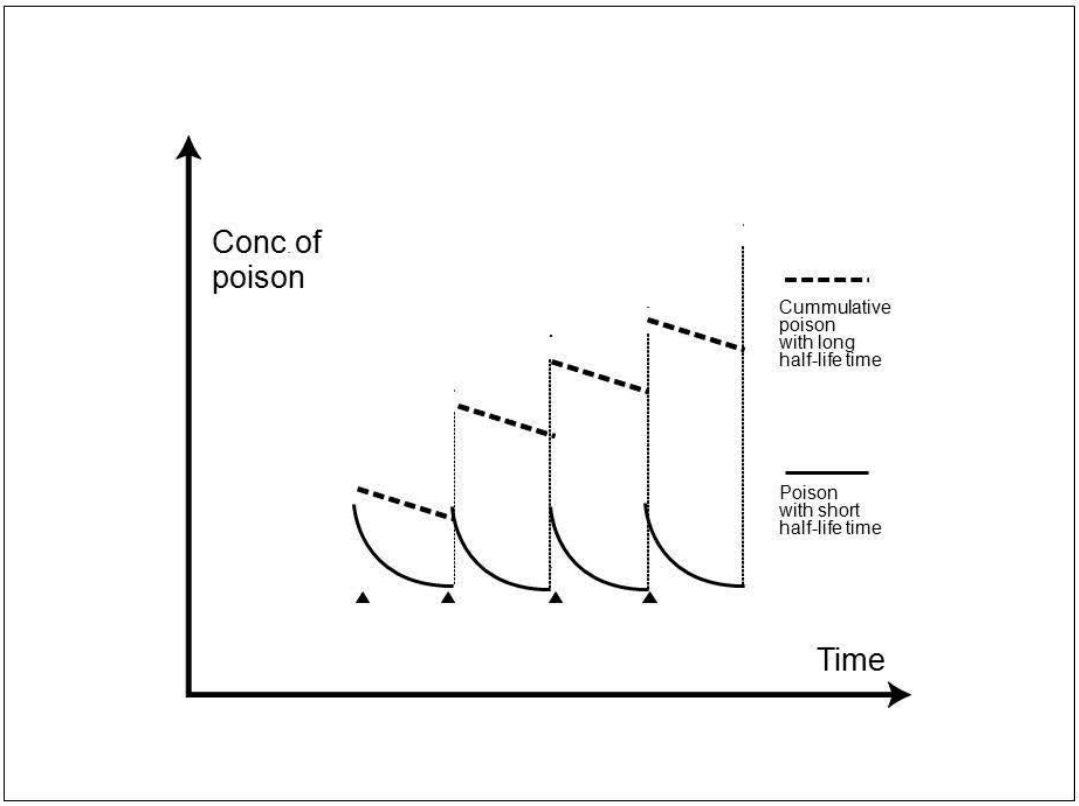


Figure 33.4. Shows a diagram of the difference between repeated exposures to substances with short and long biological half-lives.

The releasing of accumulated hazardous substances from the deposit site may sometimes pose a problem.

Example: Shortly after childbirth, a woman embarks on a weight-reducing diet to quickly regain her former shape. However, this releases toxic substances accumulated in fatty tissues, which may endanger the breast-fed baby.

33.4.7. Toxicodynamics – "What the Hazardous Substance Does with the Body"

Site, mode and time profile of effect

Some poisons act specifically on certain organs (e.g. lead and lead compounds impair haemoglobin synthesis, cadmium impairs renal tubules, etc.), other toxic substances act throughout the entire organism (e.g. carcinogens, mutagens, etc.).

Dose-effect

The relationship between the dose (i.e. the amount of toxic substance to which the human is exposed) and the size (i.e. significance) of negative toxic effects does not only depend on the size of the dose but also on a variety of other factors and circumstances.

Time profile and duration of exposure	Acute (shorter than 24 hrs), chronic (longer than 3 months).
Mode of exposure/route of poison entry into the body	Ingestion - GIT, inhalation - lungs, transdermal - skin, snuffing, intravenous, intramuscular – injections, etc.
Mode of poison action	Local (caustic agents), general (carcinogens, mutagens).
Site of effect	Hepatotoxic agents are not activated until they enter the liver; nephrotoxic agents into the kidneys; CNS depressants into the brain, while irritating agents already become activated when they come into contact with the skin or lungs, and haemotoxic as soon as they enter the bloodstream.

Mode of biotransformation (metabolism)	Usually a process where lipophilic substances change into hydrophylic, which are better eliminable from the body, mostly in two phases: Phase I: Metabolic transformation includes oxidation, reduction and hydrolysis. Phase II: e.g. conjugation with glucuronic acid.
Route of poison elimination from the body	Either directly as a poison or in the form of various metabolites, in urine, stools, exhaled air, bile, sweat, hair, breast milk, etc.
Form of poison	Physical and chemical properties such as solubility play an important role.
Enhancement or elimination of poison effect	A big proportion of poisonings is not caused by one poison but a mixture of interacting substances.
Poison habit	The basis of drug addiction consists in the habit and need to increase doses to obtain the same effect.
Resistance (susceptibility)	Children are generally more susceptible to poison effects than adults. Women are usually more sensitive than men.
Health condition	If an individual is compromised by poison and another disease, the adverse effects of poisoning are more serious. E.g. impaired kidney function will reduce the capacity to eliminate poison from the body.

Table 33.8. Examples

Methods of evaluation (quantification)

There are different methods and terms to quantify toxicity and characterize a specific substance in terms of toxicity.

Lethal Dose – usually denoted as LD_{50}	The dose that causes one half of individuals from an exposed set to die
Toxic Dose	The lowest amount which causes poisoning
Therapeutic Width of Drugs	The difference between the toxic dose and therapeutic dose required for treatment
Maximum Allowable Concentration (MAC)	The highest concentration in the air acceptable in the workplace
Maximum Permissible Level in Biological Exposure Tests (BET)	The highest allowable concentration in the biological material of exposed organisms

Table 33.9. Examples

Differences in toxicity of different substances are in orders of magnitude.

Alcohol	10.000
NaCl	4.000
Morphine	900
Paracetamol	142
Strychnine	2
Nicotine	1
Curare	0.001
Botulinum toxin	0.00001

Table 33.10. Examples. Comparison of LD_{50} (mg/kg) of known substances

33.5. Toxicological Tests

The spectrum of toxicological tests is very wide and varied. The media publish information regarding analyses of suspected substances (landfills, contaminated food, baby toys, etc.), and/or about toxicological tests of the environment (outdoor or indoor air, drinking or waste water, contaminated soil), etc.

However, these indications usually do not occur in clinical laboratories. Toxicological tests in clinical laboratories

generally include analyses of different biological materials taken from potentially intoxicated persons.

33.5.1. Reasons for Toxicological Tests

Acute intoxications	Accidental (unlabelled or wrongly labelled containers), intentional (suicides, crimes) intoxication, unconscious patient with acute signs of intoxication. Examples: Serum ethylene glycol after ingestion of Fridex from a soda water bottle. Serum paracetamol or salicylates after intoxicating a child or a suicide attempt.
Chronic intoxications	Differential diagnosis of signs of long-term exposure to hazardous substances. Examples: Delta-aminolevulic acid and coproporphyrin III in urine after a long--term exposure to lead.
Drug abuse (toxicomania)	Capture of new addicts, check of compliance with withdrawal treatment rules, checking whether people in control groups of studies do not use drugs. Examples: Drugs and drug metabolites (THC, methamphetamine – pervitine, opiates, cocaine, ecstasy, buprenorphine, benzodiazepines), etc. in urine.
Professional intoxications	Biological exposure tests, indemnification for occupational diseases. Examples: Mandelic acid in urine after professional exposure to styrene, or trichloroethanol and trichloroacetic acid after exposure to trichloroethylene.
Crimes	Examples: Serum ethanol and osmolality in drivers, or demonstration of drug abuse at work.
Therapeutic Drug Monitoring (TDM)	Monitoring of administered drugs (especially those with a narrow therapeutic scope). Examples: Digoxin, methotrexate, tacrolimus, sirolimus, theophylline, vancomycin, gentamicin, phenytoin, valproate, carbamazepine, cyclosporin A, phenobarbital, etc. in the serum.
Experimental intoxications	Animal research. Example: Detection of Cd in tissues after conducting experimental research cadmium intoxication on a set of mice.

Table 33.11. Examples

33.5.2. Types of Toxicological Tests

33.5.2.1. Unknown Intoxication (Toxicological Screening)

Example: Two hours after midnight, a man unable to communicate, with acute signs, is brought to the hospital. It is necessary to find out as soon as possible whether he is intoxicated.

From an analytical point of view, it is one of the most difficult tasks. The condition may involve any toxic substances at any concentration. The spectrum of the most common poisonings slowly changes with newly emerging drugs and substances and their availability. If a substance is positively demonstrated, it is important to keep in mind that mixed poisonings with multiple poisons often occur, and then it is necessary to decide which component dominates. Any additional information that one is able to procure is very important (e.g. the estimated time of intoxication, containers, residues, tablets found, texts, drugs given as first aid, etc.). In general, it is advisable to take the widest spectrum of materials (blood, urine, hair) as possible in sufficient amounts.

This difficult task requires extensive laboratory equipment with suitable analyzers.

The speed of response of the laboratory is very important in this type of toxicological test. Clinicians often make do with the qualitative finding, which may be quantified at a later stage.

33.5.2.2. Monitoring of Known Exposure

Examples: Therapeutic drug monitoring before and after administering another dose of the drug or professional exposure to a known substance such as organic solvent toluene, or alcohol level analysis in drivers and workers.

The biological half-life of the monitored hazardous substance should be taken into account and suitable material for capturing the poison should be selected accordingly. Example: It is useless to take urine to ascertain the toluene metabolite level one week after intoxication, because the half-life of toluene is about 2 - 3 hours. On the other hand, THC metabolites from chronic abuse can be demonstrated in urine up to even several weeks after the last application.

This test can be scheduled beforehand; the speed of response is not very important, while adequate accuracy, i.e. trueness and precision, is important (refer to the Analytical Properties of the Methods).

33.5.3. Materials Used for Toxicological Tests

The spectrum of materials used for toxicological tests in clinical laboratories is wide-ranging.

Blood, serum	Invasive sampling (problematic in very little children), suitable only a couple of hours after intoxication.
Urine	Non-invasive; poison is usually concentrated in urine, appears in urine later than in blood; it is possible to conduct the test for several days; problems associated with “adulteration”.
Hair, nails	A unique opportunity to look at the history (1 cm of hair corresponds to about 1 month). If possible, the hair should not be chemically damaged and contaminated (dyes, shampoos).
Exhaled air, saliva	Useful to demonstrate ethanol or drug ingestion.
Other (gastric content, tissues)	Uncommon in clinical laboratories - this is the domain of forensic medicine and pathology.

Table 33.12. Examples

33.5.4. Focus of Toxicological Tests

The focus (objective) of toxicological tests depends, among other things, on the time passed from intoxication to biological material sampling.

Hazardous substance detection	Usually in blood (serum, plasma) immediately after intoxication.	Examples: Toluene, lead, drugs before and after administration of the next dose, etc.
Hazardous substance metabolite detection	Usually in urine; the test can be made even up to several days after intoxication takes place.	Examples: Hippuric acid, mandelic acid, drug metabolites, etc.
Determination of effects (negative effects of the hazardous substance)	Targeted tests depending on the type of damage. Examples: Liver tests, haematopoiesis tests, kidney tests, etc.	Examples: Liver tests after paracetamol poisoning, porphyria tests after lead poisoning, etc.

Table 33.13. Examples

33.5.5. Forensic Issues

Forensic purposes generally require results confirmed by at least two methods based on different principles (e.g. gas chromatography and Widmark test for ethanol).

Toxicological tests made in clinical laboratories of biochemistry are mostly intended for acute clinical cases (e.g. a non-communicating person with signs of intoxication, requiring fast therapeutic action). They are not very useful for forensic purposes. One of the reasons is insufficient safeguarding the sample against accidental or intentional tampering (swapping or contaminating the sample, etc.).

33.6. Analytical Techniques

Toxicology requires very specialized equipment, not only for the analysis itself but also for the pre-analytical preparation of specimens. The analysis itself requires the availability of a wide spectrum of analyzers and devices based on different principles.

Spectral systems	UV, VIS and IR spectrophotometry, flame photometry, AAS, ICP-MS, etc.
Chromatographic systems	TLC, GC-MS, LC, HPLC-MS, etc.
Immunochemical systems	FPIA, MEIA, AMIA, CMIA, EIA, ELISA, etc.
Other	Homogenization, lyophilization, isolation, mineralization, pre-concentration, etc.

Table 33.14. Examples

Selection of suitable analytical technique initially depends on the task, which can be very varied in toxicology, and usually involves individual situations rather than an extensive series of similar samples typical for routine biochemistry.

Toxicological screening for unknown substance intoxication often starts with the application of simple immunochemical test cards. These systems usually only perform qualitative analyses, their specificity is rather problematic (cross reactions) and they are available only for a certain limited amount of substance types (e.g. 10 basic types of classical drugs), but they are very easy and fast to use at any time of the day and night, because they do not need the presence of a skilled specialist. The qualitative result, which is sufficient to clinicians at the first stage, can be quantified at a later stage, if necessary, using a suitable chromatographic or spectrophotometric method. Chromatographic techniques (GC-MS, HPLC, TLC, etc.) have a good specificity and make it possible to determine practically any organic substance, however they are lengthy, complicated, and require an experienced and qualified chromatographer. ICP-MS is the most suitable technique for screening inorganic substances such as heavy metals, though it is not used in clinical laboratories. Collaboration with specialized laboratories is required in indicated cases.

The situation is simpler, if **monitoring of the known substance intoxication** is needed. If the substance is known, the most adequate analytical technique can be used. Different commercially available quantitative immunochemical procedures such as FPIA, MEIA and others are used for therapeutic drug monitoring (TDM). Biological exposure tests (BET) in industrial toxicology and hygiene (test of toxic substances directly in the biological material of exposed persons) employ various techniques such as targeted AAS for a specific heavy metal, etc.

33.6.1. Pre-Analytical Phase (Sampling and Sample Preparation)

Since samples are varied, they require different types of preparation before performing an analysis:

Solid samples such as hair must firstly be decontaminated (washed alternately in the aqueous and organic phases and dried), homogenized (by mixing, grinding, cutting, freezing), sometimes dried to a constant weight, while the amount of sample taken for analysis must always be weighed on the analytical balance. The sample has to be subsequently converted into a solution, i.e. mineralized (today most often by microwave-assisted mineralization under pressure in a closed system with a mineralizing agent admixture). Pre-concentration is sometimes required to obtain a measurable concentration (e.g. almost complete evaporation of the mineralization residue and subsequent dissolving of the mineralization residue in a small amount of a suitable solvent).

Even **liquid samples** often have to be treated before toxicological testing (e.g. adjustment of a sample pH by adding suitable agent). Some substances (e.g. drugs) have to be isolated before the test from a complex organic matrix of the biological specimen by extracting them into an adequate organic solvent, or alternatively chromatographic separation is used. This will reduce the number of potentially interfering substances and simplify identification. Solid-phase extraction (using silica gel, cation exchange, anion exchange) has become more and more widely used of late.

Gas and volatile substances are isolated by distillation or headspace technique, in which the equilibrium gas phase above the liquid sample is analyzed.

33.6.2. Quality Control

The control of toxicological analysis quality is specific in some aspects. Toxicological tests mostly do not contain a large series of more or less identical samples, which is usual in routine clinical biochemistry. Requirements are often unique, or repeat at long or irregular time intervals. It is sometimes problematic to obtain suitable reference material

with the same matrix. Interlaboratory control systems are not commonly available.

Nevertheless, quality control is important and absolutely essential, and as a minimum it is necessary to verify the quality parameters of the method used (precision, trueness, accuracy, sensitivity, specificity, recovery, etc.) (Refer to Analytical Properties of Methods). It is clear what the negative consequences can be for the tested person due to a falsely positive (innocent person is included in the list of drug abusing persons) or, conversely, a falsely negative finding (no therapy for the late after-effects of intoxication caused by an unrecognized substance).

33.7. Most Common Forms of Intoxication

33.7.1. Haemoglobin Disorders

The best known and most common include **carbon monoxide (CO)** intoxication caused by incomplete coal or gas combustion (gas boilers in bathrooms) or exhaust gas accumulation in an enclosed space (garage). The frequency of poisonings is presently growing due to tightly sealed plastic windows. During intoxication (signs include headache, weakness, fatigue, dizziness), CO binds to haemoglobin to form **carboxyhaemoglobin (COHb)**. The fact that COHb is unable to carry oxygen will manifest itself in most severe poisonings by tissue hypoxia with lactate acidosis. The CO bond to haemoglobin is reversible however is roughly 200 times stronger than the oxygen bond. Serious cases are therefore treated under higher oxygen pressure in the hyperbaric chamber. COHb is tested by acid-base balance analyzers.

COHb is also present in the blood of non-intoxicated people, but not higher than 5%. COHb levels over 10% can be found in smokers. So the information as to whether the patient is smoker or non-smoker is very important in checking for possible CO intoxication.

Another possible poisoning, disabling the transport function of haemoglobin, is the oxidation of Fe^{2+} in haemoglobin to Fe^{3+} through the effect of **nitrites, nitrates and some drugs** to form methaemoglobin (MetHb). This is especially dangerous in newborns, which have an immature MetHb-reducing system. This is why the permissible level of nitrates in drinking water for infants is much lower than the level for adults.

There are also other, luckily rare, types of poisoning such as **sulfhaemoglobin (SulfHb)** formation due to **hydrogen sulphide** poisoning, or **cyanhaemoglobin (CNHb)** due to hydrogen cyanide and other cyanine poisoning.

33.7.2. Heavy Metal Poisonings

Heavy metals may be defined in a variety of ways. Some define heavy metals in terms of density, others in terms of atomic weight or toxicity. Heavy metals used by humans contaminate the external and working environments. Most of them have long biological half-lives, so they slowly accumulate in exposed organisms. Typical heavy metals include lead, cadmium, mercury, copper, chromium, manganese, aluminium, vanadium, nickel, etc.

In terms of the history of heavy metals, lead has most probably been used the longest (PbO as a sweetener, waterworks, leaded petrol, accumulators, paints, radiation protection and other uses). Pb has a strong affinity to bones, where it accumulates. Pb is able to substitute other biologically important elements (Ca, Fe, Zn) in the bond to -SH, $-NH_2$ and -COOH groups, in proteins and other molecules. When bound to Pb, these substances work differently, leading to many negative effects (toxicity).

Inhibition	Effect
d-ALA - dehydratase	Impaired haemoglobin synthesis
NMDA receptors in the brain	Worsened long-term memory

Table 33.15. Examples

Many “famous” Pb intoxication stories can be found in history:

- Ludwig van Beethoven suffered from chronic abdomen pains from youth; his bones and hair were analyzed in 2002 to find high lead concentrations.
- Francisco Goya – painter, intoxicated by lead paints.
- Sir John Franklin – polar explorer, died along with all his team members in Greenland due to tins sealed

with lead.

Lead poisonings do continue to occur today (both at home and abroad), often in various unexpected circumstances:

- A doctor family became acutely intoxicated from a ceramic pot with Pb glaze, in which they made lemon tea because they did not want to drink chlorinated tap water.
- A young woman had Pb poisoning from using an over-the-counter product, Femikalp. Unfortunately, dried Indian plants contained in the product to enhance vital functions were heavily contaminated with lead.
- A large group of relatives was extremely exposed to lead because they confused red paprika powder with red lead, a powder primer (triplumbic tetroxide) of exactly the colour of red paprika.
- A group of restorers was exposed to Pb while restoring 300 year old wall paintings in the Teplá monastery. The old historical paints contained Pb.

Gradual improvement of knowledge about toxicity of substances, here the specifically discussed Pb and Pb compounds, leads to the gradual lowering of maximum permissible levels. *For example, the maximum permissible level of Pb in the blood of children was reduced from 600 µg/l in the 1960 - 70 to 100 µg/l used at present.*

Treatment of intoxicated people primarily involves eliminating accumulated lead from the body by injections of chelates that release Pb deposited in bones and excrete it through urine.

Cadmium is associated with the globally known itai-itai syndrome, an effect of Cd intoxication from factory waste dumping to seawater and the consequent consumption of contaminated fish. Damage to proximal renal tubule cells is typical and may even lead to tubular necrosis. Cd concentration is tested in serum, urine and hair; optionally beta-2--microglobulins may be tested in urine as an indicator of tubular injury.

Mercury is also connected with the Japanese name Minamata as a synonym for human poisoning by Hg from the environment. Toxic actions of inorganic and organic mercury compounds differ. Neurological injuries are most commonly reported, but the kidneys and lungs are also often damaged. Early signs are non-specific and include weakness, fatigue, headache, later the characteristic shiver and tremor (handwriting test is recommended). Mercury concentration in blood, urine and hair is measured.

33.7.3. Alcohol Intoxication

The use of **ethanol** is widespread in the CR with generally known symptoms.

The use of greater doses can be considered intoxication with acute and chronic effects. Like all forms of alcohol, which generally metabolize to aldehyde and then to acid, ethanol metabolizes due to the effect of an alcohol dehydrogenase enzyme to form acetaldehyde and subsequently acetic acid. A markedly greater serum osmolality occurs in ethanol and, analogically, other forms of alcohol intoxication. Osmolality can be measured or computed from the following formula: 2 x Na + urea + glucose.

The difference between the measured and computed values (osmolal gap) is evaluated. Refer to the chapter about osmolality.

Methanol intoxication is relatively frequent and is due to substituting ethanol for methanol. Methanol metabolizes to formaldehyde and then to formic acid. It is a highly toxic substance (a lethal dose for adults amounts to 50 ml of methanol). The typical permanent consequence of methanol intoxication is optic nerve injury leading to blindness. Severe metabolic acidosis also develops. Analogical metabolism is used as a form of treatment, while ethanol is administered as an antidote. The alcohol dehydrogenase enzyme then “busies itself” with excess ethanol and forms limited amounts of formaldehyde responsible for optic nerve injury.

Similarly, confusions (unlabelled containers) or suicides involve **ethylene glycol** contained in anti-freezing fluids (Fridex). Ethylene glycol metabolizes to oxalic acid. This harms the kidneys. Severe metabolic acidosis develops. A flood of calcium oxalate crystals can be found in urine. The increase to serum osmolality is significant. Ethanol is administered as an antidote here as well in addition to other treatments such as dialysis in serious cases.

33.7.4. Drug Intoxication

Drug intoxication occur not only as a result of suicide but also due to mistakes (a child confuses coloured pills for sweets) and errors (a different dose or higher drug concentration is administered by mistake). The drug concentration

measured in blood is evaluated in relation to the time passed between intoxication and sampling. There are various nomograms for this purpose.

For example, it is possible to determine by drawing a line in the nomogram for the evaluation of significance of paracetamol concentration that 50 mg/l (50 µg/ml) paracetamol measured in a sample taken 8 hours after intoxication rules out significant liver injury.

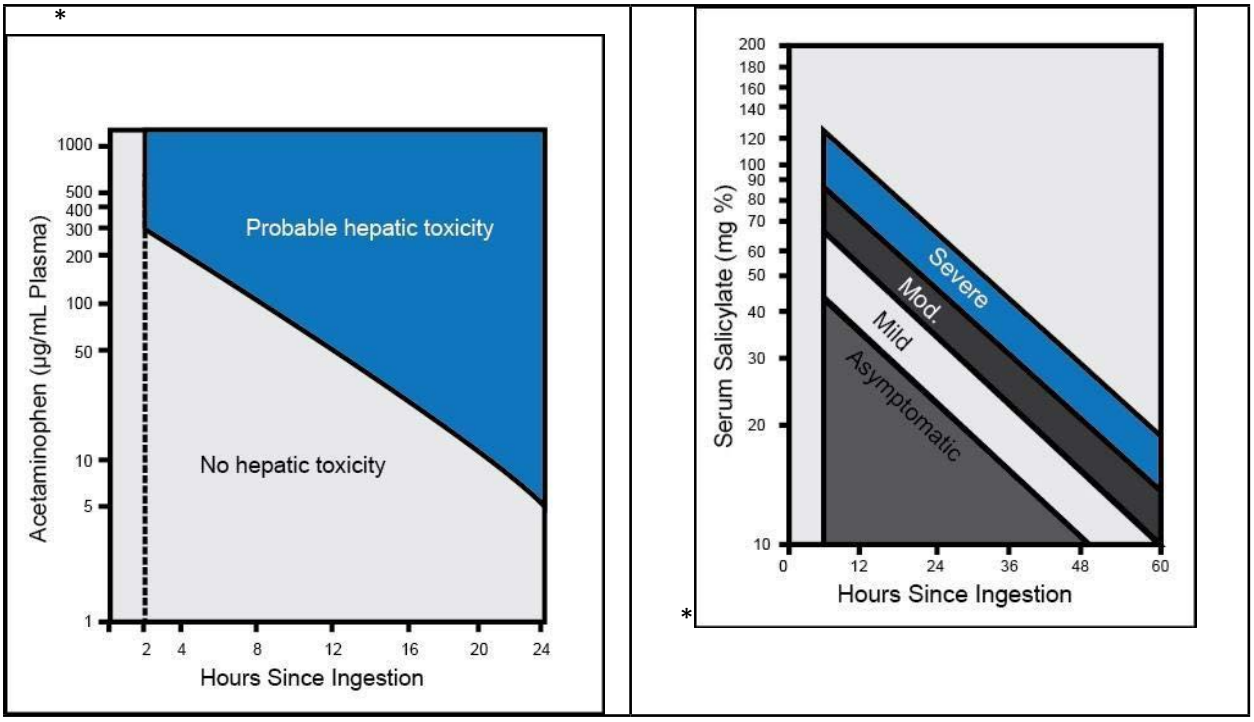


Figure 33.5. Nomograms evaluating the significance of the probable toxic effects of paracetamol and salicylates at the measured level and time from intoxication

Paracetamol (acetaminophen) contained in many drugs such as Paralen may cause severe injury to the liver due to overdosing.

Salicylates (acetylsalicylic acid) contained in drugs such as Aspirin cause metabolic acidosis, after penetrating into the CNS irritate the respiratory centre, which leads to respiratory alkalosis.

(Refer to the chapter about the ABB.)

33.7.5. Drugs

The issue of drug-taking among humans has been present throughout history, i.e. often involving abuse and habit-forming substances. Each society has adopted some basic rules for how to handle "their" drugs over time; hence the situation was generally tolerable, although personal tragedies did of course occur. Problems escalated worldwide when drugs were introduced to new places as part of globalization.

Example: We know from history the damage alcohol (commonly used in Europe) caused to Indians in America. Conversely, opiates from Asia or cocaine from South America are now causing huge problems in Europe.

The situation gets even more complicated by the fact that new, synthetic drugs are emerging in addition to long-known, natural drugs. The new drugs are produced for gain, irrespective of the consequences faced by addicts (ruined health, crime, negative social impact on family and society).

Classic drugs

"Classic" drugs can be generally divided into four basic groups distinguished from each other by appearance, effects, course of addiction and extent of risk presented to the user.

Some of these substances are also used for therapeutic purposes (e.g. opiates as painkillers), other drugs are purely abused (e.g. stimulating substances from the family of amphetamins).

Laboratories commonly use simple aids (cards) making it possible to detect the presence of drugs on the immunochemical principle.

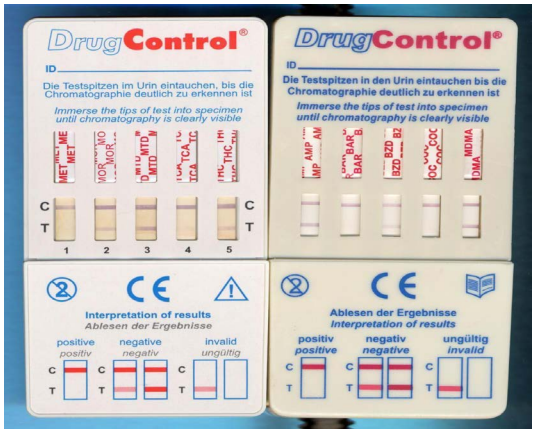


Figure 33.6. Example of a screening card to detect 10 types of abused substances found in urine. The evaluation is simple: 2 lines = negative, 1 line = positive

Cannabis drugs

Probably the best known is **marijuana**, a mixture of leaves, flowers and stalks of the Cannabis sativa (indica) plant. It is applied by smoking (joints), and symptoms include disorders of perception, emotions, thinking and short-term memory. It is also referred to as **THC** after the active substance tetrahydrocannabinol, and ranks among soft drugs. There has been a long ideological struggle between those who wish to legalize the drug (analogically to ethanol and nicotine), and their opponents, who argue by saying that THC is often an initial phase for using hard drugs.

Hashish is a hemp resin with similar but much stronger effects than marijuana.

Opiates

Opioids represent a wide group of alkaloids with a chemical structure similar to that of **morphine**, which can be found along with other substances (**codeine**, **thebaine** and **papaverine**) in the juice from unripe Papaver somniferum poppy-heads.

In addition to these purely natural opioids, there are a lot of semi-synthetic substances such as **heroin** (semi-synthetic morphine derivative) or **buprenorphine** (semi-synthetic thebaine derivative), known as SUBUTEX, a drug used for replacement therapy.

Another substance given to addicts for replacement therapy is the synthetic preparation **methadone**.

Opioids are applied primarily by intravenous injections, but also by smoking or sniffing. Application initially induces euphoria and pain control. Intoxication by a higher dose leads to sedation, somnolence, vomiting, contraction of the pupils or even respiratory arrest (naltrexone - a clinically used antidote is used).

As a result of addiction, discontinuation of habitual dose leads to withdrawal symptoms such as sweating, malaise, anxiety, depression, insomnia and fever. The addict will do anything to get another dose, which has to be stronger and stronger.

Stimulants

The best known stimulant is alkaloid **cocaine** isolated from the Erythroxylon coca bush growing in South America. It is applied by powder inhalation or chewing coca leaves. Cocaine is a strong CNS stimulant, induces euphoria and a sense of tirelessness. However, it causes a strong psychological dependence.

In addition, this group includes the synthetic drug **methamphetamine** (**meth** known as “glass”, “crystal” or “tik” and in CR as „pervitine“).

Ecstasy (party drug), a derivative of methamphetamine applied in the form of tablets, is very widespread among young people.

Hallucinogens

There are natural hallucinogens such as **psilocybin** contained in **magic mushrooms** (Psilocybe bohemica) or **tryptamines on toad skin**, and synthetic hallucinogens such as the best known **lysergic acid (LSD)**. As the name suggests, substances from this group cause euphoria and hallucinations.

For example, an urban legend popularizes the belief that after licking a toad a girl may think she sees a prince (like in fairy tales).

Regrettably, this also brings about depression, paranoia, liver and kidney damage.

New drugs

New, fully synthetic drugs so far unidentifiable by methods available to routine laboratories keep emerging.

GHB (gamma-hydroxybutyrate) known as “liquid ecstasy”

- 1,4-butanediol known as “thunder nectar”
- Ketamine, a veterinary anaesthetic, often mixed with ecstasy
- Tryptamines, synthetic and natural serotonin-like substances

Problem of “adulteration”

When monitoring people suspected of using addictive drugs, it is necessary to pay attention to “adulterating” efforts, i.e. attempts to make drug detection impossible by diluting the urine sample or by adding bleach, acid, base or other substances. This is why special detection cards exist, which, in addition to the drug detection test, also contain auxiliary tests demonstrating unwanted intervention in the sample by changed colour of the check field.

33.7.6. Other Poisonings

As mentioned in the introduction, one can be poisoned practically by anything, if the dose is large enough. However, covering the whole topic would exceed the scope of this text.

This is why only some examples have been provided:

- **Organophosphate** poisonings are common. Organophosphates are not only used as warfare agents (sarin, soman, etc.), but they are also commonly used in farming, for rodent control and other purposes. The maximum permissible level for professional activities is 30% depression of cholinesterase activity. People with acute intoxication die in convulsions as a result of acetylcholinesterase enzyme inhibition at the neuromuscular junction.
- Frequent chronic and also acute intoxications by **organic solvents** (toluene, xylenes, benzene, trichloroethylene, styrene, etc.) used for dying, painting, laminating, solving, cleaning and other purposes. The test involves the detection of urinary metabolites such as mandelic acid for styrene, hippuric acid for toluene, phenol for benzene, trichloroethanol and trichloroacetic acid for trichloroethylene and other substances. Organic solvent intoxication causes liver damage, abortions, addictions and other effects.
- There are many **toxic plants** in the environment and at home, which are responsible for many poisonings, most commonly of children.

Plant	***Toxic Substance (Symptoms)***	**Photo** (source: Internet – Wikipedia)
Giant hogweed	Phototoxic coumarins (causes difficult-to-treat photodermatitis in contact with the skin)	 *
Dieffenbachia	Oxalic acid (acidosis, kidney injury)	 *

Datura	Atropine alkaloid (abused – excitation, hallucinations)	 *
Belladonna	Tropine and atropine alkaloids (even death due to respiratory paralysis)	 *
European yew	Buxamine alkaloid (spasmodic poison, vomiting, colic, muscle pain)	 *
Euphorbia (crown-of-thorns)	Irritating latex (skin, mucosa and eye irritation, even loss of vision)	 *
Death cap	Cyclic peptides amanitine and phalloidine (liver and kidney damage after 3 - 4 days)	 *

Table 33.16. Examples

- A huge amount of various products are commonly used in households.

Product	***Comments***
Potassium permanganate	Available in the form of violet crystals under the name of HYPERMANGAN
Sodium hypochloride	In detergents and bleaches such as SAVO
Rat poison	Coloured granules (anticoagulant).
Acids	Lime-scale removers
Lyes	Drain cleaners
Bluestone	Copper(II) sulphate, protects plants against fungus, products such as KUPRIKOL

Table 33.17. Examples

33.7.7. Causes: Mix-Ups, Unlabelled or Wrongly Labelled Containers

One of the most frequent causes of poisoning is wrongly labelled or unlabelled containers. This leads to mistakes being made, often with tragic outcomes.

Examples:

Yellowish hydrochloric acid stored in a cognac bottle.

Trichloroethylene poured into a soda water bottle.

Also manufacturers often choose an inappropriate label design.

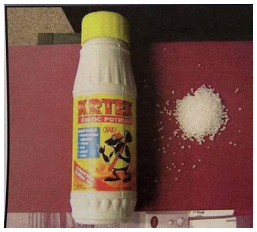
Product	***Photo***
Drain cleaner (KRTEK) – sodium hydroxide (picture of a mole is attractive for children and the contents looks like a powdered soft drink)	
Gel candles looking like drinks (resemble a fruit jelly)	

Table 33.18. Examples

70 - 90% of child poisoning cases occur in the home. Toddlers (1 - 3 years) are at the greatest risk as their motor skills develop, they can get where they could not get before, are inquisitive, and explore their surroundings by putting things into their mouths.

CHAPTER 34

34. Laboratory investigation of Ovarian and Testicular Disorders

Author: MUDr. Tomáš Franěk.

Reviewer: MUDr. Richard Pikner, Ph.D.

34.1. Hormone Tests for Ovarian and Testicular Disorders

Sex hormone or sex hormone metabolite testing forms an inherent part of gynaecological examinations, assisted reproduction, urology and endocrinology. Women are often tested for the diagnosis of infertility, menstrual cycle disorders, and differential diagnosis of hirsutism and virilization. Sex hormone testing is a supplementary examination for male gonadal diseases, infertility and cancer of the male genito-urinary system. As regards children, the major indication group is the differential diagnosis of hypogonadism and hyperfunction syndromes associated with symptoms of precocious puberty. Some specific aspects typical of this group have to be taken into account when testing sex hormones. One should keep in mind that hormones are excreted in a circadian rhythm, frequently in the pulse. The phase of the woman's current menstrual cycle also has to be taken into account. In addition, one should not ignore analytical differences in the reference values of each laboratory, so it is very advisable to use one laboratory for the tests and work only with the predefined, constant reference range within the relevant laboratory.

34.1.1. LH Tests

34.1.1.1. LH - Introduction

LH, luteinizing hormone or lutropin is a glycoprotein with a relative molecular weight of about 32 kDa, which is secreted through basophilic gonadotrophic cells of the anterior pituitary. Like the FSH, LH consists of an α-subunit common to some other anterior pituitary hormones, and a β-subunit specific to LH, which defines the biological and immunological specificity of LH and makes it possible to distinguish the LH from other glycoprotein hormones. LH secretion is controlled by hypothalamic decapeptide GnRH produced in the hypothalamus, from where it is released in the pulse (every 30 - 60 minutes) through the portal venous system to the pituitary gland. LH secretion is regulated centrally by CNS neurotransmitters and also peripherally through the action of sex steroid hormones. LH secretion is thus affected by a complex mechanism of positive and negative feedback. LH receptors rank among a group of receptors coupled with adenylate cyclase activated by G-proteins. The function of LH is to act on ovarian theca cells where stimulation of steroid synthesis takes place. The steroids subsequently transform into oestradiol with the contribution of FSH. In addition, LH is responsible for the final maturation of ovarian follicles and also invokes ovulation and early development of the corpus luteum with subsequent progesterone secretion. In males, LH stimulates testosterone synthesis in Leydig cells in the interstitial tissue of the testicles. FSH and LH tests can be used as part of carving out a differential diagnosis between peripheral and central hypogonadism. LH and FSH levels are increased in peripheral hypogonadism (known as hypergonadotropic hypogonadism), while the levels are normal or more often decreased in central hypogonadism. Determination of the LH and FSH also plays an important role in diagnosing polycystic ovary syndrome where the FSH level decreases or is normal, and the LH level increases, so the LH/FSH ratio increases in contrast to healthy individuals.

34.1.1.2. Analytical Phase of LH Testing

The most suitable material for LH testing is blood serum, so sampling sets with accelerated haemocoagulation should be used for blood sampling. The menstrual cycle phase is important in fertile women, so the day of the cycle should be recorded. The sample must be transported to the laboratory on the day of sampling, and the material processed within 8 hours of sampling. Tightly covered samples are stored at 7 ±5 °C no longer than 48 hours. Where longer storage is required, samples should be stored at -20 °C. Thawed samples should not be frozen again. The most common methods used for testing are RIA and CMIA.

34.1.1.3. Indications for LH Testing

Amenorrhoea, dysmenorrhoea, oligomenorrhoea, infertility, primary and secondary hypogonadism, conditions following cytostatic treatment or radiation, atypical sexual maturation in children, pituitary traumas and tumours, gonadal dysgenesis, selected chromosomal aberrations and hypothalamic-pituitary-gonadal axis disorders.

34.1.1.4. Reference Ranges

Gender	Age, (phase of the menstrual cycle in women)	Reference ranges
Both genders	0 - 1 year	0.02 - 8 IU/l
both gender	1 - 10 years	0.04 - 3.9 IU/l
Males	10 - 15 years	0.56 - 7.8 IU/l
Males	over 15 years	1.2 - 10 IU/l
Fertile women	Follicular phase of the cycle	2 - 13 IU/l
	Ovulation peak	15 - 80 IU/l
	Luteal phase of the cycle	1 - 19 IU/l
	Menopausal women	10 - 60 IU

34.1.1.5. Interpretation

34.1.1.6. Elevated levels

Primary ovarian insufficiency, Turner syndrome, climacterium praecox, gonadotropinomas, physiologically elevated values in the menopausal period, primary hypogonadism in males.

34.1.1.7. Reduced levels

Pituitary traumas and tumours, anorexia nervosa, secondary ovarian insufficiency, pseudopubertas praecox.

34.1.2. FSH Tests

34.1.2.1. FSH - Introduction

FSH, a follicle stimulating hormone or follitropin is a glycoprotein with a relative molecular weight of about 30 kDa, which is secreted by basophilic gonadotrophic cells of the anterior pituitary. The LH structure consists of two non-covalently bound subunits, α and β. The α-subunit is common to all glycoprotein hormones of the hypophysis (TSH, LH). The β-subunit is specific to the relevant hormone and defines its biological and immunological specificity. FSH production is controlled by a complex mechanism of positive and negative feedback. FSH is secreted in the pulse through the action of secretion of the LH-FSH releasing hormone from the hypothalamus. FSH production is inhibited by inhibin, a dimeric glycoprotein from the family of cytokines. FSH receptors are also common to LH. These are snake-like receptors coupled with adenylate cyclase by G-proteins. The function of FSH is to stimulate the growth and maturation of ovarian follicles; FSH together with LH support oestrogen secretion and participate in endometrium transformation in the proliferative phase of the menstrual cycle. When the ovarian follicles stop working during menopause, steroid production decreases and feedback FSH secretion inhibition is slowly reduced. This is why FSH levels during menopause are markedly higher than FSH levels in the fertile period. The stimulation of Sertoli cells by the FSH helps maintain male spermatogenic epithelium, and affects the synthesis of inhibin and binding protein for androgens (SHBG) in seminiferous tubules. FSH and LH tests can be used for differential diagnosis of hypogonadism, i.e. to distinguish between the peripheral and central causes. LH and FSH levels are increased in peripheral hypogonadism (known as hypergonadotropic hypogonadism), while the levels are normal or more often decreased in central hypogonadism. The LH and FSH assay is also important for diagnosing polycystic ovary syndrome where the FSH level decreases or is normal, and the LH level increases, so the LH/FSH ratio increases in contrast to healthy individuals.

34.1.2.2. Analytical Phase of FSH Testing

The most suitable material for FSH testing is blood serum, so sampling sets with accelerated haemocoagulation should be used for blood sampling. The menstrual cycle phase is important in fertile women, so the day of the cycle should be recorded. The sample must be transported to the laboratory on the day of sampling, and the material has to be processed within 8 hours of sampling. Tightly covered samples are stored at 7 ±5 °C no longer than 48 hours. Where longer storage is required, samples should be stored at -20 °C. Thawed samples should not be refrozen. Methods most commonly used for testing are RIA and CMIA.

34.1.2.3. Indications

Amenorrhoea, dysmenorrhoea, oligomenorrhoea, infertility, primary and secondary hypogonadism, conditions following cytostatic treatment or radiation, atypical sexual maturation in children, pituitary traumas and tumours, gonadal dysgenesis, selected chromosomal aberrations and hypothalamic-pituitary-gonadal axis disorders.

34.1.2.4. Reference Ranges

Gender	Age, (phase of the menstrual cycle in women)	Reference ranges
Both genders	0 - 1 year	0.02 - 8 IU/l
Both gender	1 - 10 years	0.04 - 3.9 IU/l
Males	10 - 15 years	0.56 - 7.8 IU/l
Males	over 15 years	1.2 - 10 IU/l
Fertile women	Follicular phase of the cycle	2 - 13 IU/l
	Ovulation peak	15 - 80 IU/l
	Luteal phase of the cycle	1 - 19 IU/l
	Menopausal women	10 - 60 IU/l

34.1.2.5. Interpretation

34.1.2.6. Elevated levels

Primary ovarian insufficiency, Turner syndrome, climacterium praecox, gonadotropinomas, physiologically elevated values in the menopausal period, primary hypogonadism in males.

34.1.2.7. Reduced levels

Pituitary traumas and tumours, anorexia nervosa, secondary ovarian insufficiency, pseudopubertas praecox.

34.1.3. Sex Hormone-Binding Globulin (SHBG) Tests

34.1.3.1. Introduction

SHBG or sex hormone-binding globulin is a protein with a molecular weight of 90 kDa, which consists of two subunits, is produced in the liver and has a biological half-life of about 7 days. SHBG can be found in the zone of β-globulins when tested by electrophoresis. The main function of SHBG is to transport steroid hormones to target cells within the entire body. The transport in the SHBG-hormone complex is applied primarily in the transportation of testosterone, oestradiol, oestriol, oestrone, dehydroepiandrosterone (DHEA) and 5-dihydrotestosterone (DHT). Due to the high affinity of hormones to SHBG, about 98% of circulating hormones are bound to this protein. As the affinity of hormones to SHBG is about 100 times higher than their affinity to albumin, hormones bound to albumin are more easily available to the target cells.

34.1.3.2. Analytical Phase

The most suitable material for SHBG testing is blood serum, so sampling sets with accelerated haemocoagulation should be used for blood sampling. Blood should be sampled on an empty stomach. The sample must be transported to the laboratory on the day of sampling, and the material processed within 8 hours of sampling. Tightly covered samples are stored at 7 ±5 °C no longer than 48 hours. Where longer storage is required, samples should be stored at -20 °C. Thawed samples should not be refrozen. Methods most commonly used for testing are RIA and ECLIA - electrochemiluminescence immunoassay.

34.1.3.3. Indications

Hyperandrogenism; the assay is also recommended for cases of high or low clinically mute testosterone levels. The SHBG assay is important for testosterone free index calculation (see 2.3.6 TFI Index).

34.1.3.4. Reference Ranges

Gender	Age	Reference ranges
Males	18 - 60 years	14.5 - 48.4 nmol/l
Females	18 - 55 years	26.1 - 110 nmol/l
Post - menopausal women (without hormone substitution)		14.1 - 68.9 nmol/l

34.1.3.5. Interpretation

34.1.3.6. Elevated levels

Hyperthyroidism, use of hormonal contraceptives and anti-epileptic drugs, physiologically during pregnancy and in elderly males, hepatic cirrhosis, hepatitis.

34.1.3.7. Reduced levels

Obese patients, hypothyroidism, elevated androgen levels, alopecia, polycystic ovaries, hirsutism.

34.1.3.8. TFI Index

Determination of SHBG is important for calculating the testosterone free index (TFI index), which provides better information as to the potential increase in the biologically active free testosterone fraction.

TFI = Testosterone (nmol/l) / SHBG (nmol/l)

34.1.4. Testosterone Tests

34.1.4.1. Introduction

Testosterone is one of the principal male sex hormones. This hormone has a molecular weight of 288.4 kDa. Testosterone synthesis starts by creating pregnenolone from cholesterol in a reaction catalyzed by 20α-hydroxylase. Testosterone is produced in males under the influence of the luteinizing hormone in Leydig cells of the testicles; the main testosterone producing organ in females are the ovaries, and smaller amounts are also produced in the adrenal cortex. However, on a daily basis females produce only about 5 - 10% of the amount produced by males. Circulating testosterone has three different forms. There is a small amount (1 - 2.5%) of free testosterone, a fraction bound by a weak bond to albumin, and the last fraction forms a strong bond to SHBG. The different strengths of the testosterone-protein bond also define the biological availability of testosterone. This is why the weak testosterone-albumin bond and the free testosterone fraction form bioavailable testosterone. Testosterone levels exhibit large age and gender-dependent differences during development. Testosterone levels in male newborns are slightly higher than levels found in female newborns after childbirth. For the next three months, testosterone levels rise and then decrease in boys until the end

of the first year, although the levels are slightly higher compared with girls. At the age of 6 - 7, the onset of puberty is connected with nocturnal pulses of LH, which lead to an increase in testosterone levels. Increased testosterone level in this period is important for proper development of secondary sex characteristics and bone density. At the age of 16 - 19, puberty is complete; testosterone levels become stabilized and remain about the same throughout adult age. At about 50 years of age, testosterone slowly decreases in males, but there are great interindividual differences. This decrease is responsible for lower sexual activity and lower physical performance. This condition is referred to as the andropause, a condition equivalent to the female menopause. Like most other sex hormones, testosterone exhibits a diurnal rhythm, with morning levels about 25 - 30% lower than evening levels. Male testosterone is responsible for the development of secondary sex characteristics, voice, affects bone density and muscle mass, and plays a unique role in sexual functions and libido. Female testosterone affects the growth of pubic and axillary hair, and affects libido like male testosterone. Pathologically decreased testosterone level is responsible for the condition called hypogonadism, which comprises 3 large subgroups: Hypogonadotropic hypogonadism, hypergonadotropic hypogonadism and “testicular feminization syndrome”. Hypogonadotropic (secondary) hypogonadism is due to decreased function of intact gonads, most often caused by congenital or acquired panhypopituitarism, GnRH deficiency, hyperprolactinaemia, malnutrition or anorexia nervosa. Decreased LH and FSH levels are typical. Hypergonadotropic (primary) hypogonadism is caused by gonad dysfunction, and patients have higher LH and FSH levels. Causes of hypergonadotropic hypogonadism include gonad injury (radiation, diseases, developmental defects, tumours, drugs), poor gonad development in patients with chromosomal aberrations, and enzyme defects of androgen synthesis. Testicular feminization syndrome is caused by an androgen receptor defect; afflicted individuals are feminine in appearance, whose genitals consist of a blind-pouched vagina and present testicles. Testosterone levels in these patients are normal or slightly heightened. High testosterone levels in females cause virilization characterized by typical enlargement of the clitoris and excess body hair connected with increased proliferation of and secretion from sebaceous glands. Acne, facial hair growth and deepening of the voice are also often present.

34.1.4.2. Analytical Phase

The most suitable material for testing testosterone is blood serum, so clean sampling sets or sampling sets with accelerated haemocoagulation should be used for blood sampling. The sample has to be transported to the laboratory on the day of sampling, and the material has to be processed within 8 hours of sampling. Samples can be stored at 2 - 8 °C for 2 days, or frozen to -20 °C for longer periods of storage (1 month). Thawed samples may not be refrozen. The most commonly used methods for testing are RIA and CMIA.

34.1.4.3. Indications

Males: Hypogonadism, infertility, diff. dg. of pubertas paecox in boys, replacement therapy monitoring.

Females: Androgenic conditions, hirsutism, menstrual disorders, acne.

34.1.4.4. Reference Ranges

Gender	Age	Reference ranges
Males	15 - 60 years	10 - 30 nmol/l l
Males	over 60 years	0 - 10 nmol/l
Females	over 15 years	0.15 - 2.6 nmol/l

34.1.4.5. Interpretation

34.1.4.6. Elevated levels

Short-term intensive physical strain, tumours of the testes and ovaries, adrenal tumours, use of anabolic agents.

34.1.4.7. Reduced levels

Enzyme defects of testosterone synthesis, primary and secondary hypogonadism, long-term physical strain, drug abuse, liver disease, drugs such as ketoconazole, Klinefelter syndrome.

34.1.5. Progesterone Tests

34.1.5.1. Introduction

Progesterone is a C_{21} steroid hormone (molecular weight: 314.5 kDa) produced by the corpus luteum, placenta and small amounts also by ovarian follicles and the adrenal cortex. It is an important intermediary product in steroid hormone biosynthesis in all steroidogenic tissues, and small amounts are transferred to the blood in the testicles and in the adrenal cortex. Circulating progesterone occurs in a free form (about 2% of the total amount) as well as in a form bound to albumin (18% of total progesterone) or SHBG (80% of total progesterone). Progesterone has a relatively short half-life. The first step of inactivation is chemical reduction, in which the following substances may be formed depending on the degree of reduction: 1) pregnanediones, 2) pregnanolones and 3) pregnanediols. These reduced compounds subsequently conjugate with glucuronic acid and are excreted in urine as water-soluble glucuronides. Progesterone secretion is primarily regulated by the luteinizing hormone. Progesterone levels are low in the second half of the follicular phase of the menstrual cycle, however levels dramatically increase during the luteal phase, i.e. 6 - 8 days after ovulation, i.e. on the 20th to 22nd day of the menstrual cycle. Progesterone reaches its peak values from 30 to 65 nmol/l. This rise in progesterone is caused by secretion from the corpus luteum. If fertilization does not occur, the levels subsequently decrease. In the case of fertilization, progesterone secretion from the corpus luteum is maintained by human choriogonadotropin. As the pregnancy progresses, progesterone is also produced in the placenta. The most important target organs for progesterone are the uterus, breasts and the brain. Progesterone in the uterus stimulates vascularization of the endometrium; endometrial glands become coiled and produce serous secretion. As soon as corpus luteum regression occurs, hormonal stimulation of the endometrium declines and the endometrium turns into the menstrual phase. Progesterone has an anti-oestrogenic effect on myometrium cells, reduces their irritability and sensitivity to oxytocin, which is why it is essential for maintaining pregnancy. Progesterone in the breasts stimulates the development of lobules and alveoli. Progesterone supports differentiation of duct tissue ready for oestrogens and supports the secretory function of lactating breasts. Due to its thermogenetic effect, progesterone also causes an increase to the basal temperature.

34.1.5.2. Analytical Phase

The most suitable material for progesterone testing is blood serum, so sampling sets with accelerated haemocoagulation should be used for blood sampling. The menstrual cycle phase is an important factor in fertile women, meaning the day of the cycle should be recorded. The sample must be transported to the laboratory on the day of sampling, and the material processed within 8 hours of sampling. Tightly covered samples are stored at 7 ±5 °C no longer than 48 hours. Where longer storage is required, samples should be stored at -20 °C. Thawed samples should not be refrozen. The most commonly used methods for testing are RIA and CMIA.

34.1.5.3. Indications

Infertility, evaluation of the corpus luteum function, ovulation cycle demonstration.

34.1.5.4. Reference Ranges

Gender	Age	Reference ranges
Both genders	0 - 10 years	0.2 - 1.7 nmol/l
Males	10 - 15 years	0.3 - 4.3 nmol/l
Males	over 15 years	0.89 - 3.88 nmol/l
Females	10 - 15 years	0.3 - 41 nmol/l
Females	over 15 years	0.48 - 89.14 nmol/l

34.1.5.5. Interpretation

Progesterone levels reach their maximum 6 to 8 days after ovulation. This rise is subject to the secretion of progesterone from the corpus luteum, i.e. proving that ovulation occurred. If progesterone levels are under 25 nmol/l on these days, it is a sign that most likely there was no ovulation. Lower progesterone levels produced by the corpus luteum may also be a sign of luteal phase deficiency (LPD) associated with infertility and early spontaneous abortions. Women on oral contraceptives have decreased progesterone levels.

34.1.6. Oestradiol Tests

34.1.6.1. Introduction

Oestradiol is one of the most biologically active substances from the group of female steroid hormones responsible for developing and maintaining female sex organs and secondary sex characteristics. In conjunction with progesterone, oestradiol plays a role in the menstrual cycle, breast and uterus development, and is involved in maintaining pregnancy, when oestradiol concentration steadily grows. Oestrogens also play a role in calcium homeostasis and support good bone density. A long-term decline in oestrogens therefore leads to bone density reduction with consequent osteoporosis. Oestrogens also have a protective effect on lowering cardiovascular risk, because they increase the catabolism of circulating LDLs, probably by increasing the number of LDL receptors in the liver. Healthy women produce the most part of oestradiol in the ovarian follicle and corpus luteum, and in the placenta during pregnancy. Small amounts of oestradiol are also produced in the adrenal cortex and male testicles. Unlike the testicles, ovaries have a highly active aromatase system, which quickly converts androgens such as testosterone into oestrogens. More than 20 oestrogens are known today, however laboratory tests may be clinically used for 17β-oestradiol and oestriol only. Plasma oestradiol concentration has two peaks of secretion: one immediately before ovulation and the second in the middle of the luteal phase. The rate of oestradiol secretion is 36 µg/day (133 nmol/day) during the early follicular phase, 380 µg/day immediately prior to ovulation and 250 µg/day in the middle of the luteal phase. It is the high levels prior to ovulation that affect LH secretion at the pituitary level, leading to massive secretion of the LH which triggers ovulation. Oestrogen secretion drops down to low levels after the menopause. At least 97% of oestradiol circulating in the blood plasma is bound to plasma proteins such as SHBG and albumin. The main site of oestrogen inactivation is the liver, where oestrogens are subject to hydroxylation, oxidation, reduction and methylation to form oestrone, which further conjugates with glucuronic acid to form water-soluble glucuronides excreted in urine. The function of oestrogens consists of endometrium proliferation stimulation. The uterine mucus gets thinner and more alkaline due to oestrogens, which are changes helping the sperm survive and move, enlarge uterine muscle mass and increase the content of contractile proteins. Oestrogens make the uterine muscle more active and sensitive; action potentials propagate in individual fibres, increasing uterine sensitivity to oxytocin.

34.1.6.2. Analytical Phase

A test tube with accelerated haemocoagulation or lithium heparin. The sample must be transported to the laboratory on the day of sampling, and the material processed within 8 hours of sampling. If samples are not processed within 8 hours of sampling, they should be stored tightly covered at 7 ± 5 °C no longer than 48 hours. Where longer storage is required, samples should be stored at -20 ± 5 °C. Thawed samples must not be refrozen. The most commonly used methods for testing are RIA and CMIA.

34.1.6.3. Indications

Males: Differential diagnosis of gynaecomastia; can also be used as a tumour marker in germinal tumours of testicles, choriocarcinoma, hepatoma and oestradiol-producing adrenal tumours.

Females: Monitoring of ovulation induction and ovarian hyperstimulation, diff. dg. of ovarian insufficiency, menstrual cycle disorders; can also be used as a tumour marker in germinal tumours of testicles, choriocarcinoma, hepatoma and oestradiol-producing adrenal tumours.

34.1.7. Reference Ranges

Gender	Age, (phase of the menstrual cycle in women)	Reference ranges
Males	All age groups	29 - 94 pmol/l
Females	Follicular phase of the cycle	46 - 607 pmol/l
	Ovulation phase	315 - 1,828 pmol/l
	Luteal phase	161 - 774 pmol/l
	Post-menopausal period	<201 pmol/l

34.1.7.1. Interpretation

34.1.7.2. Elevated levels

Oestrogen-producing tumours, slight increase in hepatic cirrhosis and obesity.

34.1.7.3. Reduced levels

Primary ovarian insufficiency, anovulatory cycles, corpus luteum insufficiency.

34.1.8. Dehydroepiandrosterone Sulphate - DHEAS

34.1.8.1. Introduction

Dehydroepiandrosterone sulphate (DHEA-SO_4, DHEA-S) is a steroid hormone secreted physiologically exclusively in the adrenal gland, only a small part comes from male testicles. Given the sole production of this steroid hormone in the adrenal gland, increased DHEAS levels are a major indicator of adrenal pathology. DHEAS only has slight androgenic effects, however may metabolize to stronger androgens such as androstendione and testosterone, thus indirectly causing hyperandrogenism in hirsutism or virilization. DHEAS concentrations start to increase at the beginning of puberty, at about 6 - 7 years of age, when the entire complex of changes is initiated by the presence of nocturnal pulses of LH. DHEAS production becomes reduced from the age of about 30. Increased DHEAS levels in women with signs of virilization or hirsutism allow us to distinguish between adrenal and ovarian causes of hirsutism or virilization. Higher DHEAS levels are caused by adrenal system pathology, e.g. adrenal tumours, bilateral adrenocortical hyperplasia or congenital adrenal hyperplasia connected with 21-hydroxylase, 11β-hydroxylase or 3β-HSD deficiency.

34.1.8.2. Analytical Phase

A test tube with accelerated haemocoagulation. The sample has to be transported to the laboratory on the day of sampling, and centrifuged after delivery to the laboratory. The sample has to subsequently be processed within 8 hours of sampling. If samples are not processed within 8 hours of sampling, they should be stored tightly covered at 7 ± 5 °C no longer than 48 hours. Where longer storage is required, samples should be stored at -20 ± 5 °C no longer than 2 months. Thawed samples must not be refrozen.

The most commonly used methods for testing are RIA and ECLIA.

34.1.8.3. Indications

Differential diagnosis of hirsutism and virilization, suspected adrenal tumour, congenital adrenal hyperplasia.

34.1.8.4. Reference Ranges (in µmol/l)

Gender	Age, (phase of the menstrual cycle in women)	Reference ranges
Both genders	Day 0 - 7	2.93 - 16.50
Both gender	Day 7 - 28	0.86 - 11.70
Both genders	Day 28 - 365	0.09 - 3.35
Both genders	1 - 5 years	0.01 - 0.53
Both genders	5 - 10 years	0.08 - 2.31
Males	10 - 15 years	0.66 - 6.70
Males	15 - 20 years	1.91 - 13.40
Males	20 - 25 years	5.73 - 13.40
Males	25 - 35 years	4.34 - 12.20
Males	35 - 45 years	2.41 - 11.60
Males	45 - 55 years	1.20 - 8.98
Males	55 - 65 years	1.40 - 8.01
Males	65 - 75 years	0.91 - 6.76
Males	75 - 150 years	0.44 - 3.34
Females	10 - 15 years	0.92 - 7.60
Females	15 - 20 years	1.77 - 9.99
Females	20 - 25 years	4.02 - 11.00
Females	25 - 35 years	2.68 - 9.23
Females	35 - 45 years	1.65 - 9.15
Females	45 - 55 years	0.96 - 6.95
Females	55 - 65 years	0.51 - 5.56
Females	65 - 75 years	0.26 - 6.68
Females	75 - 150 years	0.33 - 4.18

34.1.8.5. Interpretation

34.1.8.6. Elevated levels

Adrenal tumours, bilateral adrenocortical hyperplasia or congenital adrenal hyperplasia, polycystic ovary syndrome.

34.1.8.7. Reduced levels

Adrenal cortex insufficiency.

34.1.9. AMH Testing

34.1.9.1. Introduction

Anti-Müllerian hormone is a dimeric glycoprotein from the cytokine class XVI, which includes other inhibin and activin glycoproteins that play an important role in the growth and differentiation of tissues. The AMH molecule contains 2 monomeric units linked by disulphide bridges, forming a unit with a molecular weight of 72 kDa. The gene coding human AMH is located on the short arm of chromosome 19. The AMH is produced in male Sertoli cells of the testicles, and females produce the AMH only in the postnatal period in ovarian granulosa cells. As the name of the AMH makes clear, the AMH regulates sexual differentiation of a male foetus. Specifically, the role of this hormone is to induce and maintain physiological regression of Müllerian ducts, which is completed in about the 9th week of intrauterine development. It is the absence of the AMH, which is essential in the development of female sex organs in the female embryo. The AMH has no function in adult males, so the AMH level remains relatively low. However, the situation is different in adult females. It has been proved that the AMH inhibits primordial follicle formation and inhibits the sensitivity of growing follicles to the FSH at the same time. It has been proven that the AMH level is therefore directly proportional

to the number of antral follicles and is a very good indicator of ovarian reserve.

34.1.9.2. Indications

Differential diagnosis of intersexual conditions, tumours of Sertoli cells of the testes, gonadal dysgenesis, tumours of ovarian granulosa cells, polycystic ovary syndrome, disorders of metabolism and testosterone production, diff. dg. and therapy of infertility, tests for IVF.

34.1.9.3. Analytical Phase

The AMH can be tested in human serum or plasma. Blood is therefore collected into test tubes without additives or test tubes with a gel separator, lithium heparin or EDTA. The sample must be transported to the laboratory on the day of sampling, and centrifuged after delivery to the laboratory. The serum or plasma samples can then be stored at 2 - 8 °C for 24 hours. Where longer storage is required, samples should be frozen at < -18°C. The test method is enzyme immunoassay (EIA), for example.

34.1.9.4. Reference Ranges

Sample – Serum	Median Age	Reference Interval (µg/L)
Boys	4.8	3.8 - 160
Girls	5.0	ND - 8.9
Males	38	1.3 - 14.8
Females	30	ND - 12.6
Post-menopausal females	71	ND

Note: ND - Not Detected; the values were taken over from the instructions for use of the Beckman Coulter AMH gen II ELISA assay.

34.1.9.5. Interpretation

34.1.9.6. Elevated levels

Tumours of Sertoli cells of the testes, tumours of ovarian granulosa cells, polycystic ovary syndrome, testicular dysgenesis, testotoxicosis in boys, some biosynthesis disorders and androgen insensitivity syndromes.

34.1.9.7. Reduced levels

Cryptorchidism, anorchia, testicular dysgenesis, precocious menopause.

34.1.10. Inhibin Tests

34.1.10.1. Introduction

Inhibins belong to the cytokine class XVI. These glycoproteins have a unique function in human reproductive physiology. The basis of the molecule of both inhibin types is a glycosylated α-subunit with a molecular weight of 18,000, and a non-glycosylated subunit βA for inhibin A and βB for inhibin B. Both β-subunits are coupled with the α-subunit by means of disulphide bridges. Inhibins can also be found in the complex with binding proteins follistatin or α2-macroglobulin. The activin-follistatin bond affects its biological and immunochemical activity. Inhibins are produced in female granulosa cells within the follicle and in male Sertoli cells of the testes.

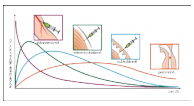
34.1.10.2. Inhibin A

The inhibin A molecule is a heterodimer consisting of two subunits, α a β_A. This inhibin is produced primarily in ovarian granulosa cells, in the placenta during pregnancy, and small amounts in male Sertoli cells of the testes. The major function of inhibin A is feedback FSH secretion inhibition, although inhibin B is the major inhibitor. Inhibin A is used in clinical practice for trisomy 21 diagnosis. If included in the prenatal screening with AFPm hCG and non-conjugated E3,

Down syndrome detection rate increases by up to 20%. Other potentially useful indications for inhibin A investigations are include diagnosis for pre-eclampsia, premature birth and the differential diagnosis of infertility. Inhibin A may also be used as a marker of embryo transfer in IVF.

34.1.10.3. Inhibin B

The inhibin B molecule structure is the same as that of inhibin A, it only contains β_B-subunit instead of β_A-subunit. The main site of hormone production of this type in males is Sertoli cells of the testes, but recent research has shown that Leydig cells are also capable of producing inhibin subunits, especially after the end of sexual maturation. This fact is also confirmed by studies in which inhibin B levels increased following administration of recombinant FSH to hypogonadotropic males. The main role of this hormone is to regulate gametogenesis influenced by feedback mechanism by means of the FSH. Inhibin B is produced in females primarily in small antral follicles, which corresponds to the fact that inhibin B is a sensitive marker of the current number of follicles of the ovarian reserve. This is why inhibin B investigations are very important in the differential diagnosis of infertility and is a good marker of precocious menopause at the same time. Inhibin B investigations are clinically important in males for the recognition of bilateral cryptorchidism and anorchia, in the diagnosis of Sertoli cell tumours and the differential diagnosis of infertility and pubertas praecox.



CHAPTER 35

35. Therapeutic Drug Monitoring

Authors: doc. RNDr. Kristian Šafarčík, Ph.D.

Reviewer: prof. MUDr. Jaroslav Racek, DrSc.

35.1. Introduction

Drugs act at different levels: at the level of the entire organism, at the level of organs and tissues and at the level of cells and molecules. The determining factor of pharmacological action is drug concentration at the site of action. The aim is to create an objective basis for rational drug dosing and also for individual treatment, taking the individual differences of patients into account. Making an effort to adjust doses to individual needs is very important in ensuring effective treatment and reducing the rate of adverse drug effects.

Most drugs show a characteristic relation between the dose and the effect. A more detailed analysis has shown that the intensity of the pharmacological action of drugs better correlates with the drug concentration in system circulation than the drug dose. These basic premises are essential for Therapeutic Drug Monitoring (TDM), aimed at the clinical use of the measured drug concentration in the body with the objective of adjusting the dose individually. Individual dose adjustment relies on reaching the specific concentration of the target drug. This reduces the main source of drug concentration variability caused by individual differences in pharmacokinetics at the level of each patient and in the relevant group of patients.

The aim is to understand more precisely the pharmacokinetics and pharmacodynamics of a drug in individual patients (Figure 38.1). Drug concentrations do not have to be measured in situations where easy clinical monitoring of the drug effect is possible (antihypertensive drugs, analgetics, hypnotic drugs) or simple laboratory tests are available (anticoagulants, hormones, hypolipidemic agents).

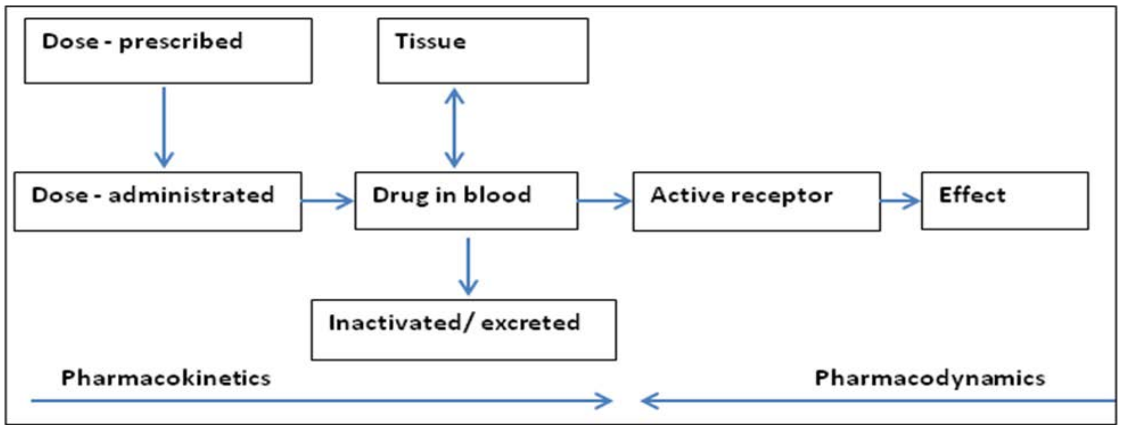


Figure 35.1. Processes after drug administration

35.2. Indications for Drug Level Determination

The following properties are drug characteristics suitable for monitoring:

- **Narrow therapeutic width**
- **Risk of toxicity**
- **High interindividual variability**
- **Steep dependence between the dose and the effect**

- **Concentration-effect relationship is closer than the dose-effect relationship**
- **Non-linear pharmacokinetics**

TDM is beneficial for the patient in the following cases:

- **Elimination of patient non-compliance**
- **Changes during physiological conditions (pregnancy, childhood, old age)**
- **Changes during pathological conditions (fever, renal, hepatic or cardiac failure)**
- **Drug interactions**

TDM is currently used for the following groups of drugs:

- **Antibiotics** – aminoglycosides (gentamicin, amikacin), vancomycin
- **Bronchodilatants** – theophylline
- **Antiepileptics** – phenobarbital, primidone, phenytoin, ethosuximide, carbamazepine, valproic acid, clonazepam, lamotrigine, topiramate, levetiracetam
- **Cytostatics** – methotrexate, busulfan
- **Cardiacs** – amiodarone, digoxin
- **Immunosuppressive drugs** – cyclosporin A, tacrolimus, sirolimus, everolimus, mycophenolate
- **Psychoactive drugs** – lithium, diazepam, antidepressants, antipsychotics

In certain cases it is advisable to determine the basic metabolite of the parent active substance:

- ***Carbamazepine*** – 10,11 epoxycarbamazepine
- ***Primidone*** – phenobarbital
- ***Amiodarone*** – desethylamiodarone
- ***Diazepam*** – desmethyldiazepam
- ***Metoprolol*** – hydroxy metoprolol

TDM has been performed on a routine basis since the mid-1970s as part of the concept of clinical pharmacology as an independent branch together with the fundamental principles of pharmacokinetics and pharmacodynamics. In 1967, Baastrup and Schou demonstrated the relationship between the concentration and pharmacological effect of lithium, and Buchtal described the relationship between the plasma concentration of phenytoin and frequency of seizures in patients being treated for epilepsy in 1960.

35.2.1. TDM in Practice

TDM in practice comprises three basic parts: **The analytical determination of drug concentration**, usually by immunoassays or chromatographic methods. Immunoassays are more commonly used, and include homogeneous methods (no need to separate the bound indicator) and nonhomogeneous, mostly competitive methods of determination. There are many benefits to immunoassays including simple handling of the sample without the need to extract the sample. Regular instruments found in a biochemical laboratory can be used. Homogeneous immunoassays can be performed on biochemical analyzers with photometric detection and heterogeneous immunoassays on immunoassay analyzers from different manufacturers. Drug concentrations can be determined in the random access mode, also for a whole series of samples for pharmacokinetics profiles. The drawback of immunoassays is certain overestimation of the drug concentration due to the cross reaction of metabolites, which invariably occur when the administered drug is metabolized and are present in the systemic circulation. The structure of metabolites is often very similar to the parent substance, so they can bind to the detection antibody used in the immunoassay, which leads to overestimation of the concentration found. However, this does not matter in many cases.

Chromatographic methods primarily use gas chromatography (GC) and high-performance liquid chromatography (HPLC) with different types of detection, recently with the detector based on mass spectrometry, which makes it possible to detect not only the administered parent substance but sometimes also major metabolites of the administered drug.

The next step is **drug concentration interpretation**, sometimes with pharmacokinetic analysis. The laboratory result is interpreted by the clinical pharmacologist (clinical biochemist); the numerical result is supplemented by recommendations to the clinician concerning dose or dosing interval (schedule) adjustment.

Feedback between the clinician and the result interpretation is essential, since acceptance of the recommendations and further follow-up based on this recommendation ensures a change of the dosing and rational pharmacotherapy. Failure to take heed of the recommendation and merely inserting the result to the medical record only brings about unnecessary increases to the overall cost of treatment.

35.2.2. Information Required for TDM

Samples of blood for testing drugs which create stable levels (drugs with a long elimination half-life) should be taken before administering the next dose. Cyclosporin A is such drug. Testing digoxin level is recommended 6-11 hours after administration, at which point blood digoxin concentration better reflects the levels in the myocardium. In any case, it is necessary to wait until equilibrium has been reached. Equilibrium has been reached provided the time equal to 4-5 elimination half-lives has passed since the beginning of administration.

The concentration of drugs, the level of which changes quickly (drugs with a short elimination half-life), is determined:

- Before administration – basal value (trough value) - t_0
- After administration when the maximum concentration is reached (peak value) - t_{max}

The route of administering the drug is also important – maximum concentration in the blood following i.v. application is reached within 30 minutes, following i.m. application in an hour, following oral administration even after a longer time depending on the pharmaceutical form of the drug administered (tablets, capsules, matrix tablets, film-coated tablets with extended release, etc.). The entire dose is absorbed after i.v. application, while different portions are absorbed after oral administration, and depend on the type of the drug, galenical form of the drug and condition of the GIT.

Correct interpretation in terms of the concentration of the drug thus requires filling in a lot of information on the request for laboratory analysis:

Information about the drug type and form

- Dose history, current dose, compliance, length of therapy
- Patient's condition (age, weight, height, organ functions)
- Time of administering the dose and time of blood sampling
- Type of biological material (blood, plasma, serum)
- Medication history (concurrently administered drugs)

35.2.3. Pharmacokinetic Analysis

Pharmacokinetics is the study of the fate of drugs in the body including drug absorption, distribution, biotransformation (metabolism) and elimination. These processes can be described by mathematical methods and a pharmacokinetic analysis can be used to determine the pharmacokinetic parameters. Pharmacokinetic information acquired when registering the drug registration and further information gained at subsequent stages of clinical testing form the basis to understand the parameters of the population pharmacokinetic curve for the relevant drug and its pharmaceutical form. It is possible to compare this information to even just one measured drug concentration for the examined patient. Together with the drug type, the measured drug concentration, dose, time of administration and time of biological sample collection, patient's weight and age and other data are entered onto a special computer program, which evaluates whether the measured concentration complies with the average population pharmacokinetics.

In the event unexpected concentrations are found, it is necessary to consider whether the wrong dose has been taken, drugs have interacted with one another, the drug has been used inappropriately, whether absorption disorders are present, collection took place at the wrong time, kidney and liver disease are present, there has been a change in the strength of the bond to serum proteins or the possibility of patient non-compliance (use of a higher or lower dose).

35.2.4. Target Drug Concentration Range

If the measured values do not range within certain therapeutic limits and there is no doubt that all the entered data is correct, the program allows for a change of dose to be simulated or for the dosing interval to comply with requirements for optimum method of treatment.

The drug concentration should be within the target range. The concept of the target drug concentration range (replaces the older term “therapeutic range”) is based on two concepts – the minimum effective concentration and maximum safe concentration for the relevant medicine. Values between these concentrations represent the maximum therapeutic effect and at the same time a minimum risk of toxicity and side effects for most patients. Regrettably, this simple theory does not have to be fully valid in all aspects and the target range has to be always connected with, not replaced by, a clinical assessment of the patient’s condition. The objective is not to reach the “target range” but to treat patient’s health problems.

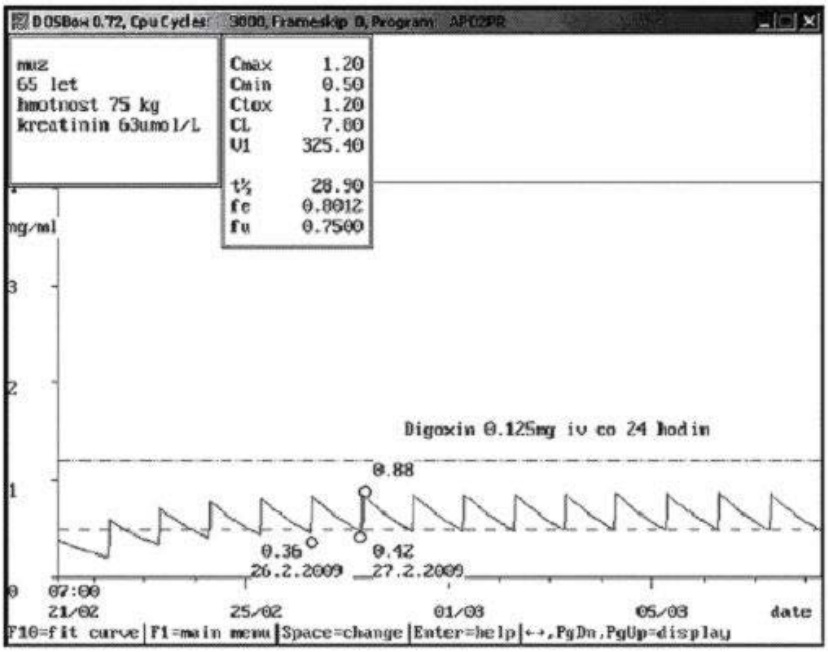


Figure 35.2. An intravenous dose of 0.125 mg of digoxin was administered to a patient (65 years) over a period of 24 hours. The digoxin level prior to administration was low – 0.36 ng/ml. Digoxin levels were normal 1 hour after administration. Prepared by the MW-PHARM program for pharmacokinetic analysis.

35.3. ADTM (Advanced Therapeutic Drug Monitoring)

Further development in this area leads to ADTM, which introduces the determination of free (pharmacologically effective) drug fraction, metabolites of the drug, determination of drug concentration in target tissues (e.g. immunosuppressive drugs in lymphocytes), and each patient is tested while taking into account genetic and phenotype predispositions for the metabolism of each drug group.

35.3.1. Biotransformation of Drugs

The study of administered drug movement within the body also includes the monitoring of processes which change the structure of the drug. The number of medicines in use is big yet their biotransformation in the body is based on a relatively small number of reactions. The aim is to create more hydrophilic, more polar metabolites, excreted from the body more easily than the original lipophilic substance. It is important to have knowledge regarding enzymes producing biotransformation activity localized primarily in the endoplasmic reticulum and cytosol of the cell. A greater part of biotransformation oxidation reactions is catalyzed by the cytochrome P450 monooxygenase system. Although more than 50 different cytochrome P450 proteins have been described in humans, drugs are metabolized primarily using five to six major enzymatic pathways. The main forms of cytochrome in the human liver are CYP2D6, CYP1A2, CYP2A6, CYP2C9, CYP2E1 and CYP3A4.

The capability of the body to metabolize and eliminate the drug is often affected by the genetic predisposition of the body. About 40% of drugs are eliminated using a genetically polymorphic pathway, of which cytochrome P450 2D6 (CYP2D6) is the most important. Metabolic pathway polymorphism means that the population includes individuals in which the relevant enzymatic activity is inactive (poor metabolizers – PMs), individuals with incomplete enzyme activity deficiency (intermediate metabolizers – IMs) and a group of extensive metabolizers (EMs) with normal metabolic transformation activity. For enzyme 2D6 cytochrome P450 there is also a group of ultrarapid metabolizers (UMs), i.e. people with unusually high CYP2D6 metabolic pathway activity.

If usual drug doses are administered, poor metabolizers accumulate the substances. High drug concentrations in target tissues may manifest themselves as toxic and adverse effects of drugs. Conversely, ultrarapid metabolizers do not reach equilibrium and drug concentrations in target tissues may be lower than minimal effective concentrations. This is why these people are resistant to common therapeutic dosages.

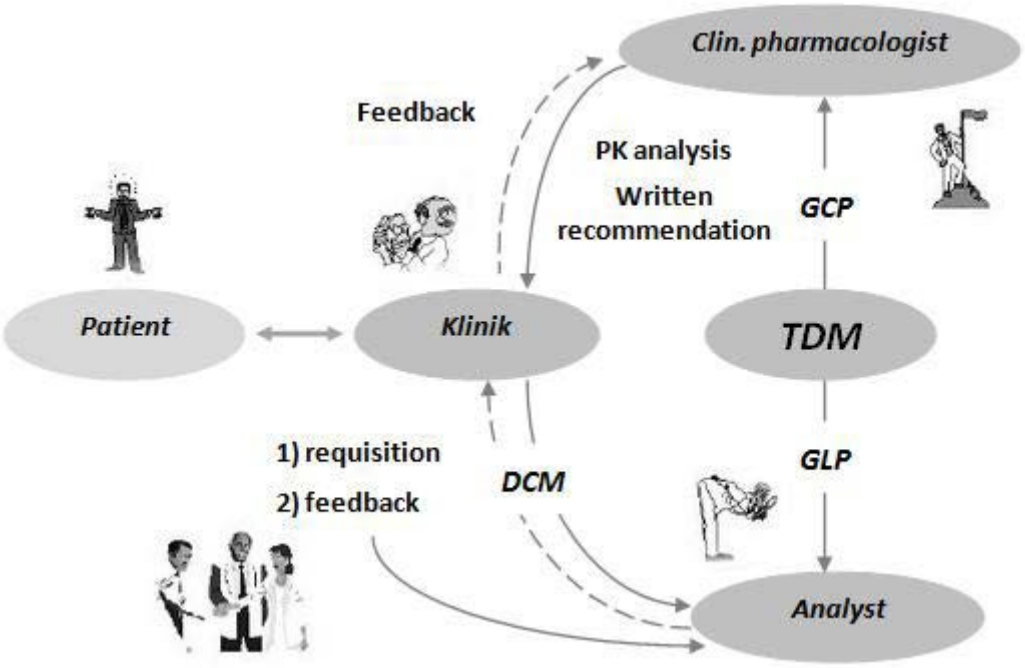


Figure 35.3. Diagram of correct therapeutic drug monitoring

35.4. Pharmacogenetics – Polymorphism – Genotype – Phenotype

Some promising methods focus on discovering genetic polymorphism in drug metabolism based on gene mutations. Phenotype and genotype are determined *in vivo*. The genotype can be determined by DNA analysis using the PCR (polymerase chain reaction). The analysis is based on restriction fragment length polymorphism (RFLP). DNA is cleaved to the fragments by endonucleases at the site of the sequence of bases specific to each enzyme.

Pharmacogenetics not only attempts to describe these determinants but also use this knowledge to personalise therapy and predict metabolic transformation and the patient's response to the specific drug. Present knowledge implies that many other external factors are decisive for current metabolic activity (**phenotype**) in addition to the genotype of cytochrome P450 forms. These factors involve processes of metabolic induction and inhibition of metabolizing enzymes triggered by other concurrently administered drugs or food components. An important role in internal factors is played by diseased conditions such as hepatic cirrhosis, which lead to the reduced synthesis of many enzymes. The resulting metabolic activity of each enzyme – drug metabolism phenotype – is tested using model substrates (Table 38.1).

1A2	**2C19**	**2C9**	2D6	3A	CYP450
korein	omeprazol	losartan	dextromethor-phan	midazolam	
teofylin	metfenytoin	tolbutamid	debrisochin	simvastatin	Model substances
		diklofenak	metoprolol	chinin	

Table 35.1. Examples of specific substrates (model substances) for determining metabolic activity

The principle of phenotype determination comprises administering a substance, which is a highly specific substrate for the tested enzyme. The rate of model substance biotransformation has to depend primarily on the enzyme activity studied, not on the flow rate of blood through the liver. The parent substance/metabolite ratio is decisive for putting the tested subject to the relevant group referred to as "poor metabolizers" or "extensive metabolizers".

For more details about pharmacogenomics and genetic polymorphism of selected enzymes, see Chapter 39, Trends in Laboratory Medicine, Section 4.6. Pharmacogenomics.

35.5. Personalized Pharmacotherapy

Clinical pharmacology utilizes results of phenotyping and genotyping to detect interindividual differences in drug metabolism, which contributes to treatment optimization and a lower incidence of adverse effects. Phenotyping is more important than genotyping to determine the current activity of each cytochrome for individualized drug dosing; genotyping is important for qualitative rather than quantitative use.

It is to say that therapeutic drug level monitoring is an indispensable, yet not fully clinically appreciated method for personalized medicine in the area of personalising drug dosage. TDM supplemented by ADTM, phenotyping and/or genotyping reaches an even a higher level.

Drug	Cammon target range: minimum effective concentration $C_{ss,min}$ or maximum safe concentration $C_{ss,max}$ (in mg/l)		Beginning of steady state
Aminoglycoside antibiotics			
	$C_{ss,min}$	$C_{ss,max}$	
Amikacin	3-5	20-30	8 hours
Gentamicin	1-2	5-10	8 hours
Netilmicin	1-2	5-10	8 hours
Tobramycin	1-2	5-10	8 hours
Antiarytmics			
Amiodaron	1,0-2,5		> 1 month
Disopyramid	2-5		24 hours
Chinidin	1-5		24 hours
Lidokain	1,5-5		12 hours
Mexiletin	0,5-1,9		2 days
Prokainamid	3,6-10		16 hours
Sotalol	1-2,5		2 days
Antiepileptics			
Ethosuximid	50-100		8 days
Fenobarbital	15-40		2 weeks
Fenytoin	10-20		> 2 weeks
Karbamazepin	5-12		2 weeks
Klonazepam	0,025-0,075		5 days
Valproic acid	50-100		40 hours
Primidon	5-15		2 days
Psychopharmaceuticals			
Amitriptylin	0,1-0,25		3 days
Haloperidol	5,2-15		3 days
Imipramin	0,12-0,3		2 days
Lithium	5,5-7		3 days
Nortriptylin	0,05-0,15		5 days
Other			
Cyclosporin A	0,08-0,25		3 days
Digoxin	1-2 $\mu g\ l^{-1}$		7 days
Salicylates	150-300		2-5 days
Theofylin	10-20		36 hours

Table 35.2. Common target therapeutic drug ranges

CHAPTER 36

36. Trends in Laboratory Medicine (POCT, Automation, Consolidation, EBM, Omics)

Authors: Mgr. Alena Krnáčová, Mgr. Radka Šigutová; Mgr. Tomáš Karlík; Ing. Vladimír Bartoš, Ph.D.; RNDr. Pavlína Kušnierová, Ph.D.,

Reviewer: prof. MUDr. Jaroslav Racek, DrSc.

36.1. POCT (Point-Of-Care Testing)

36.1.1. Definition of POCT

POCT (Point-Of-Care-Testing, Near-Patient Testing, Bedside Testing…) is defined as laboratory testing performed at or near the site of patient care, with the result leading to a potential change in patient care. The test is performed outside the laboratory, by personnel without primary laboratory education. POCT also includes patient self-monitoring at home.

POCT offers an improvement of the diagnostic process by integrating measurement, monitoring and physiological observation at the site of patient care where required and necessary.

In the 1950s, the POCT system saw considerable growth in terms of technological development and widespread use. As it turned out, the use of POCT is very convenient wherever no laboratory is available to provide results sufficiently fast given the importance of treatment. In conjunction with this development, a systematic method of monitoring POCT analysis quality was created to avoid significant differences between measurement results from central clinical laboratories and POCT analyses. Regrettably, many countries have omitted this systematic way of quality monitoring, and as a result of market-moving development of POCT, they continue to struggle with the problem of POCT quality and its links to the laboratory medicine analytics today.

Nowadays, POCT is an important part of laboratory diagnosis and has a firm place in the organisational structure of *laboratory medicine*.

36.1.2. POCT Instruments

POCT tools and equipment are usually operated by people who do not have primary training or qualifications for laboratory work. The reaction principle, operating procedure, overall technical solution and the instrument design are all adjusted to this fact. The instruments may be portable or fixed. The test is usually performed on whole blood, without centrifuging, which would delay the time before the result is obtained. Hence the benefits of these instruments include the use of whole blood, very small sample volumes (micro-volumes), a short time of analysis and a shorter turnaround time (TAT – time between sample collection and getting the result). Internal quality control as well as an independent external quality control programme (control sera, samples with specific concentration of the analyte measured) should be available for the POCT instrument.

It is true, however, that many POCT systems have not been verified sufficiently and reliably enough and their documentation is insufficient as regards basic analytical features of the measurement method used. NOKLUS/SKUP, a Norwegian organisation for POCT system control, has revealed that based on measurement results compared with results from clinical laboratories, many of these analyzers do not provide reliable tests which could be used for clinical decisions. As the spectrum of POCT systems with their varied measurement technologies keeps developing and growing, it has become more and more imperative to rely on standards and other documents recommending adequate methods for their implementation and use.

There are many POCT instruments of varying quality on the market, so great attention should be paid to selecting the right POCT instrument.

It follows from the definition that like all laboratory instruments, POCT instruments rank among the category of in vitro diagnostic medical devices covered by European Directive 98/79/EC and Czech Government Regulation No. 453/2004 Coll., as amended. The user's basic responsibility here is to verify POCT system performance parameters specified by the manufacturer. This verification is impossible without collaboration with a clinical laboratory!!!

Examples of POCT instruments (blood gas, ion and metabolite analyzers; CRP, lactate, glucose, haemoglobin, ACT, troponin, alcohol analyzers, etc.)

36.1.3. Purchase and Implementation of POCT Instruments

Where is POCT equipment most commonly used?

- At the hospital
  - Intensive care units, metabolic units, operation theatres, post-operative wards, central admission, ...
- Decentralised hospital sites
- Health-care facilities other than hospitals
  - Outpatient offices, nursing institutes, hospices
- At the patient's home
- Field use
  - Rescue and emergency services, police

When is the right time to buy a POCT instrument and how to do it?

POCT instruments are needed in particular where tests for important parameters are poorly available or unavailable in the required time. Given the purpose of acquisition, the new equipment has to help demonstrably improve patient care and has to be analytically acceptable.

The POCT TAT has to be shorter than the lab TAT. Depending on the nature of the test and considering the local conditions, one should determine which difference between the two TATs can be medically and economically justified. The only exception to this rule are situations whereby a lab test result is not available round the clock (in the afternoon, at night, on weekends and holidays, etc.), and the test is "acute" – leading to an immediate and urgent reaction of the attending physician.

The following questions should be answered *when deciding about purchasing a POCT instrument:*

Clinical character (physician):

1. TAT, TTAT (turnaround time, therapeutic turnaround time)
2. Purpose of use (diagnosis, monitoring, screening and other aspects)
3. The range of tests measured
4. Expected number of tests within an interval of time (frequency)
5. Critical limits and the way of reacting to results exceeding them
6. Test reporting to healthcare payers
7. Patient benefit to cost ratio

The following issues should be assessed *before purchasing a POCT (analyst):*

h. Pre-analytical issues:

- Method of requesting the test (order slip), identification and documentation of requests
- Type of collection kit
- Type of anticoagulant and other admixtures (potential measurement interference)
- Type of sample (capillary, venous or arterial blood, saliva, urine, ...)

- Number of steps in handling the sample before analysis (sample treatment – the more steps, the more likely a pre-analytical error is)
- How fast the sample is prepared for analysis (influence of time on the resulting value)
- Suitability of the POCT for use in the target group of patients
- Storing of reagents

i. Analytical issues:

- Sufficient scope of the method for the intended purpose
- Same measurement units on the POCT and laboratory instruments
- Comparable cut-offs of the POCT and laboratory instruments
- Assessing the influence of calibration to the plasma or whole blood, inclusion of correction factors in the electronics of the instrument
- Assessing the influence of interfering substances
- Verification of basic analytical features of the measurement method (LoD, LoQ, analytical sensitivity, bias, precision, measurement range, calibration curve, comparison with the routine laboratory method, interferences), or a reference to relevant independent and renowned sources of this information
- Quality management (internal quality control, external quality control, comparability of the same test results (if exists) with an associated clinical laboratory, etc.) If an independent external control programme is unavailable for the POCT system, then it cannot be recommended for use as the quality of the measurement results cannot be checked on a regular basis.
- Evaluation in line with good laboratory practice principles

j. Post-analytical issues:

- Possibility to connect the POCT instrument to the LIS and NIS (hospital)
- Unique identification of measurement results from POCT equipment in order to distinguish them from the same test performed in the lab
- Possibility to report to insurance companies
- Required use of the correction factor specified by the manufacturer (whole blood to plasma conversion, or correction to haematocrit)
- Method of keeping and archiving technical and laboratory documents (maintenance, repairs, findings, results, etc.)

**laboratory assistance is required; option to ask the dealer for a demo version of the instrument for some time to assess the pre, post and analytical phases of the POCT instrument.*

Implementation of the POCT Instrument

k. Hospitals and their field sites

l. Outpatient departments and other health-care facilities

m. Self-monitoring

a) Hospitals and their field sites

Since POCT forms an integral part of laboratory diagnosis and has its firm place in the organisational structure of laboratory medicine, the hospital should also have a multi-disciplinary POCT team, whose main goal it is to unify, improve and streamline the processes of bedside laboratory diagnosis and, if possible, unify POCT instruments within the healthcare facility.

The POCT team should have permanent members and flexible members. Permanent members should include a POCT team leader (ideally laboratory personnel), supervisors for each laboratory specialisation (laboratory personnel), a POCT quality manager (laboratory personnel), requesting site representative, technical and operating representative, financial and analysis representative, legal representative, IT representative, the POCT physician also plays an important role, and/or other members. Flexible members should include representatives of clinical sites where the POCT instrument is located.

Example of POCT instrument purchase and implementation:

The prospective POCT buyer – clinical site representative addresses the POCT team leader. The POCT team leader instructs the supervisor, and assesses the benefit of POCT instrument implementation with the clinician (to discuss the clinical, analytical, pre and post-analytical aspects of the request). Following consensus, the supervisor provides a demo version of the POCT instrument and prepares specifications for the public contract. The POCT team leader coordinates the technical and operating representative and the legal representative to make the purchase, while also instructs the IT representative to provide IT to connect the instrument to the network (if the POCT instrument allows it). The requesting site representative familiarizes himself/herself with the new consumable item. The supervisor with the technical and operating representative and the IT representative ensure installation of the instrument. Before commissioning, the supervisor ensures that the instrument is verified in the clinical laboratory and prepares the relevant documents. The POCT quality manager sets up the quality management system. Together with the supervisor, the dealer demonstrably trains users. The supervisor determines the trial period, and after that gives permission for regular use. The clinical site appoints a "coordinating nurse", who is responsible for the instrument and communicates with the supervisor. The supervisor oversees the whole POCT process throughout the entire use of the POCT at the hospital.

Successful use of POCT equipment to the patient's benefit is dependent on the physician's (user's) collaboration with the clinical lab!!!

b) Outpatient departments and other health-care facilities

After answering clinical questions, the prospective POCT instrument buyer (doctor) should choose and ask a clinical laboratory for assistance on assessing the pre-analytical and analytical phases. Following the purchase and training on the instrument, the instrument should be sent to the associated clinical laboratory for verifying the measurement method. The laboratory issues a verification protocol and recommends a quality management system (as recommended by the manufacturer and professional society). The doctor is responsible for successful internal and external quality control. Prior to introducing the POCT method, POCT users are required to attend a single introductory course on the POCT topic, if such a course exists.

c) Self-monitoring

A POCT instrument lent by the healthcare facility to the patient in home care should be checked on a regular and demonstrable basis through internal controls (at least before lending and after returning) and verified in the associated clinical laboratory. The healthcare facility has to keep the required documents for the POCT instrument. The patient has to be duly and demonstrably trained.

Prescription POCT instrument or POCT instrument gained by the patient from other sources:

The doctor should check whether the patient was trained and can handle the POCT instrument correctly. The doctor should inform the patient about the importance of internal controls and the possibility of verifying the POCT instrument in the clinical laboratory when in doubt about any measurement.

36.1.4. Regular Education

Regular education is the basis for obtaining a correct result, especially when the operating personnel do not have a specialised laboratory qualification. Education is guaranteed by supervisors or special courses containing:

- Theoretical background for the relevant test and technology used;
- Character, collection, storage and method of biological material application;
- Procurement and storage of required consumables, reagents and replacement parts;
- Equipment operation, calibration and measurement procedure;
- Sharing results, transfer to the LIS/NIS, documenting and archiving;
- Evaluation of results and clinical significance of the test;
- Sources of errors at all phases of the measurement process;
- Quality management, documenting and archiving of internal and external quality control results;
- User maintenance and service of POCT instruments.

Absence of verification, internal and external quality control, inadequately trained user and poor knowledge of the pre-analytical phase of the test are the most common causes of incorrect results, causing potential damage to

the patient. The user must not be satisfied with any result; they must require the correct result!

36.2. References

14. ČSN EN ISO 22870:2006 Vyšetření u pacienta (VUP) – Požadavky na kvalitu a způsobilost.
15. ČSN EN ISO 15189:2007 Zdravotnické laboratoře – Zvláštní požadavky na kvalitu a způsobilost.
16. The Norwegian Quality Improvement of Primary Care Laboratoriem. http://www.skup.nu
17. Doporučení ČSKB ČLS JEP – Správné zavádění a používaní POCT (duben 2011). http://www.cskb.cz
18. Program externího hodnocení kvality v oblasti klinické biochemie a nukleární medicíny, hematologie, transfůzní služby, alergologie a imunologie. http://www.sekk.cz

36.3. Automation and Consolidation

36.3.1. Definition of Automation

The use of control systems for equipment and process control.

36.3.2. History

In the early days classical quantitative analytical techniques were used in biochemistry laboratories. The only instruments available were pH-meters, polarimeters, simple photometers, chromatographic techniques were only at the stage of development. First multi-channel analyzers were developed in the 1960s, first microprocessors converting the analytic signal to analyte concentration in the 1970s, and the first protocol allowing communication between the analyzer and the laboratory information system emerged. The 1980s are an important milestone in automation – first biochemistry analyzers to process larger numbers of samples appeared. In early 1990s a trend of laboratory test decentralisation emerged; biochemistry analyzers appear at the patient's bedside, in operation theatres or physicians' offices (POCT technology).

36.3.3. Types of Automation

- *Total laboratory automation* means that all processes (from the receipt of samples to releasing the result) are automatic and part of a complex line. Major benefits of total automation include the processing of a large number of samples at a defined rate, reduced number of errors due to human factor limitation, higher personnel safety, lower number of operators and related lower personnel costs.
- *Modular laboratory automation* means the automation of some laboratory processes only. These are basically stand-alone pre-analytical and analytical systems.

36.3.4. Pre-Analytical Phase Automation

The pre-analytical phase of laboratory testing is the most common source of errors caused mainly by the human factor. Pre-analytical modules were developed with the aim of minimizing these errors.

Capabilities of the pre-analytical system:

- Identification of material with a barcode
- Primary test tube type discrimination
- Sorting of samples
- Sample amount and quality check
- De-capping of test tubes
- Preparation of aliquot tubes
- Overview of sample movement within the lab
- Traceability of samples (e.g. when additional parameters are required)
- Capping of primary tubes
- Archiving of primary tubes

- Increased laboratory staff safety (minimal staff exposure to biological risks)
- Reduced risk of confusing samples
- Saving of primary test tubes
- Faster turnaround time (TAT)

Pre-analytical phase automation also includes pneumatic tube transport, transport equipment connecting the lab and other departments for transporting material automatically. Speed is the greatest asset of pneumatic tube transport. The acquisition cost is a drawback.

36.3.5. Analytical Phase Automation

Currently the most sophisticated phase thanks to high-performance biochemistry and immunochemistry analyzers, a growing number of which are capable of connecting the pre-analytical and post-analytical system. Automated analyzers can perform up to thousands of different analyses per hour.

The following activities have been automated:

- Sample and patient identification
- Sample and reagent dosing
- Incubation of mixtures
- Method calibration
- Measuring the resulting analytical signal
- Recording, analysis and saving of data

Analyzers are categorised based on different criteria, basically into a) continuous flow analyzers and b) discrete analyzers. Depending on the sample analysis arrangement, analyses are divided into sequential, serial, parallel and random-access analyses, where any method in any sequence of any sample can be selected.

Sample and reagent consumption has also decreased significantly. Sample consumption in the analyzer usually does not exceed 5 µl (range 1–30 µl); total reagent consumption does not exceed 250 µl. Methods are generally designed as 1-reagent or 2-reagent methods.

The automation of the analytical part of the test guarantees accurate, comparable and reproducible results.

36.3.6. Post-Analytical Phase Automation

The post-analytical phase has an inter-disciplinary character of collaboration between the lab and the indicating physician. Among other things, it consists of the interpretation of results in relation to physiological values, other test results and the patient's clinical condition. The post-analytical phase of clinical lab testing includes methods of keeping and storing the biological material after the test, validation, release, interpretation and comments on results, consulting and education. The tools of process automation include computer programs facilitating the result check, which are part of the laboratory information system and give the user an option to preset different criteria for checking the results (e.g. setting reference, pathological, warning and critical limits for each parameter). Results sent to the laboratory system following the analytical check by the lab staff and which fulfil the preset criteria proceed to a subsequent automatic medical check and are released to the requester. Results which do not fulfil the preset criteria are not released and have to be reviewed by either a physician or authorized staff with university qualification. Only after this review has been made are the results released from the lab.

The data and information transfer between the lab and the requesting physician is most often done via electronic communication between the laboratory and hospital information systems. Results can be sent to external requesters by mail or secured electronic communication.

36.3.7. Major Advantages of Automation

- Less collection tubes – better usage of the biological material – more careful for the patient)
- Shorter turnaround time
- Elimination of errors

- Reduced costs
- Higher efficiency and productivity of work
- Higher occupational safety

Further development leads to a greater use of POCT, miniaturisation of laboratory tests using microfluid and nano-chip technologies, or further development of molecular biology.

36.3.8. Consolidation

- Combining different analytical technologies or strategies in one instrument or a group of instruments –the idea of a centralised laboratory
- Integration – connection of analytical instruments or instrument groups with pre-analytical and post-analytical equipment
- Consolidation may combine sites of biochemistry, genetics, haematology, immunology, molecular biology, microbiology, parasitology, virology and in vitro nuclear medicine methods
- This is not any "liquidation" of the disciplines, but interdisciplinary collaboration
- Analytical parts of the disciplines are consolidated, not interpretation parts, which, conversely, benefits from the consolidation – comprehensive picture of the patient (all test results together) and thus the comprehensive interpretation of results
- The benefits include lower operating costs (collective purchase of reagents and other material, lower personnel costs), uniform material receipt and processing, fewer collection tubes, all of which is more considerate for the patient, shorter turnaround time, fewer errors, one single collection site
- It is also associated with certain problems – consolidation often ends by collective receipt and processing of material; the disciples sometimes have different opinions on the sense of consolidation

36.4. References

19. Průša, R., a kolektiv: Kazuistiky a stručné kapitoly z klinické biochemie, ÚKBP, UK, FN Motol, Praha 2009
20. Racek J. a kol. Klinická biochemie. 2. vydání. Galén, Praha, 2006.
21. Zima T. a kol. Laboratorní diagnostika. 2. vydání. Galén, Praha, 2007

36.5. EBM

36.5.1. Introduction

Evidence-Based Medicine (EBM) is "conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients".

Moreover, every patient is unique and the human genome is also unique. As the newest part of modern medicine, genetic analysis is based on examining the individual characteristics of each gene. Personalized medicine provides each patient with the best, highly effective tailored treatment with minimal side effects. This leads to a streamlined treatment process and, last but not least, less suffering. Individual treatment is one of the most sought after improvements offered by predictive genetics in modern medicine.

36.5.2. Principles of EBM

Evidence-based medicine integrates:

- External findings (evidence) of the clinical research; this mainly involves exploring the scientific basis of validity of the problem to be solved (diagnosis, treatment, prognosis, etc.) in line with latest published and critically reviewed results of clinical research;
- The physician's clinical experience comprising the physician's expertise and ability to use new knowledge and skills properly to assess the patient's clinical condition;
- Patient's expectations, firstly respecting the patient's wishes, opinions, preferences and needs.

Evidence-based medicine forms part of the managerial decision-making process. This approach requires the understanding of inconsistent results and the assessment of quality and cogency of evidence (published or not published). In the final phase, the physician has to be able to apply this approach to a specific patient in the context of that

particular health system.

The EBM decision-making process can be divided into six inter-related steps:

- Transforming a problem into the clinical question;
- Use of “internal evidence” to answer a clinical question;
- Searching for “external evidence” to answer a clinical question;
- Critical assessment of external evidence;
- Integrating external and internal evidence;
- Evaluating the decision-making process.

EBM is a practical continuation of clinical epidemiology because when making the decision, the physician obtains the required scientific evidence from published epidemiological studies, among other things. Clinical decision-making involves many open questions due to the difference between the theory of each medical discipline and clinical practice. Sometimes the opinion of an “authority” can be considered, but these people sometimes contradict each other. Another option is to seek the answer in professional literature, i.e. assess the validity of research results. Every physician, although they do not wish to be involved in research work, should understand the principles of research to be able to interpret the research results.

36.5.3. Sources of Information

To answer questions arising in connection with patient care (clinical questions), physicians use different sources of information:

22. Their own experience;
23. Experience of their colleagues;
24. Scientific literature.

MEDLINE, the global bibliographic database freely available on the Internet at the PubMed (www.pubmed.gov) information portal, is the best known source. The database offers tools to categorise found articles according to the study type, called EBM filters. Selected items may include review articles, controlled clinical trials, recommended procedures, meta-analyses, etc. A method for the processing of a new type of review articles, the systematic review, has been created in the system for EBM purposes. The evidence pyramid (Fig. 1) ranks systematic reviews among the most significant types in the hierarchy of publications as they seek to answer a specific clinical question in a comprehensive way. Systematic reviews spare the time for searching, study and the reader’s own critical assessment of a large amount of literature.

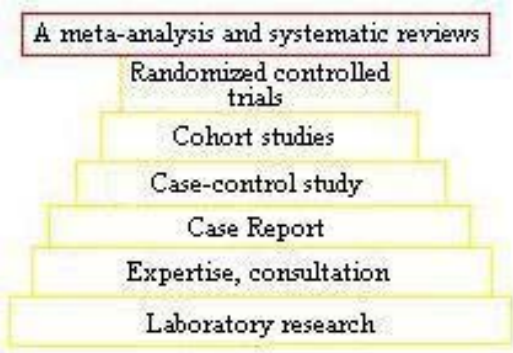


Figure 36.1. Hierarchy of Scientific Evidence in Biomedicine

36.5.4. Personalised Medicine

Personalised medicine is a multidisciplinary branch of science which studies the optimisation of therapeutic and diagnostic procedures in a specific patient using modern methods. It is often referred to as “The right diagnostics, right treatment to the right patient at the right time”. It has evolved in reaction to the modern development of drugs and diagnostic/therapeutic procedures, including drugs which are often marketed as soon as the condition of efficiency verified in phase III clinical studies has been met with standard summary significance. These procedures do not deal

with specific patient treatment, and before personalised medicine emerged, nobody was interested in why some drugs do not work in a fraction of patients.

While classical medicine aims to cure most patients (and puts up with a drug being efficient in 97% patients), personalised medicine asks “How to do it if I wish to increase the likelihood of curing a specific patient?”

36.6. References

25. Sackett, D.L.; Rosengerg, W. M. C.; Gray, J. A. M., Hayes R. B.; Richardson W. S. Evidence based medicine: what it is and what it isn‘t. BMJ, 1996, roč. 312, čís. 7023, s. 71-2.
26. Claridge, J. A.; Fabian, T. C.. History and development of evidence-based medicine.. J Word Surg, 2005, roč. 29, čís. 5, s. 547-553.
27. Kniha v původním znění „How to read a paper – the basics of evidence based medicine“ vyšla již ve dvou vydáních (1997, 2001 v nakladatelství British Medical Journal). Získala cenu „Prix du Livre Médical et Pharmaceutique 2000 – La revue Prescire, Francie“. Byla přeložena do pěti jazyků. Je hodnocena jako bestseller – vynikající učebnice základů „evidence based medicine“ – celosvětového trendu v medicíně.
28. Janout V., Klinická epidemiologie – základ pro medicínu založenou na důkazu. Časopis lékařů českých 2003, 142(2): 67-69.
29. Porzsolt F. et al., 2003, Evidence-based decision making − the 6-step approach. ACP Journal ClubNov-Dec, 139:A11-2.
30. Mihál V. et al., Application of evidence based medicíně process to undergraduate and postgraduate medical education. 3rd International Conference of Evidence-Based Health Care Teachers & Developers „Building bridges between research and teaching“.Taormina (Italy), 2nd-6th November, 2005. Abstract Book, p.71, Online: http://www.ebhc.org/2005/Abstract%20book.pdf
31. Greenhalgh, T.. Jak pracovat s vědeckou publikací. Praha : Grada, 2003. S. 201.
32. Potomková J., Význam systematických přehledů pro medicínu založenou na důkazu. Pediatrie pro praxi 2004, č.2:105-106. Online: http://www.pediatriepropraxi.cz/pdfs/ped/2004/02/17.pdf
33. Pritchard, D. J., Korf, B. R.: Základy lékařské genetiky, Galén, 2007
34. Nussbaum, R. L., McInnes, R. R., Willard, H. F.: Klinická genetiky, Thompson & Thompson, TRITON 2004
35. Murray, R. K., Granner, D. K:, Mayes, P. A., Rodwell, V. W., Harperova biochemie, Nakladatelství a vydavatelství H & H, 2002
36. http://www.fnol.cz, http://www.inno-med.net, http://public.fnol.cz/www/tm/innomed/letaky/EBM.pdf

36.7. OMICS

36.7.1. INTRODUCTION

Research teams consisting of geneticists, organic chemists and biochemists, computer and other specialists deeply investigate into the discipline called functional genomics. This discipline studies concurrent action of hundreds to thousands of genes and develops software for such studies. Structural genomics aimed at determining the structure of most proteins coded by certain genome is one of subdisciplines of functional genomics. Three-dimensional protein structures and functions are the key for targeted drug designs. Transcriptomics is a subdiscipline determining the connection and disconnection of a set of mRNA transcripts in a certain type of cells or tissues, and changes in the expression of genes caused by diseases. Proteomics establishes which protein sets are present in the cell and what their mutual interactions are. Finally, metabolomics analyzes the metabolome, a complete set of small molecules in the cell, and identifies its changes.

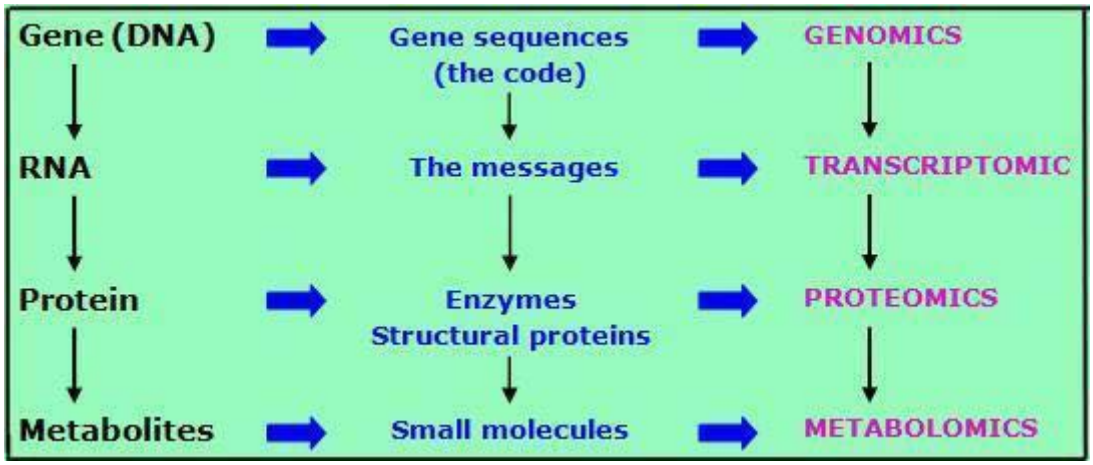


Figure 36.2. Genomics, proteomics, metabolomics, transcriptomics

36.7.2. 39.4.2. GENOMICS

Genomics is a discipline in molecular genetics which studies the genome of organisms. The discipline involves DNA sequencing, genetic mapping and genome annotations, i.e. searching for genes and other functional elements (effort to identify the meaning of the sequence), and intergenomic research – effort to compare the genomes of different organisms to better understand the process of evolution. Molecular biology and bioinformatics methods are widely used in genomics.

Research areas in genomics:

- Structural genomics – research into the structure of genes and whole genomes (distribution of genes within the genome).
- Functional genomics – particularly the annotation of genomes, searching for genes, determining the gene function.
- Comparative genomics – comparing the genomes of different organisms.
- Population genomics – comparing genomes within a particular population of a particular organism, researching the effects of specific conditions.
- Computational genomics – work with genomics data using computing methods in bioinformatics.
- Personal genomics – sequencing genomes (or genome parts) of human individuals and analysing these; effort to provide doctors with precise information about the patient's genetic predispositions in the future.

Basic methods and approaches in genomics include:

- Sequencing – Sanger sequencing method
- Bioinformatics

36.7.3. TRANSCRIPTOMICS

The study of transcriptomics examines the expression of genes within the whole genome (genome-wide expression profiling). The gene expression is identified indirectly from analysing the mRNA converted into complementary cDNA.

Transcriptome – a set of genome-wide messenger RNA (mRNA) referred to as "transcripts" in one type or a population of cells, tissue or organism. Like the proteome, and unlike the genome, the transcriptome is dynamic; changes based on external influences.

Genome-wide methods of the gene expression study include:

- cDNA microarrays and oligo-microarrays
- SAGE (Serial Analysis of Gene Expression)
- cDNA-AFLP (Amplified Fragment Length Polymorphism)

36.7.4. PROTEOMICS

A scientific discipline which studies (analyses) the expression of genetic information at the protein level (proteome) in a sample (organelle, cell, tissue, fluid, etc.). In addition, proteomics examines the three-dimensional structure of proteins and protein interactions.

Proteome – the entire set of proteins in a cell (sometimes in a sample) under specific external conditions. The first studies in proteomics date back to 1975 when O'Farrell, Klose and Scheele began mapping proteins from *Escherichia coli* and *mammals* using 2D gel.

More generally, proteome can be seen as cellular and complete proteome.

Cellular proteome is a set of proteins found in a particular cell or a particular cell type under a particular set of conditions. Example: Hematopoietic cells in the anaphase splitting into the proteome.

Complete proteome subsequently refers to the proteome of an organism, a sum of proteomes of all cell types. This term also includes proteins which may be present in the organism but are not synthesized at the given moment.

Proteome is larger than the genome, particularly in eukaryotes. This is mainly the result of alternative splicing and post-translational protein modifications such as phosphorylation or peptide chain splitting.

While the genome is relatively stable, proteins change very fast and in different ways in the cell. This makes the proteome analysis more difficult. All the cell's proteins are analysed at the same time.

Types of proteomics:

- Protein expression proteomics – approach comparing changes in protein expression between samples (such as in the study of causes of diseases, study of pathogenesis, study of virulence factors in microorganisms, etc.)
- Structural proteomics – analysing the expression and function of proteins in sub-cellular structures (such as organelles or complexes).
- Functional proteomics
- Post-translational modification – phosphoproteome
- Protein-protein interaction (yeast two-hybrid assay – interactomics)
- Proteomics in drug research

Two major approaches in proteomics:

- Classical proteomics – 2D separation techniques – protein separation – identification (MALDI-TOF)
- "Shotgun" proteomics – separation (usually by ESI-HPLC) of protein peptides (fragments) following proteolysis of the complex protein mix by trypsin, identification using LC-MS/MS analysis

Separation in proteomics:

- Gel-based vs. non-gel-based separation
- [HPLC (High Performance Liquid Chromatography), GC (gas chromatography), CE (capillary electrophoresis)]

Identification in proteomics:

- *Mass spectrometry* (MS), tandem MS
- (MS/MS, MSn) – the MS units separate collision cells where the selected fragment is split by inert gas

Quantification:

- In gel or MS/MS

Ettan™ DIGE System – Fluorescent staining of two protein samples with two dyes in gel vs. iTRAQ using MS

36.7.5. METABOLOMICS

Metabolomics is a scientific study of metabolic (low-molecular weight) profiles of specific cellular processes. It examines dynamic changes in the presence (quantity) of metabolites in the biological system as a result of exogenous

stimuli, pathophysiological changes or genetic disposition – "metabolic fingerprint".

Metabolome is the (quantitative) collection of metabolites (metabolic intermediates, hormones, metabolites of endogenous and exogenous agent biotransformation, secondary metabolites, etc.) in a biological sample (whole organism) analysed by a specific method.

It closely follows the functional state of a biological system.

In 2007, scientists from the University of Alberta completed the human metabolome in urine, blood and cerebrospinal fluid. They catalogued 6500 metabolites, 1500 drugs and 2000 food components (see the **Human Metabolome Database (HMDB)** at http://www.hmdb.ca)

Metabolomic approaches and analytical methods:

37. Extraction of metabolites from biological samples
38. Separation techniques
39. Detection
40. Identification and quantification

Separation methods:

- Gas Chromatography – very effective for volatile, low-molecular weight substances, especially in connection with MS
- High Performance Liquid Chromatography (HPLC) – wider spectrum of analytes
- Capillary Electrophoresis (CE)

Detection methods:

- Mass Spectrometry (MS) – mainly in connection with GC-MS, HPLC-MS.
- Nuclear Magnetic Resonance (NMR)
- Fourier Transform Infra-Red spectrometry (FT-IR)
- Electrospray Ionisation Mass Spectrometry (ESI-MS)

36.7.6. Pharmacogenomics

Pharmacogenomics study the expression and sequencing of nucleic acids with the aim of contributing to the development of new medicines and identification of their mechanisms of action. A much wider spectrum of drug response variability causes than mere variability of metabolizing enzymes have been described such as variability of transport mechanisms, receptors and post-receptor mechanisms for drugs, all of which play a role in pharmacokinetic and pharmacodynamic parameters of the drug. Pharmacogenetic procedures contribute to an ever more complex study of these phenomena.

Genetic Polymorphism

Genetic changes in the activity of drug-metabolising enzymes are a substantial source of a variety of the drug's pharmacokinetic parameters between individuals leading to a variety of different therapeutic responses from patients. Inter-individual differences may be significant, from failure of therapy to toxicity manifestations and a marked incidence of drug interactions. The variability of metabolic activity is caused by the presence of several types of alleles for the particular enzyme, which predetermine the varied activity of the coded enzyme. Hence there are alleles in the population, the carriers of which are likely to have a fast metabolism of specific enzyme substrates (Fast Metabolizers, FMs), and alleles causing substantial reduction of enzyme's metabolic capacity of their carrier (Slow Metabolizers, SMs). For the FM phenotype alleles are dominant which means that the SM phenotype is expressed only in homozygotes for defective alleles, yet enzyme activity in heterozygotes is either normal or slightly reduced (Intermediate Metabolizers, IMs). An enzyme is referred to as polymorphic, if more than 1% of SMs exist within the population. If the incidence of variant alleles varies in different populations, the enzyme may be polymorphic in one population, and non-polymorphic in another.

The frequency of each allele occurring in populations is a result of natural selection. According to the Darwin model, the heterozygous genotype is more favourable than the homozygous genotype in terms of survival and reproduction. Alleles are thus maintained in the population owing to the balance between the incidence of a favourable and

unfavourable genotype. The glucose-6-phosphate dehydrogenase polymorphism is a well-known example: the variant alleles expose hemizygous males and homozygous females to a high risk of severe damage (or death) by neonatal jaundice and haemolysis, while heterozygous individuals are more resistant to malaria.

Polymorphism of Some Enzymes

Cytochrome P-450

Cytochromes P-450 are proteins which play a major role in the catabolism of most xenobiotics, metabolism of some endogeneous substrates (thromboxanes, prostaglandins) and in steroid biosynthesis.

Cytochromes from 14 families and 20 subfamilies have been found in the human body. As regards drug metabolism, the representatives of 4 families are of the greatest importance. Cytochromes are bound in membranes of smooth endoplasmic reticulum and mitochondria, particularly in the liver, but also the kidneys, lungs, small intestine mucosa, brain, adrenal glands, gonads and other organs.

Reactions catalysed by cytochromes in humans lead to easier excretion of xenobiotics from the body to prevent their accumulation. In addition to detoxifying the organism, they are also important for the processing of carcinogenesis and activation of procarcinogens. Isoenzymes metabolising drugs and xenobiotics are not strictly substrate-specific. Several metabolic pathways catalysed by multiple cytochromes have been described for many substances.

CYP2D6

This cytochrome P-450 isoenzyme ranks among enzymes, whose potential clinical impact of polymorphism on pharmacotherapy is generally accepted as it is a catalyzing enzyme of many commonly used drugs from groups including antidepressants, antipsychotics, analgesics, beta-blockers and artiarrhythmics.

CYP2C9

The polymorphism of CYP2C9 is primarily important for the metabolism of hypoglycaemizing antidiabetics, which may cause serious hypoglycaemias in the slowed-down metabolism of SMs, and anticoagulants (S-warfarin).

CYP2C19

CYP 2C19 isoenzyme substrates include drugs such as omeprazole and some antidepressants. Area Under the Curve (AUC) values for omeprazole have been reported as twelve times as high in SMs vs. FMs . As omeprazole is an inductor of another isoenzyme, CYP-450 1A2, following omeprazole administration SMs may show considerably elevated CYP1A2 activity levels as well as accelerated metabolism of drugs and other substances acting as a substrate for this enzyme.

At the same time, CYP2C19 is an enzyme catalyzing the conversion of a proquanil antimalarial drug into an active metabolite, which may lead to insufficient therapy for malaria, especially in some Asian populations with up to 20% of SMs.

CYP2E1

CYP2E1 is especially known for its role in ethanol metabolism (also its inductor), but it is also involved in the metabolism of many other xenobiotics. Among drugs, anaesthetics such as halothane, enflurane, isoflurane are substrates for this isoenzyme, and together with CYP3A4, CYP2E1 is one of the ways to form reactive hepatotoxic metabolites of paracetamol.

Other Cytochrome P-450 Enzymes

The activity of the other two enzymes (CYP3A4 and CYP1A2) largely involved in drug metabolism features a marked inter-individual variability. Rather than genetic polymorphism, this variability is probably caused by many non-genetic influences, to which the two proteins are quite sensitive (induction and inhibition by drugs, physiological and pathological processes of the body, diet, etc.).

Cholinesterase

Cholinesterase exhibits polymorphism important for succinylcholine hydrolysis. The gene for this enzyme has a series of relatively rare alleles, which lead to the expression of an enzyme with reduced activity in roughly 2% of people, while the activity is completely undetectable in roughly 1% of people. This metabolic defect may result in an apnoea

prolonged for up to several hours.

Glucuronyltransferases

Glucuronidation is the most common conjugation reaction in drug metabolism. There are also congenital glucuronication disorders such as the Gilbert's syndrome characterised not only by the affected bilirubin metabolism, but also reduced clearance of many drugs including tolbutamide, acetaminophen, rifampicin and irinotecan. The genetic basis of the syndrome in patients with milder symptoms is the mutation of the non-coding region of glucuronidase gene leading to lower gene expression. The other form with more serious symptoms is probably caused by a mutation encoding sequences of the gene for glucuronidase.

N-Acetyltransferases

Polymorphism of N-acetyltransferase 1 (NAT1) and N-acetyltransferase 2 (NAT2) have been described.

The molecular basis of NAT2 polymorphism is the presence of more than 15 allelic forms, most of which reduce enzyme activity. The inability to acetylate drugs leads to adverse effects such as peripheral neuropathy in patients using isoniazid, or hypersensitivity reactions following administration of sulphonamides. By contrast, fast acetylators are at risk of leukopenia following the administration of amonafide, a pro-drug used in chemotherapy.

Increased activity of the polymorphic enzyme NAT1 leads to an increased risk of colon cancer, but the importance for drug metabolism is not clear so far.

Glutathione S-Transferases

The polymorphism of glutathione S-transferases (GSTs) is very pronounced.

GSTs represent important detoxification pathways; their polymorphism has been primarily studied in relation to tumour chemotherapy. For example, carriers of one variant allele (GST P3) are less susceptible to cytotoxic symptoms of cisplatin treatment than other patients.

Methyltransferases

This is a group of enzymes catalyzing many reactions where variability or polymorphism is present. The importance of polymorphism in this group of drugs has been identified for substances such as thiopurine S-methyltransferases (TSMT) where 4 types of alleles have been described. TSMT represents an important catalytic pathway of cytotoxic agents 6-mercaptopurine, 6-thiopurine and azathioprine. Fatal myelosuppressions caused by high drug concentrations have been described repeatedly in patients with a TSMT deficiency. There is also an elevated risk of secondary malignancies (brain tumours and acute leukaemias) endangering patients with some allele variants.

Transport Proteins

P-glycoprotein (P-gp) is the first described and best-known protein from the large group of membrane transporters referred to as ABC proteins, whose activity depends on ATP hydrolysis. Transported substrates are excreted from the intracellular space by their action. The localisation of P-gp in the liver, kidney and intestine cell membranes leads to the elimination of substrates in bile and urine, and to reduced biological availability of drugs from the GIT. In addition, P-gp is largely expressed in endothelial CNS cells, which prevents penetration of substrates through the blood-brain barrier. Increased P-gp expression in tumour cells leads to the known phenomenon of multiple drug resistance.

The gene encoding P-gp is highly polymorphic. Many alleles do not lead to a change of P-gp activity, however it is likely that not all variants are known so far. In addition to this, P-gp inducibility is variable, while its extent depends on the presence of allelic variants.

The efflux system of P-gp interacts with other pathways leading to the elimination of xenobiotics from the body. For example, CYP3A4 and P-gp have many common substrates, and also have a high activity in some similar localisations as P-gp (liver, intestine).

All of these factors point to the fact that the clinical importance of P-gp is still very poorly described, and understanding this area is an outlook for the future which is sure to bring about some therapeutic procedures. It is already known, however, that the induction of P-gp by rifampicin leads to a significant decrease in biological availability of digoxin. The biological availability of cyclosporin also depends on P-gp activity.

Pharmacogenetics in Cardiovascular Disease Treatment

The introduction of facts known from pharmacogenetics in cardiovascular disease treatment has advanced from the phase of theoretical assumptions to the phase of practical use, especially in antithrombotic therapy. The basis of inter-individual differences in the drug effects is usually the polymorphism of transformation enzymes, transport systems ensuring the drug's fate within the body, and receptors, enzymes or signal molecules affecting the drug's effect itself.

The polymorphism of efflux P-glycoprotein P (ABCB1 3435C→T) in drug resorption or polymorphism of CYP2C19 isoenzyme in bioactivation of clopidogrel are clinically important in antithrombotic therapy. P-glycoprotein (P-gp) genotyping is important for determining the biological availability of clopidogrel; the 3435CC genotype is related to low P-gp activity, higher availability of clopidogrel and therefore a more reliable antithrombotic effect; by contrast, the 3435TT variant is related to high activity and insufficient availability. The detection of the CYP2C19 genotype, namely CYP2C19*2 or CYP2C19*3, is also important; clopidogrel bioactivation is slower in these allele carriers, and higher amounts of the prodrug are degraded by esterases. Vitamin K reductase genotyping is similarly important in endangered people. VKORC1AA carriers constitute a warfarin-sensitive subpopulation, and the same portion of individuals with the VKORC1BB variant are less sensitive to warfarin, meaning they are a warfarin-resistant subpopulation. Similarly, there is a subpopulation with varying speeds of warfarin biodegradation by isoenzymes CYP2C9 and CYP4F2. Resulting combinations of responses at the vitamin K reductase level are co-responsible for the resulting therapeutic dose of antivitamin K.

Given the high clinical importance of different types of polymorphism in the drug effect, there is a trend to introduce new drugs (prasugrel, ticagrelor) with pharmacokinetics independent of the different activity of metabolic and transport systems.

Tumour Biology

Many different biomarkers have been tested in clinical practice to personalise cancer treatment. In other words, based on laboratory testing of these biomarkets to establish whether further cancer treatment would be beneficial for the patient when compared with mere monitoring after radical surgical resection, indicate what type of chemotherapy should be selected from the standard options considering the potential toxicity of chemotherapy, and whether or not biological treatment is indicated in addition to chemotherapy. The rising costs of tumour disease pharmacotherapy are another important reason for properly targeted treatment.

Examples of some biomarkers used

According to the results of many studies, adjuvant fluoropyrimidine-based chemotherapy for stage II of colorectal carcinoma with MSI (microsatellite instability) is seen as beneficial. MSI and other biomarkers such as allelic imbalance on chromosome 18q are currently being tested as part of prospective clinical trials in adjuvant chemotherapy indication.

Other biomarkers are being sought for predicting chemotherapy toxicity. One of them is UGT1A1*28 polymorphism, which is accompanied by a significant decrease in enzyme activity and thereby an increased risk of severe neutropenia.

K-ras mutations are currently a biomarker used on a routine basis. K-ras mutation identification is a unique marker of the absence of a response to the EGFR inhibitor (cetuximab or panitumumab) therapy. As shown by OPUS study results, the cetuximab/FOLFOX combination worsens therapy results when compared with chemotherapy alone in a population with a mutant K-ras form. The same results were confirmed by analyzing CRYSTAL and PACCE studies, where irinotecan was used instead of oxaliplatin. An assay for the presence of K-ras mutation is a minimum condition for initiating cetuximab or panitumumab therapy.

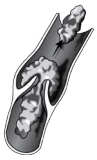
Breast cancer is the most common cancer in women worldwide. Luckily enough, despite the high incidence of this disease, more and more patients are successfully treated thanks to continuous improvement in diagnosis and treatment. Great hope is given to targeted biological treatment able to find tumour cells and stop their growth or destroy them. One of the new drugs is Herceptin, which, as a specific antibody, is effective in treating breast cancer with an elevated expression of epidermal growth factor 2 (HER2 gene). In this HER2-positive tumour type, Herceptin is a specific HER2 receptor blocker, which stops the tumour from growing and may even make it recede.

Monitoring the parametric profile of patients with breast cancer treated with Herceptin therefore presents a unique opportunity to describe the epidemiological situation of HER2-positive patients in the Czech Republic, which may be used by professionals not only for modelling the relationship of risk factors and survival achieved, but also for

economic estimates such as planning antitumour therapy costs.

36.8. References

41. Snustad, P.,D., Simmons, M., J.: Genetika, Brno: Masarykova univerzita, Nakladatelství, 2009
42. Šeda, O., Šedová, L. : Farmakogenomika a nutrigenomika: komplexní interakce genů s prostředím. Klin Farmakol Farm 2005.
43. Proteomická sekce ČSBMB Co je proteomika? http://proteom.biomed.cas.cz/proteomics/proteomics.cs.php
44. Šeda, O., Liška F., Šedová, L., Aktuální genetika - Multimediální učebnice lékařské biologie, genetiky a genomiky, http://biol.lf1.cuni.cz/ucebnice/
45. OMICS: Transkriptomika, proteomika, genomika, metabolomika - komplexní přístupy studia funkce buňky. http://genetika.upol.cz/files/predmety/kbbmtsb_omics_finalni_2010.pdf



CHAPTER 37

37. Anticoagulant Therapy Monitoring

Author: MUDr. Ivana Hadačová

Reviewer: MUDr. Jitka Šlechtová

37.1. Blood Coagulation Physiology

The haemostatic system has two basic functions:

- Haemostasis if a vessel is injured
- Maintaining blood in a liquid state, i.e. preventing blood from clotting inside the vessel

Four systems are used to maintain the equilibrium:

- Primary haemostasis system
- Plasma clotting system
- Inhibitors of coagulation
- Fibrinolytic system

Trauma causes the vessel to contract and restrict the blood flow. Platelets at the site of injury adhere to collagen in the subendothelial ligament of the injured vessel, aggregate, and the content of cytoplasmic granules is released to form a **primary haemostatic plug**.

Blood clotting begins at the same time. Fibrin coagulum is formed by gradual activation of coagulation factors. Coagulation factors are glycoproteins primarily produced in the liver. The coagulation capacity of factors II, VII, IX and X depends on the presence of vitamin K. Most factors have a character of serine proteases, are present in plasma in their inactive form under steady state conditions, and gradually become activated when clotting is triggered – the coagulation cascade is activated. The coagulation system becomes activated in two possible ways: the intrinsic pathway through factor XII (FXII) and kallikrein, and the extrinsic pathway through factor VII. Starting from factor X, the two systems join to a common pathway. Calcium ion and platelet phospholipids have to be present for proper activation. The aim of activation is to produce a sufficient amount of factor II (thrombin). Thrombin releases fibrin monomers from fibrinogen, and fibrin monomers crosslink to form polymers. Factor XIII reinforces the fibrin network (scaffold) to form **final fibrin coagulum**. Coagulation system activation using the extrinsic and the intrinsic pathways separately is mainly performed in laboratory conditions. The extrinsic coagulation pathway is the primary pathway for *in vivo* blood coagulation. The action of extrinsic tenase (FVIIa + tissue factor) initiates blood clotting and a small amount of thrombin is produced. The intrinsic coagulation pathway is activated by thrombin feedback through intrinsic tenase (FIXa+FVIIIa). The process is amplified and prothrombinase (FXa + FVa) is produced. Prothrombinase cleaves inactive prothrombin to active thrombin, which is capable of converting fibrinogen to fibrin coagulum. The activation of clotting factors leads simultaneously to activation of regulatory system – **coagulation inhibitors**. Major coagulation inhibitors include antithrombin (AT), protein C (PC) and protein S (PS). All three of them are produced in the liver, and the function of PC and PS depends on the presence of vitamin K. Thrombin formed in the coagulation process binds to transmembrane glycoprotein thrombomodulin on the vessel endothelium, which activates C protein. The complex of activated PC, PS, thrombin and thrombomodulin, bound to the endothelial protein C receptor, inhibits activated factors V and VIII (Va, VIIIa). AT activity is directed against activated factors II and X (IIa, Xa) and also against activated factors IX, XI and XII (IXa, XIa, XIIa). AT is a slow inhibitor; the reaction of AT with target factors is greatly accelerated by the presence of heparin. The **fibrinolytic system** has a regulatory function. As soon as a clot is formed inside the circulatory system, inactive plasminogen is converted to active plasmin. This conversion is enabled by a tissue plasminogen activator or urokinase. The resulting plasmin cleaves fibrin to form degradation products – the clot is dissolved. The resulting plasmin is eliminated by α2-antiplasmin, and the fibrinolytic system is inactivated by plasminogen activator inhibitors (PAI) at the same time.

37.2. Laboratory Tests

Laboratory tests are used in haematology to examine haemorrhagic conditions, conditions with a tendency for excessive blood clotting (thrombophilic conditions) and to monitor anticoagulant therapy. Most haemocoagulation tests are performed from plasma. Plasma is obtained from blood taken in a test tube with an anticoagulant. The anticoagulant binds calcium ions; the coagulation process is inhibited if they are absent. Most haemostasis tests use 3.8% sodium citrate (1 portion of citrate to 9 portions of blood) as an anticoagulant. Commercially available collection tubes have to be filled with blood up to the mark to keep this ratio. The principle of **coagulation tests** is to measure the time from activation of the coagulation system to the formation of fibrin coagulum. **Chromogenic tests** employ the enzymatic action of coagulation factors to measure the amount of pigment released from the chromogenic substrate.

Tests for monitoring anticoagulant therapy

Activated partial thromboplastin time (APTT) test – coagulation test depending on the level of intrinsic system factors (XII, XI, IX, VIII, prekallikrein, high-molecular-weight kininogens). The test is less sensitive to the deficiency of common pathway factors (X, V, II, fibrinogen). APTT is extended if the above factors are deficient or in the presence of specific and non-specific inhibitors, high titres of products of fibrin and fibrinogen cleavage and heparin. The test is used to monitor anticoagulant therapy with standard heparin. The test result is shown as the patient's time, mentioning normal plasma time (normal plasma is a pool of plasma from healthy donors), or as an APTT-R, i.e. patient's time to normal plasma time ratio.

Prothrombin time (PT) (Quick test) – coagulation test depending on the level of the extrinsic system and common pathway factors (VII, X, V, II, fibrinogen). PT is extended if the factors listed above are deficient, rarely in the presence of specific or non-specific inhibitors. Like APTT, PT is expressed as the patient's time, mentioning normal plasma time, or as a ratio of the two times. PT is also expressed as the international normalized ratio (INR), which considers the sensitivity of the reagent used to reduce vitamin K-dependent factors, referred to as the international sensitivity index (ISI). The test expressed by the **INR is used to monitor anticoagulant therapy with antivitamins K (Warfarin)**.

Anti Xa test – a chromogenic test measuring the X factor inhibition in the plasma of a patient on heparin therapy. The resulting value is read from the calibration curve and expressed in U/ml. **The test is used to monitor anticoagulant therapy with low-molecular-weight heparin** and can also be used to monitor therapy with standard heparin.

37.3. Laboratory Monitoring of Anticoagulant Therapy

Objective: Reach such decrease of blood coagulation capacity that minimizes the risk of thrombosis on the one hand, and the risk of bleeding on the other hand, i.e. **reach the therapeutic level**.

37.3.1. Warfarin

Mechanism of action: Warfarin interferes with the synthesis of vitamin K-dependent factors (II, VII, IX, X) and inhibitors (protein C and protein S), thereby decreasing their function activity.

Indications: Prophylaxis and treatment of venous thrombosis and pulmonary embolism.

Dosage:	Adults	Children
	5 - 15 mg/day; dose is adjusted based on INR	0.1 – 0.35 mg/kg/day; (maximum dose: 15 mg) the dose is adjusted based on INR

Laboratory control:	INR
Therapeutic and preventive range:	2 – 3 2.5 – 3.5 for patients at high risk of thrombosis (mechanical heart valve, repeated pulmonary embolism)

37.3.2. Standard Heparin

Mechanism of action: Standard heparin acts through antithrombin and prevents the conversion of fibrinogen to fibrin coagulum. Standard heparin is not thrombolytic itself, but prevents fibrin coagulum from growing.

Indications: Treatment of venous and arterial thrombosis, prevention of blood clotting during extracorporeal circulation and dialysis.

Dosage: (thrombosis treatment)	**Adults**	**Children**
Bolus:	75 U/kg IV	75 U/kg IV
Maintenance dose:	18 U/kg/hr IV - for continuous infusion	20 - 28 U/kg/hr IV - for continuous infusion
	Dose is adjusted based on APTT	Dose is adjusted based on APTT/anti Xa

Laboratory control:	APTT R	Anti Xa test (more adequate for children)
Therapeutic range:	1.5 – 2.5	0.4 – 0.7 U/ml

37.3.3. Low-Molecular-Weight Heparin (LMWH)

Mechanism of action: By binding to antithrombin, LMWH enhances the inhibitory effect on the activated X (Xa) factor and, to a lesser extent, factor II (IIa).

Indications: Prevention and treatment of venous and arterial thrombosis.

Dosage:	Adults	Children
Prevention	2000 - 4000 aXa IU (20 - 40mg)/24 hrs s.c.	< 2 months: 75 aXa IU (0.75 mg)/ kg/12 hrs s.c. 2 – 12 months: 50 aXa IU (0.5 mg)/ kg/12 hrs s.c. >12 months: 100 aXa IU (1mg)/kg/24 hrs s.c.
Treatment	100 aXa IU (1mg)/kg/12 hrs or 150 aXa IU (1.5mg)/kg/24 hrs s.c.	< 2 months: 150 aXa IU (0.75 mg)/ kg/12 hrs s.c. >2 months: 100 aXa IU (1mg)/kg/12 hrs s.c.
		Dose should be adjusted based on antiXa test result

Laboratory control:	Anti Xa test (first sampling 4 hours after 2nd dose; if dose is changed, 4 hours after the next dose)
Preventive range:	0.2 – 0.4 U/ml

Therapeutic range:	0.5 – 1.0 U/ml

CHAPTER 38

38. Clinical Nutrition and Metabolic Balance

Authors: MUDr. Milan Dastych; doc. MUDr. Milan Dastych, CSc., MBA

Reviewer: MUDr. Daniel Rajdl, Ph.D.

38.1. General Nutrition Disorders

Two extreme nutrition disorders are encountered in clinical practice: undernutrition on the one hand and obesity on the other. While obesity is, for a greater part, a problem of the developed, civilized world, the opposite condition – undernutrition – occurs both in countries with a low social and economic standard as well as in developed countries. Malnutrition in developed countries is usually connected to an underlying disease.

Individuals at risk of or already facing malnutrition can be seen in everyday hospital practice. The course of the disease is worse and is often complicated by fatal impacts of malnutrition. The text below deals with malnutrition experienced by hospitalized patients, the causes, impact and possible options for treating malnutrition.

38.2. Causes of Undernutrition

In general, undernutrition results from taking an unbalanced diet in which certain nutrients are lacking. This imbalance (negative balance) occurs in the following situations:

- Lower food intake (anorexia nervosa, diseases of the teeth and oral cavity, swallowing problems, tumours of the oesophagus, stomach and pancreas, loss of appetite, poverty of the homeless and tramps connected to both malnutrition as well as alcoholism);
- Digestion and absorption disorders (gastrectomy, pancreatopathy, cystic fibrosis, hepatopathy, biliary pathology, lactase deficiency, celiac disease, conditions after bowel resection, blind loop syndrome, inflammatory bowel diseases, intestinal fistulas);
- Metabolic diseases (hepatic and renal insufficiency);
- Increased losses and catabolism due to a disease (fistulas, operations, sepsis, tumours, traumas, infections, abscess).

38.3. Incidence of Hospital Malnutrition

Back in 1885, Florence Nightingale warned at her nursing lessons that “Thousands of patients are annually starved in the midst of plenty, from want of attention to the ways which alone make it possible for them to take food”. Screening studies confirmed this assumption, and about 15 - 20% of hospital deaths are correlated with malnutrition. Malnutrition occurs in 30 - 35% of hospitalized patients. Table 1 shows the incidence of malnutrition in some diseases.

Geriatrics	50%
Inflammatory bowel diseases	75%
Tumours	20 - 80%
Pulmonary diseases	45%
Malnutrition developed during hospitalization	20 - 25%
Total hospital malnutrition	30 - 35%

Table 38.1. Incidence of malnutrition in selected diseases

38.4. Risks and Impacts of Hospital Malnutrition

Malnutrition due to catabolism with increased proteolysis and gluconeogenesis leads to the loss of muscle mass and visceral protein in the body. Reduced protein pool, or reduced functional capacity of the protein system within the body, results in a weaker defensive and reparative immune response. This situation leads to the increased incidence of infectious complications (urinary and respiratory tract infections, early infections) and considerably worse healing of wounds. Another major impact is the loss of muscle mass and strength leading to long-term immobilization, development of decubitus and respiratory failure with consequent dependence of artificial pulmonary ventilation. All these changes result in a longer stay in hospital, higher morbidity and mortality rates as well as increased cost of treatment. Early diagnosis and treatment of malnutrition may prevent these effects.

38.5. Diagnosis of Malnutrition

38.5.1. Patient History

From historical data, the question about weight loss over a certain period of time is essential. Clinically significant weight loss is a loss of 5% in one month or 10% in 2 - 6 months. Another important question is about the amount of food taken in the past ten days. A food intake under ¾ of portions signals a risk of malnutrition. Supplementary questions include analysing dietary habits, questions about indigestion related to food intake, the presence or absence of diarrhoea, social status, or the state of teeth.

38.5.2. Anthropometry and Physical Examination

Basic anthropometric data include weight (kg) and height (cm). However, practice reveals that about 30% of hospitalized patients are not weighed and height is measured only very occasionally. This wrong approach contributes to underestimation of malnutrition, which can be corrected by simple organisational measures. Weight and height values are used to compute the **body mass index** BMI = weight [kg] / height [m]^2, to classify patients (Table 2) and serve to carry out an objective nutritional status. Malnutrition in the population aged 18 - 70 is defined as BMI < 18.5, life-threatening malnutrition as BMI < 12.5. The BMI < 20 is considered a malnutrition limit in patients over 70.

BMI kg/m^2	
Higher than 40	Class III obesity
35 – 40	Class II obesity
30 – 35	Class I obesity
25 – 30	Overweight
20 – 25	Normal range
Lower than 18.5	Malnutrition
Lower than 12.5	Life-threatening malnutrition

Table 38.2. BMI Classification

Another useful index found from the difference of height (cm) – 100 is the **Broca index** to calculate ideal body weight. This simple calculation is used in practice to determine the dose of proteins and calories for enteral and parenteral nutrition. Example: A patient weighing 120 kg, 160 cm tall has a Broca index of 60, which means that his/her ideal weight is 60 kg. The dose of energy and nutrients is then calculated for this weight so as not to „feed the fatty tissue".

Upper-arm circumference (UAC) (characterizes the muscle reserve) and **triceps skin fold (TSF)** (characterizes the adipose tissue) measurements are special examinations that are not commonly used to diagnose malnutrition. There is an exception of assessing malnutrition in oedema conditions such as hepatic cirrhosis. Values between the 5th and 15th percentiles are indicative of malnutrition even with a normal BMI; values under the 5th percentile suggest severe malnutrition as shown in Table 3.

Males	UAC (cm)	TSF (mm)
15th percentile	27	7
5th percentile	25	5
Females		

15th percentile	25	15
5th percentile	23	11

Table 38.3. Limit upper-arm circumference and triceps skinfold values in a nutritional status assessment

The physical examination inquires into the main body structure types - asthenic, normosthenic, obese; aspects to be noted include oedemas of the lower extremities, ascites, alopecia, nail and skin condition, healing of wounds, haematomas or goitre.

38.5.3. Laboratory Tests

Although many laboratory indicators are available, some basic biochemical parameters are sufficient for diagnosing malnutrition. Tests for serum protein (total protein, albumin, prealbumin) and one of inflammatory reaction markers, preferably C-reactive protein, are important. It is difficult to evaluate the level of anabolic protein (albumin, prealbumin, transferrin) in patients with hepatic failure. Low levels in these conditions are indicative of reduced synthetic function of the liver rather than the degree of malnutrition. Assessing the inflammatory condition of the body based on the inflammatory marker (CRP) is essential for diagnosing stress starvation.

The biological half-life of albumin is 21 days, and albumin levels correlate not only with the amount of proteins within the body but are also greatly affected by the inflammatory response of the body, leading to higher permeability of capillaries and albumin seepage into the third space. Rapid drops in albumin level within a short period of time are therefore an indicator of the scale of the stress reaction (a new attack of sepsis, bleeding) and the severity of the disease rather than only an indicator of malnutrition. Stress-induced malnutrition, however, is always present. Albumin levels are assessed with respect to the long-term nutritional status preferably from three weeks values, because fluctuations at shorter intervals are indicative of the severity of the disease or albumin losses through proteinuria or exudative enteropathy. From this point of view, the current protein status of the body is better characterized by total protein. Albumin levels < 28 g/l measured for a long time are considered as an indicator of malnutrition. Prealbumin, an anabolic phase protein with a biological half-life of 2 - 3 days, is a good short-term indicator which can be used to evaluate the adequacy of the selected nutritional intervention. Like albumin, a prealbumin level assessment is limited to non-stress conditions, since the prealbumin level in stress metabolism decreases in direct proportion to the severity of the disease. Prealbumin level assessment is further limited in patients with renal failure, where falsely high values can be found due to decreased glomerular filtration.

The trace element (Zn) and cholesterol levels are suitable indicators of long-term nutritional status. Their levels decrease in people with poor nutrition over a long time. Blood count is a simple and inexpensive indicator detecting malnutrition expressed by typical normocytic anaemia and/or leucopenia.

38.6. Two Types of Malnutrition

38.6.1. Cachexia

The first characteristic type is **simple cachexia** – marasmus (synonyms: marasmic malnutrition, simple starvation, protein-energy malnutrition, the “skin and bone” look). This type of malnutrition occurs due to inadequate intake of energy substrates (carbohydrates, lipids) and proteins. In clinical terms, this condition is characterized by low weight and a low BMI. Laboratory findings are poor. In metabolic terms, simple starvation is characterized by the use of fat reserves as a source of energy. Lipolysis (beta-oxidation) produces active metabolite, acetyl-CoA. Acetyl-CoA is metabolized in the Krebs cycle to produce reduced coenzymes, which are further processed in the respiratory chain to generate energy in the form of ATP. Acetyl-CoA is also used for gluconeogenesis.. When the Krebs cycle capacity has been exceeded, excess acetyl-Coa leads to the synthesis of alternative sources of energy – ketone bodies (β-hydroxybutyric acid, acetoacetic acid, acetone). The reason for this metabolic change is considerable reduction of protein catabolism and maintenance of their function. During developed simple starvation, protein catabolism is limited to 10 - 20 g/day. This primarily involves slow elimination of muscle protein. Released amino acids are used for synthesizing plasma proteins, the level of which remains normal. If normal body fat percentage (20 - 25%) is taken into account, this reserve in a 70-kg individual can be the source of 126,000 kcal, i.e. an energy reserve for 60 - 70 days. This assumption corresponds to actual observations in concentration camps during World War Two.

38.6.2. Kwashiorkor

Another type of malnutrition is **kwashiorkor** (the name is derived from the African Ga language and translates as "the sickness the baby gets when a new baby comes"). In African conditions, this type of malnutrition has exogenous causes (see the note). "**Kwashiorkor-like malnutrition" = protein malnutrition = stress starvation** has endogenous causes in our conditions. Aetiopathogenesis can be varied. This type of malnutrition is due to insufficient protein intake, increased losses (proteinuria, exudative enteropathy), increased degradation (catabolism) or reduced synthesis (hepatic cirrhosis) of proteins, i.e. absolute deficiency of proteins. In clinical terms, this type of malnutrition is characterized by fluid retention, oedemas, ascites and weight gain. Laboratory tests show a marked drop in serum proteins and albumin, and elevated acute phase proteins (CRP, orosomucoid, caeruloplasmin, α_1-antitrypsin) in patients with an inflammatory reaction. In developed countries, KW-like malnutrition is most commonly related to a stress situation (polytrauma, surgery, sepsis, acute pancreatitis, burns). Metabolic changes are regulated by the hormonal environment in the body. Metabolic stress induces a drop in the production of insulin and elevated output of counterregulatory hormones such as cortisone, glucagon and adrenalin. This results in the blockade of lipolysis, accelerated proteolysis and gluconeogenesis from proteins. Proteins thus lose their phyologenetic function (i.e. structural support, signalling, transport, defence), and become the source of energy for glucose-dependent tissues (brain, erythrocytes, lymphocytes, renal cortex) as well as other tissues such as the heart and liver resulting from of the absence of ketone bodies. Their deaminated carbon chain is a reservoir of energy, like carbon chains of triacylglycerides derived from fatty acids. The organism is unable to survive if more than 50% of body protein reserve is lost. If protein catabolism with losses of 30 g N (about 180 g of protein) per day and average 20 kg of protein tissue in the body is considered, then maximum time of survival with this type of starvation is 10 - 14 days. Compared to simple starvation, stress starvation is a more severe clinical condition with rapid development and a potentially fatal outcome.

Pure protein malnutrition occurs in nephrotic syndrome. Combined malnutrition occurs in patients with hepatic cirrhosis, where the synthetic function of the liver is reduced. An inflammatory component is often present in patients with exudative enteropathy, however the drop in albumin is more pronounced than a drop corresponding to an inflammatory reaction. The basic differences between the two types of malnutrition are shown in Table 4.

Indicator	Cachexia	Stress starvation
Inflammation	Absent	Present
Albumin	Normal	Reduced
Body water and Na content	Reduced	Elevated
Body weight	Reduced	Elevated
REE	Reduced	Elevated
Fat	Reduced	Normal
Disease	Anorexia nervosa, Crohn's disease, chronic pancreatitis	Polytrauma, sepsis, operation

Table 38.4. Differences between cachexia and stress starvation

Note:

Kwashiorkor (KW) in the classical "African" form is typical of children 1-4 years old when a mother weans her child, replacing breast milk with a diet high in starches, but deficient in proteins (corn, cassava). KW pathogenesis involves high insulin secretion and conservation of muscle proteins at the expense of plasma proteins (drop in albumin, development of oedemas, reduced lipoprotein synthesis in the liver, development of hepatic steatosis). Clinically similar conditions can be found in the CR (hypoalbuminaemia, retention of fluids with oedemas, ascites) in people with a pronounced inflammatory reaction to stress (polytrauma, burns, sepsis, major surgery, severe acute pancreatitis). As this condition is clinically similar to classical KW, this malnutrition is sometimes referred to as KW-like malnutrition. The pathogenesis of this condition is different, and is induced by stress hormone elevation (adrenalin, glucagon, cortisol) and development of insulin resistance (stress diabetes). This results in rapid catabolism of muscle proteins in particular, increased capillary permeability with albumin seepage into the interstitium, and fluid retention. Unlike the classical KW (weeks), this condition develops faster, within days.

38.7. Who Requires Nutritional Intervention?

Two types of patients requiring nutritional intervention can be found in hospital. These are primary poorly nourished patients in which malnutrition can be recognized on admission (BMI < 18.5, clinically significant weight loss,

albumin < 28 g/l). The second case is malnutrition developing during hospitalization. These are typically inpatients, which have no or minimal oral food intake for 3 - 5 days, most often due to a disease and examination procedures (difficulty in swallowing – cerebrovascular diseases, dysphagia, abdominal pain, acute abdomen).

38.8. Types of Nutritional Intervention – What We Can Offer to Patients

- Instruction by dietitian
- Fortified diet
- Sipping
- Enteral nutrition
- Parenteral nutrition

Instruction from a dietitian:

Consultation with a dietitian specialist often leads to a substantial impact on food intake. The dietitian specialist selects the appropriate type of nutrition according to the individual's needs (culinary preparation, ingredients, mechanical treatment), evaluates the amount of food taken, calculates the dose of nutrients taken, recommends whether enteral nutrition should be used and instructs patients on a specific type of diet.

Fortified diet

Powder protein and oligosaccharide concentrates are added to food. Their taste is neutral so the taste of the food is minimally changed. Products available include Protifar (proteins – casein) and Fantomalt (maltodextrin).

Sipping

Sipping is the most widespread nutritional intervention. Sipping means drinking flavoured balanced enteral nutrition products by little amounts as a supplement to the food taken. They usually represent a daily intake of 300 - 600 kcal, i.e. 1 - 2 packages. There is a wide selection of products differing in energy density, protein content, presence of fibre etc. Examples: Nutridrink, Fresubin, Resource, Ensure, Prosure.

Enteral nutrition

Tube feeding is often used as a synonymic term. This nutritional intervention means the delivery of balanced enteral nutrition products by a feeding tube into the gastrointestinal tract if the patient is unable to take food. Enteral feeding is delivered to the digestive tract by a nasogastric or nasojejunal tube. Jejunostomy or gastrostomy is used as entry into the GIT for long-term enteral feeding. Enteral product solutions are balanced to cover the daily requirement of energy and all nutrients at the same time. In practice this means that if the daily requirement is 2000 kcal (75% of all patients on EN), the patient is given 2000 ml of enteral product with an energy density of 1 kcal/ml. Enteral nutrition is a preferred approach to patient feeding because it is physiological, does not have adverse pharmacological effects and is safer and cheaper compared to parenteral nutrition. Contraindications for tube feeding are sudden abdominal incidents, bleeding to the GIT, ileus.

Parenteral nutrition

This nutritional intervention refers to the delivery of nutritional substrates outside the digestive tract, intravenously. This method of feeding is non-physiological and is associated with many complications connected with administration through vein access such as the pneumothorax, bleeding and embolism, as well as subacute complications such as venous thrombosis, cannula sepsis, and hepatobiliary complications. Parenteral nutrition is administered in the form of all-in-one (AIO) bags, most commonly into the central vein (v. jugularis, v subclavia, v. femoralis). AIO bags use solutions of amino acids, glucose, lipids, ions, vitamins and trace elements. Parenteral nutrition is indicated to ensure feeding in the event that enteral nutrition is contraindicated and nutrition through the GIT is impossible.

38.9. Monitoring Nutritional Status in Hospital

38.9.1. Anthropometry

Anthropometric monitoring depends on the severity of the condition and the method of nutritional intervention. The basic parameter to be monitored is body weight. Weight of patients with severe conditions (e.g. sepsis, organ

failure, circulatory instability) hospitalized at intensive care units has to be measured on a daily basis. Daily weight fluctuations are a good indicator of water balance and easily detect any water retention. The weekly weighing of stabilized patients on enteral nutrition is sufficient. Daily monitoring of the weight of patients with short bowel syndrome and high losses through ostomies, fistulas or drains is an essential prerequisite for correct crystalloid balance. Stand-on scales, chair scales and bed scales are used for measuring the weight. Upper-arm circumference is a suitable anthropometric marker to evaluate the nutritional status of bedridden patients on aftercare beds, which are often not equipped with bed scales. Calculation: **UAC – 5 = BMI** gives approximate BMI and body weight, if the patient's height is known.

38.9.2. Biochemical Parameters

When initiating nutritional intervention, the potential development of refeeding syndrome and overfeeding syndrome has to be taken into account.

The basis of refeeding syndrome is a drop in plasma concentrations of potassium, magnesium and phosphorus. In anabolic processes these substances move intracellularly and become part of synthesized protein macromolecules. Their reduced levels and reduced availability in tissues result in the disturbed transfer of nerve and neuromuscular impulses and relative ATP deficiency with serious clinical effects. The P, Mg and K reserve is reduced in undernourished patients. Administration of nutritional substrates with insufficient coverage of the ions listed above, leads to a very fast drop in their plasma levels. Clinical signs of this syndrome include muscular weakness, muscle spasms, paralysis, disorientation, disturbance of consciousness, and may result in death due to cardiac arrhythmia or diaphragm paralysis. A fatal example of this syndrome from history is deaths of prisoners released from concentration camps. The facts above prove the necessity of regular monitoring of K, Mg and P concentrations and consistent coverage of their requirement.

Overfeeding syndrome – overloading by nutritional substrates occurs if required dose of nutrients is poorly estimated in undernourished patients. Laboratory results show development of hyperglycaemia, elevated triacylglycerols, increased liver test parameters, development of jaundice and elevated urea. This is why the laboratory parameters indicated above should be monitored repeatedly if nutritional intervention is administered, especially to undernourished patients. Clinical signs of overfeeding include elevated body temperature, development of jaundice with hepatomegaly and polyuria.

Thiamine deficiency is often important in patients with diabetes mellitus, hepatic cirrhosis, in chronic ethylic patients and severely undernourished patients (anorexia nervosa). When nutritional intervention is started in these patients, incorrigible lactate acidosis may develop as a result of reduced pyruvate decarboxylase activity with thiamine as a cofactor. The finding of metabolic acidosis and an elevated lactate level with good tissue oxygenation is an unmistakable sign of this condition.

Prealbumin level monitoring is used to evaluate the effectiveness of selected nutritional intervention and the onset of anabolism. An increase in plasma prealbumin concentration within 3 - 5 days is an indicator of an ongoing anabolism. Similarly, the monitoring of nitrogen loss reveals details as to the current metabolic situation. In general, catabolism occurs if daily nitrogen losses are higher than 15 g.

To give some practical examples, listed we have included the frequencies of specific tests for ICU patients here below:

Note:

The monitoring of biochemical indicators, their extent and frequency of tests depend on the patient's clinical condition. Urea, creatinine, ions, P, Mg, total protein, albumin, CRP and glucose assays are indicated for stabilized undernourished patients on supplementary or full enteral nutrition in the ***first week****. After the next 3 - 4 days urea, creatinine, ions, P, Mg and blood glucose profile, and after another 7 - 10 days the parameters listed above plus albumin, CRP, total protein and prealbumin. Follow-up laboratory tests are advisable after 3 weeks at 4 - 6 week intervals.*

Vitamin, trace element, iron, triglycerol and cholesterol levels are checked at the beginning of treatment and subsequently every 4 - 6 weeks.

Stricter monitoring is required for patients in intensive care, often on parenteral nutrition. Urea, creatinine, ions, P and Mg are usually monitored on a daily basis. Blood glucose is tested 4 times a day. If the glucose profile is satisfactory, testing twice a week is sufficient. Bilirubin, liver test parameters, TAG and CRP are checked twice a week; albumin, total protein, prealbumin, Zn and Se are checked once a week. Vitamin levels are not tested regularly. Vitamin D, vitamin C, vitamin A, folic acid and vitamin B12 are sometimes monitored only in complicated cases in metabolic care ICUs. The

laboratory tests listed above are used to evaluate the nutritional status, monitor the kind of nutritional intervention selected and detect possible adverse effects (refeeding syndrome, overfeeding, hepatopathy associated with parenteral nutrition). Additional laboratory tests depend on the specific patient's situation – ventilation and respiration, blood gases, acid-base balance, monitoring of inflammatory reaction development (procalcitonin, CRP), etc.

38.9.3. Nitrogen Balance

Nitrogen balance is essential in order to evaluate the metabolic situation of the body. Nitrogen balance is defined as the measure of nitrogen output subtracted from nitrogen input. Nitrogen balance can be positive, even or negative. Nitrogen losses are determined from daily losses of urea as a final metabolite of amino acid (protein) catabolism. Many studies have proven that urea losses in urine represent 80% of total nitrogen losses. So if daily urea losses are known, one can find out relatively accurate daily nitrogen losses using a simple formula (**urea losses mmol/24 hrs * 0.0336 = total nitrogen losses/g/24 hrs**).

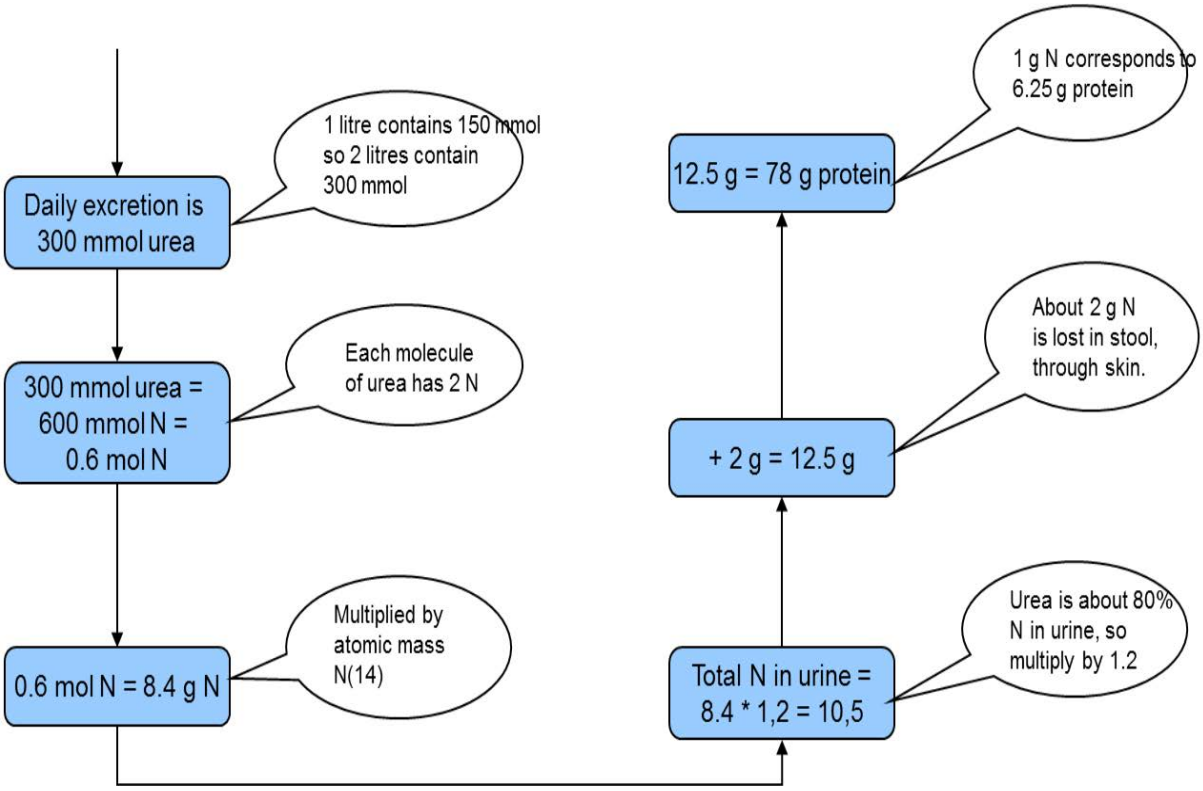


Figure 38.1. Example of nitrogen loss determination (also nitrogen intake in the food or enteral/parenteral nutrition has to be known to determine the nitrogen balance

Nitrogen metabolism is balanced when healthy, which means that the nitrogen (protein) input is in equilibrium with nitrogen losses. Nitrogen balance in anabolic situations (recovery from a disease, pregnancy, bodybuilding, training) is positive – the body produces increased amounts of proteins (covers losses, forms new tissues). In sickness, under the influence of acute (sepsis) or chronic (tumour, cardiac failure) inflammatory reaction of the body, proteolysis is stimulated, proteosynthesis suppressed, gluconeogenesis enhanced and lipolysis influenced. This results in a negative nitrogen balance with protein losses higher than protein input. It is interesting that artificial delivery of pure amino acids in the form of parenteral nutrition (PN) cannot prevent but can only reduce catabolism of endogenous proteins. Body's own proteins (muscle, albumin, immunoglobulins) are used in preference. This observation has led to abandoning the effort to reach even nitrogen balance in severe catabolism. Increased delivery of substrates in PN greatly burdens intermediary metabolism and has adverse effects (hepatic failure – icterus, cholestasis, hyperglycaemia, uraemia, hypertriglyceridaemia) which increase mortality. Therefore, if protein catabolism is found (urinary nitrogen losses of more than 15 g/24 hrs = 90 g, amino acids = 450 g of muscle), the physician should search for the cause and try to solve the situation (treatment for sepsis, abscess or autoimmune disease). Another basic procedure is selecting nutritional intervention (EN or PN) with a stress ratio of nutrients, respecting their maximum doses per kilogram of ideal weight, without even making an effort to maintain the balance. Carbohydrates and lipids are used as an energy substrate in the anabolic phase. Proteins perform specific functions. Contrary to this, proteins are used as a source of energy in stress metabolism, and so protein consumption increases and carbohydrate and lipid availability decreases. Stress-related and anabolic ratios of nutrients have been shown in Table 5.

38.9.4. Nutrient Doses

Nutrient ratio in anabolism		
Carbohydrates	Lipids	Amino acids
6 g/kg	1 - 1.5 g/kg	1 g/kg
Nutrient ratio in stress		
Carbohydrates	Lipids	Amino acids
2 - 3 g/kg	0.7 g/kg	1.5 - 2 g/kg

Table 38.5. Stress-related and anabolic ratios of nutrients

38.10. Determination of Energy Requirement

Three approaches to energy requirement determination are available:

38.10.1. Estimated Energy Requirement

Basal metabolic rate (BMR) = 1 kcal/kg/hour = 24 kcal/kg/day

Total energy requirement (BMR + AF + DF) = 30 - 35 kcal/kg/day

(BMR – Basal Metabolic Rate, AF – Activity Factor, DF – Disease Factor)

Energy requirements for the AF and DF are included in an increase in energy requirement by 5 - 10 kcal/kg/day against calculated energy requirement for the BMR. This calculated total energy requirement is usually not exceeded in clinical practice. There is an exception of convalescing patients with anabolism after overcoming catabolic conditions (polytrauma, major surgery), when the total energy requirement often amounts to 40 - 45 kcal/kg/day.

Energy dose is determined both in terms of the current and ideal weight of the patient. Ideal weight is calculated from the Broca index. For patients with a BMI under 17, energy requirement should be calculated for their current weight, and after 10 - 14 days, the dose of energy should be increased to the ideal weight in order to prevent refeeding syndrome. (See above.)

38.10.2. Calculated Energy Requirement

The Harris-Benedict equation taking the age, weight, height and gender of the patient into account is most commonly used to calculate the energy requirement.

Males: BMR = 66.473 + 13.7516 x W + 5.0033 x H – 6.755 x A (kcal/24 hrs)

Females: BMR = 655.0955 + 9.5634 x W + 1.8496 x H – 4.6756 x A (kcal/24 hrs)

(W – weight in kg, H – height in cm, A – age)

38.10.3. Measured Energy Requirement

Indirect calorimetry is used to measure energy expenditure. This method is used to calculate current resting energy expenditure equal to BMR + DF (BMR – Basal Metabolic Rate, DF – Disease Factor), which is indicative of the current metabolic state of the patient. Indirect calorimetry employs the knowledge about the nature of breathing discovered by A.L. Lavoisier (1743 - 1794): “Oxygen is consumed during respiration and oxygen consumption is proportional to energy expenditure.” This scientist stated the law of conservation of mass in 1774 and is considered father of calorimetry. Indirect calorimeter measures oxygen consumed and carbon dioxide eliminated. Energy expenditure can be calculated from O_2 consumed per minute, and the CO_2/O_2 ratio is used to calculate the respiratory quotient (RQ), which corresponds to nutrient utilization. (RQ = 1 corresponds to pure glucose utilization, RQ = 0.7 corresponds to fat utilization, RQ = 0.8 corresponds to amino acid utilization.) If nitrogen output and therefore protein losses are known, one can determine the amount of oxygen required for their metabolism and the amount of carbon dioxide produced. When these values are deduced from overall O_2 consumption measured and CO_2 expenditure, the resulting non-protein RQ is indicative of the prevailing type of utilized nutrients (carbohydrates [RQ close to 1] or fats [RQ close to 0.7]).

This method of energy requirement determination is suitable for obese people, severe cachexia and patients not responding well to the calculated dose of nutrients.

Below you can find stechiometric formulas showing nutrient oxidation and explaining the RQ values for each nutrient:

Glucose:

$C_6H_{12}O_6 + 6\ O_2 \Rightarrow 6\ CO_2 + 6\ H_2O$

RQ=1, 1gG = 0.74l O2 + 0.74 CO2, 1gG = 4kcal

260 litres of oxygen is required for a daily dose of glucose (350 g)

Fat:

$CH_3(CH_2)_{14}COOH + 23\ O_2$ => 16 CO_2 **+ 16 H_2O**

RQ=0.7, 1g fat = 2.03l O2 + 1.43l CO2, 1g fat = 9kcal

142 litres of oxygen is required for a daily dose of fat (70g)

Proteins:

$2\ C_3H_7O2N + 6\ O_2$ => $(NH_2)_2CO$ **+ 5 CO_2 + 5 H_2O**

RQ = 0.8, 1gP = 0.966l O2 + 0.782lCO2, 1g P = 4kcal

77 litres of oxygen is required for a daily dose of fat (80g)

The Weir equation is used to calculate resting energy expenditure from known oxygen uptake and carbon dioxide output:

REE = 3.95xVO_2 + 1.11xVCO_2 (kcal/24 hrs)

Oxidation of each substrate is calculated from the equation below, requiring the knowledge of daily oxygen uptake, and carbon dioxide and nitrogen output:

Protein oxidation (g/24 hrs) = nitrogen output/24 hrs x 6.25

Carbohydrate oxidation (mg/min) = 4.534xVCO_2 – 3.195xVO_2 – NUx2.0

Fat oxidation (mg/min) =1.669xVO_2 – 1.669xVCO_2 – NUx1.331

38.11. Body Composition

	Males	Females
Proteins	15 - 16	14 - 16
Carbohydrates	1 - 1.5	1 - 1.5
Fat	15 - 22	22 - 28
Water	55 - 64	50 - 58
Minerals	5	5

Table 38.6. Average normal body composition expressed as a percentage

	Males	Females
Muscular tissue	28	17
Fat tissue	15	19
Bones	12	10
Liver + GIT	2 - 3	2

Brain	1.4	1.2
Heart	0.33	0.24
Kidneys	0.31	0.27
Other	9.3	7.3
Total	70	58

Table 38.7. Body composition, by tissue types in a 70-kg man and a 58-kg woman

38.11.1. Fat Tissue and Fat-Free Tissue

Fat tissue contains 80% of fat, 18% of water and 2% of proteins. From this point of view, the human body can be divided into the **fat-free mass (FFM)** and the **fat mass (FM)**. FFM comprises all body components except lipids. FM is body fat: triglycerides, cholesterol and other lipids. A further method divides the body composition into the **adipose tissue mass (ATM)** and the **lean body mass (LBM)**. In addition to fat, ATM also comprises adipocyte membrane protein and water. LBM represents the rest of the body: bones, muscles and internal organs. Given the water and protein content, adipose tissue mass (ATM) is greater than fat mass (FM). The division above is described to be able to understand abbreviations used as an output from bioelectrical impedance instruments.

38.11.2. Bioelectrical Impedance Analysis

The bioelectrical impedance analysis (BIA) is a method that employs Ohm's law (**I = U/R**) to measure tissue resistance (bioelectrical impedance). The instrument works with a high-frequency and low-voltage current. The impedance measurement is used to determine the amount of fat mass (FM) and fat-free mass (FFM), or lean body mass (LBM). Newer instruments work with multiple frequency currents and use the fact that higher-frequency current (over 100 Hz) penetrates transcellularly and lower-frequency current paracellularly. These instruments, referred to as dual BIA, work with several electrodes on the body surface and by alternately measuring impedance between the electrodes emitting currents of different frequencies, they divide the body into different segments, the composition of each is analyzed separately. The extent of the body segment depends on the place where the electrodes are applied. The body is most commonly divided into 5 segments: upper extremities, lower extremities and the torso. The output of the measurement is %FM, %FFM or %LBM corresponding to the muscular tissue. If the body weight is known, it is easy to determine absolute values. In addition, dual BIA measures total body water (TBW), extracellular water (ECW), intracellular water (ICW), body cell mass (BCM) and segmental distribution of fat and muscle, including oedema and retained fluid detection is determined. This examination is used in clinical practice for monitoring body muscle growth during convalescence and adequate physical therapy, body composition and water content monitoring in patients with chronic renal failure for possible adequate dry weight adjustment. The determination of body composition and BCM in particular may detect incipient malnutrition in patients with hepatic cirrhosis, nephrotic syndrome or cardiac failure. The examination motivates obese patients, evaluates possible fluid retention and monitors the loss of FM. This examination may also be used to monitor correctness of selected nutritional intervention.

38.11.3. Body Density Measurement

Body density measurement is used to calculate the content of fat tissue. This method requires complete submersion of the body under water. It is known that body density is given by weight per unit of volume, so if body weight (kg) and body volume (volume of water in litres by complete body immersion) is measured, body density can be determined. Obese people usually have lower body density. Fat density is 0.9 g/ml, fat-free mass (FFM) density is 1.1 g/ml. Equation used to calculate fat percentage from known body density:

%FM = ((4.95/density)-4.5)*100

38.12. Examples – PN Specification for All-in-One Bags

38.12.1. Example 1: PN with anabolic nutrient ratio (Table 41.5.)

30-year-old patient with Crohn's disease, chronic activity, terminal ileus stenosis, subileus, weight 50 kg, height 175 cm, usual weight 65 kg, alb 30 g/l, CRP 2 mg/l.

Amino acids: 1 g/kg of ideal weight (height – 100) =1*75= 75 g = 500 ml 10% Aminoplasmal (50g amino acid) + 500 ml 5% Aminoplasmal (25g amino acid)

Glucose: 6 g/kg of current weight = 6 * 50 = 300 g G = 500 ml 40 % Glucose + 500 ml 20% Glucose

Fat: 1 g/kg of current weight = 1 * 50 = 50 g = 250 ml 20% Lipofundin

Daily doses of minerals, trace elements and vitamins. After a week if metabolically tolerated, the dose of glucose may be increased by 50 g and fat by 25 g.

38.12.2. Example 2: PN with stress nutrient ratio (Table 5)

55year-old patient with severe acute pancreatitis, weight 75 kg, height 172 cm, usual weight 75 kg, alb 22 g/l, CRP 280 mg/l

Amino acids: 2 g/kg of ideal weight = 2*75 = 150 g = 1000 ml 15% Aminoplasmal

Glucose: 2.5 g/kg of current weight = 2.5 * 75 = 200 g G = 500 ml 40% Glucose

Fat: 0.7 g/kg of current weight = 0.7 * 75 = 50 g = 250 ml 20% Lipofundin

Daily doses of minerals, trace elements and vitamins. The glucose and fat dose has to be adjusted according to the triglyceride level and blood glucose profile.