



Kód předmětu: Bi8980

MASARYKOVA UNIVERZITA

Protein expression and purification

- I. The molecular principles for understanding proteins

Lubomír Janda, Blanka Pekárová and Radka Dopitová

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Název prezentace v zápatí

This course aims to develop students' confidence in their ability to understand the way proteins function and the basis of the methods used to distinguish, identify, and characterize them.

- This lecture is intended to provide a concise summary of the principles with which you should be familiar in order to understand the structures and functions of proteins.

Literature:

Exploring proteins – Nicholas C. Price and Jacqueline Nairn

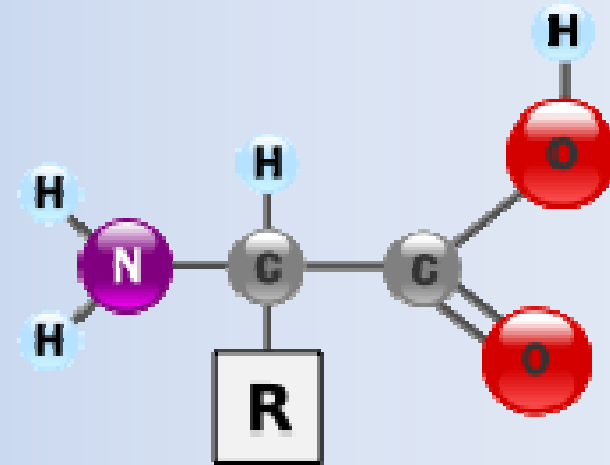
Basic Methods in Protein Purification and Analysis – Richard J. Simpson, Peter D. Adams and Erica A. Golemis

High Throughput Protein Expression and Purification – Sharon A. Doyle

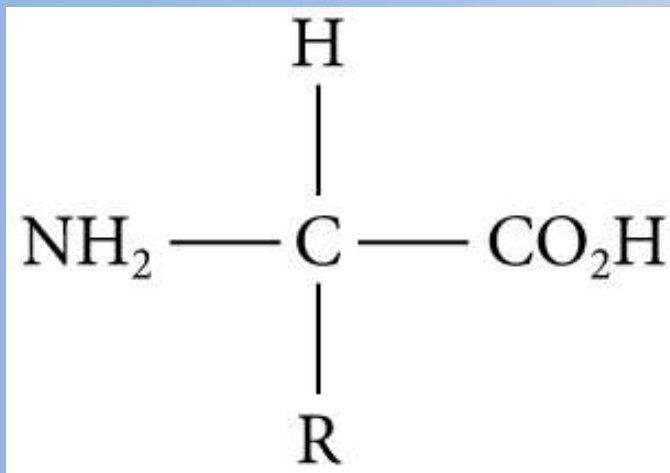
Cloning, Gene Expression and Protein Purification – Charles Hardin, Jennifer Edwards et al.

Protein Purification, Principles and Practise – Robert K. Scopes

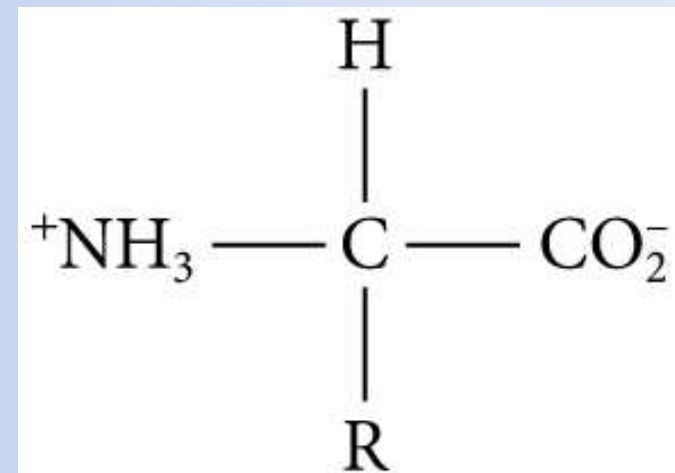
1.2. The amino acids



The constituents of proteins, the amino acids



General structure of an amino acid



The zwitterionic form of an amino acid

1.2. The amino acids

1.2.1. The variety of amino acids

1.2.2. Clasification of the amino acids in terms of polarity

Non-polar side chain

Ala, Gly, Ile, Leu, Met, Phe, Pro, Trp, Val

Polar, uncharged side chain

Asn, Cys, Gln, Ser, Thr, Tyr

Polar charged side chain

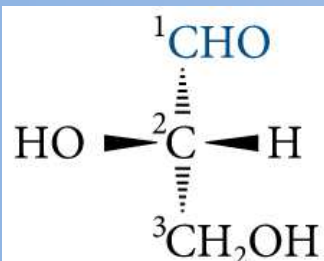
Arg, Asp, Glu, His, Lys

	NONPOLAR, HYDROPHOBIC	R GROUPS	POLAR, UNCHARGED	
Alanine Ala A MW = 89	$\begin{matrix} \text{OOC}^- \\ \\ \text{H}_3\text{N}^+ - \text{CH} - \text{CH}_3 \end{matrix}$		$\text{H} - \text{CH} - \begin{matrix} \text{COO}^- \\ \\ \text{NH}_3^+ \end{matrix}$	Glycine Gly G MW = 75
Valine Val V MW = 117	$\begin{matrix} \text{OOC}^- \\ \\ \text{H}_3\text{N}^+ - \text{CH} - \text{CH} \begin{matrix} / \text{CH}_3 \\ \backslash \text{CH}_3 \end{matrix} \end{matrix}$		$\text{HO} - \text{CH}_2 - \text{CH} - \begin{matrix} \text{COO}^- \\ \\ \text{NH}_3^+ \end{matrix}$	Serine Ser S MW = 106
Leucine Leu L MW = 131	$\begin{matrix} \text{OOC}^- \\ \\ \text{H}_3\text{N}^+ - \text{CH} - \text{CH}_2 - \text{CH} \begin{matrix} / \text{CH}_3 \\ \backslash \text{CH}_3 \end{matrix} \end{matrix}$		$\text{OH} - \text{CH} - \text{CH} - \begin{matrix} \text{COO}^- \\ \\ \text{NH}_3^+ \end{matrix}$	Threonine Thr T MW = 119
Isoleucine Ile I MW = 131	$\begin{matrix} \text{OOC}^- \\ \\ \text{H}_3\text{N}^+ - \text{CH} - \text{CH} \begin{matrix} / \text{CH}_3 \\ \backslash \text{CH}_2 - \text{CH}_3 \end{matrix} \end{matrix}$		$\text{HS} - \text{CH}_2 - \text{CH} - \begin{matrix} \text{COO}^- \\ \\ \text{NH}_3^+ \end{matrix}$	Cysteine Cys C MW = 121
Phenylalanine Phe F MW = 131	$\begin{matrix} \text{OOC}^- \\ \\ \text{H}_3\text{N}^+ - \text{CH} - \text{CH}_2 - \text{C}_6\text{H}_5 \end{matrix}$		$\text{HO} - \text{C}_6\text{H}_4 - \text{CH}_2 - \text{CH} - \begin{matrix} \text{COO}^- \\ \\ \text{NH}_3^+ \end{matrix}$	Tyrosine Tyr Y MW = 181
Tryptophan Trp W MW = 204	$\begin{matrix} \text{OOC}^- \\ \\ \text{H}_3\text{N}^+ - \text{CH} - \text{CH}_2 - \text{C} \begin{matrix} / \text{C}_6\text{H}_4 \\ \backslash \text{N} - \text{H} \end{matrix} \end{matrix}$		$\text{NH}_2 - \text{C} = \text{O} - \text{CH}_2 - \text{CH} - \begin{matrix} \text{COO}^- \\ \\ \text{NH}_3^+ \end{matrix}$	Asparagine Asn N MW = 132
Methionine Met M MW = 149	$\begin{matrix} \text{OOC}^- \\ \\ \text{H}_3\text{N}^+ - \text{CH} - \text{CH}_2 - \text{CH}_2 - \text{S} - \text{CH}_3 \end{matrix}$		$\text{NH}_2 - \text{C} = \text{O} - \text{CH}_2 - \text{CH}_2 - \text{CH} - \begin{matrix} \text{COO}^- \\ \\ \text{NH}_3^+ \end{matrix}$	Glutamine Gln Q MW = 146
Proline Pro P MW = 115	$\begin{matrix} \text{COO}^- \\ \\ \text{CH} - \text{CH}_2 \\ \quad \quad \quad \\ \text{HN} - \text{CH}_2 \quad \quad \quad \text{CH}_2 \end{matrix}$		$\text{NH}_3^+ - \text{CH}_2 - (\text{CH}_2)_3 - \text{CH} - \begin{matrix} \text{COO}^- \\ \\ \text{NH}_3^+ \end{matrix}$	Lysine Lys K MW = 146
Aspartic acid Asp D MW = 133	$\begin{matrix} \text{OOC}^- \\ \\ \text{H}_3\text{N}^+ - \text{CH} - \text{CH}_2 - \text{C} \begin{matrix} / \text{O} \\ \backslash \text{O} \end{matrix} \end{matrix}$		$\text{NH}_2 - \text{C} = \text{NH} - (\text{CH}_2)_3 - \text{CH} - \begin{matrix} \text{COO}^- \\ \\ \text{NH}_3^+ \end{matrix}$	Arginine Arg R MW = 174
Glutamine acid Glu E MW = 147	$\begin{matrix} \text{OOC}^- \\ \\ \text{H}_3\text{N}^+ - \text{CH} - \text{CH}_2 - \text{CH}_2 - \text{C} \begin{matrix} / \text{O} \\ \backslash \text{O} \end{matrix} \end{matrix}$		$\text{HN} = \text{C} - \text{CH}_2 - \text{CH} - \begin{matrix} \text{COO}^- \\ \\ \text{NH}_3^+ \end{matrix}$	Histidine His H MW = 155

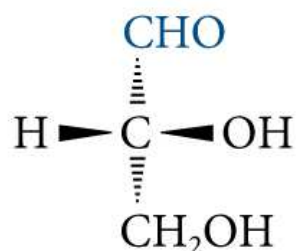
1.2. The amino acids

1.2.3. General properties of the amino acids

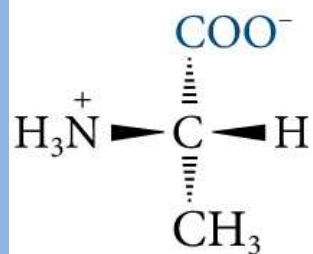
1.2.3.1. Stereochemistry



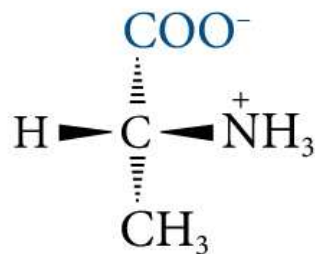
L-Glyceraldehyde



D-Glyceraldehyde



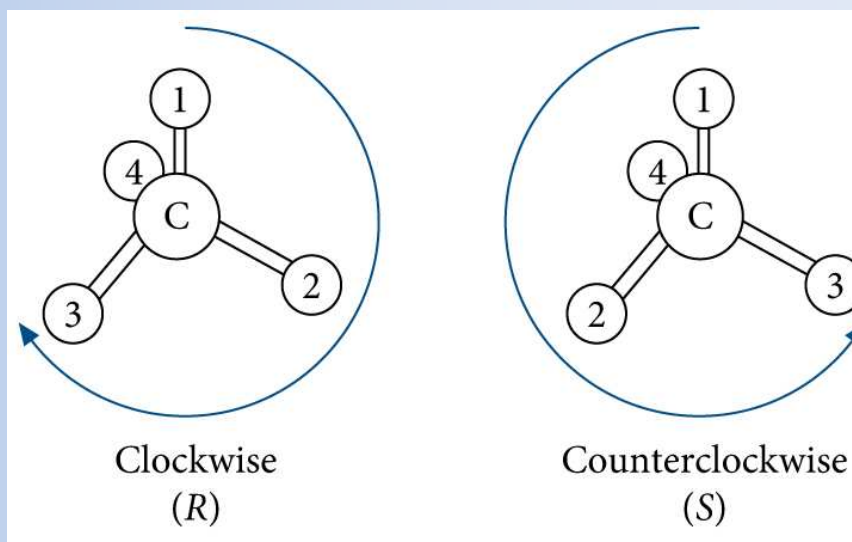
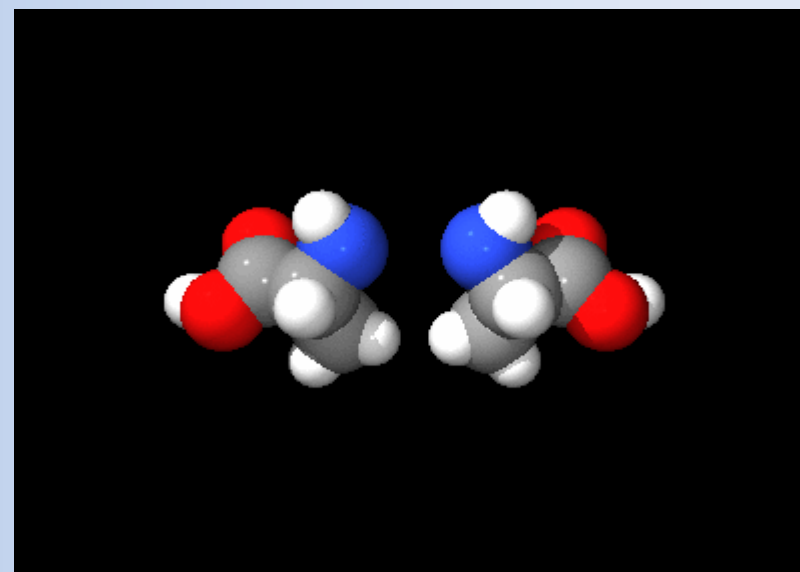
L-Alanine



D-Alanine

Left – L/S

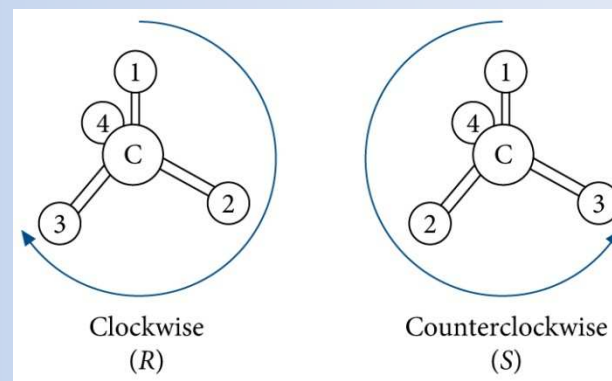
Right – D/R



1.2. The amino acids

1.2.3. General properties of the amino acids

1.2.3.1. Stereochemistry



The Sequence Rule for Assignment of Configurations to Chiral Centers.

Assign sequence priorities to the four substituents by looking at the atoms attached directly to the chiral center.

1. The higher the atomic number of the immediate substituent atom, the higher the priority.

For example, $\text{H-} < \text{C-} < \text{N-} < \text{O-} < \text{Cl-}$.

2. If two substituents have the same immediate substituent atom, evaluate atoms progressively further away from the chiral center until a difference is found.

For example, $\text{CH}_3\text{-} < \text{C}_2\text{H}_5\text{-} < \text{ClCH}_2\text{-} < \text{BrCH}_2\text{-} < \text{CH}_3\text{O-}$.

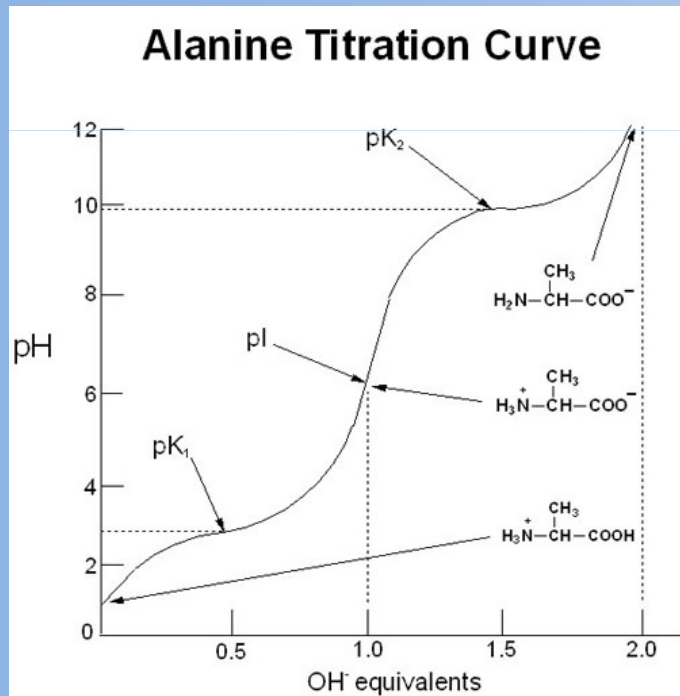
3. If double or triple bonded groups are encountered as substituents, they are treated as an equivalent set of single-bonded atoms.

For example, $\text{C}_2\text{H}_5\text{-} < \text{CH}_2=\text{CH-} < \text{HC}\equiv\text{C-}$

1.2. The amino acids

1.2.3. General properties of the amino acids

1.2.3.2. Ionization



Amino Acid	Symbol	pK ₁ (COOH)	pK ₂ (NH ₂)	pK R Group
Glycine	Gly	2,4	9,8	
Alanine	Ala	2,4	9,9	
Valine	Val	2,2	9,7	
Leucine	Leu	2,3	9,7	
Isoleucine	Ile	2,3	9,8	
Serine	Ser	2,2	9,2	
Threonine	Thr	2,1	9,1	
Cysteine	Cys	1,9	10,8	8,3
Methionine	Met	2,1	9,3	
Aspartic Acid	Asp	2	9,9	3,9
Glutamic Acid	Glu	2,1	9,5	4,1
Asparagine	Asn	2,1	8,8	
Glutamine	Gln	2,2	9,1	
Arginine	Arg	1,8	9	12,5
Lysine	Lys	2,2	9,2	10,8
Histidine	His	1,8	9,2	6
Phenylalanine	Phe	2,2	9,2	
Tyrosine	Tyr	2,2	9,1	10,1
Tryptophan	Trp	2,4	9,4	
Proline	Pro	2	10,6	

1.2. The amino acids

1.2.3. General properties of the amino acids

1.2.3.3. Hydrophobicity (Engelman et al 1986)

Amino acid		Transfer free energy kJ/mol
Phe	F	15,5
Met	M	14,2
Ile	I	13
Leu	L	11,7
Val	V	10,9
Cys	C	8,4
Trp	W	7,9
Ala	A	6,7
Thr	T	5
Gly	G	4,2
Ser	S	2,5
Pro	P	-0,8
Tyr	Y	-2,9
His	H	-12,5
Gln	Q	-17,1
Asn	N	-20,1
Glu	E	-34,3
Lys	K	-36,8
Asp	D	-38,5
Arg	R	-51,4

Kyte and Doolittle (1)	Rose, et al (2)	Wolfenden, et al (3)	Janin (1979) (4)
Ile Val	Cys	Gly,Leu,Ile Val,ala	Cys Ile Val
Leu	Phe,Ile Val Leu,Met,Trp	Phe Cys Met	Leu,Phe Met Ala,Gly,Trp
Phe Cys Met,Ala	His Tyr Ala Gly Thr	Thr,Ser Trp,Tyr	His,Ser Thr Pro Tyr Asn
Gly Thr,Ser Trp,Tyr Pro		Asp,Lys,Gln Glu,His Asp	Asp Gln,Glu
His Asn,Gln Asp,Glu Lys	Ser Pro,Arg Asn Gln,Asp,Glu		Arg
Arg	Lys	Arg	Lys

1.2. The amino acids

1.2.4. Chemical characteristic of the amino acids

1.2.4.1. Aliphatic side chains

Ala, Gly, Ile, Leu Val

1.2.4.2. Aromatic side chains

Phe, Tyr, Trp

1.2.4.3. Basic side chains

Arg, Lys

1.2.4.4. Acidic side chains

Asp, Glu

1.2.4.5. Hydroxyl side chains

Ser, Thr

1.2.4.6. Amide side chains

Asn, Gln

1.2.4.7. Sulphur-containing side chains

Met, Cys

1.2.4.8. Proline

Pro

1.2.4.9. Amid side chains

Histidine

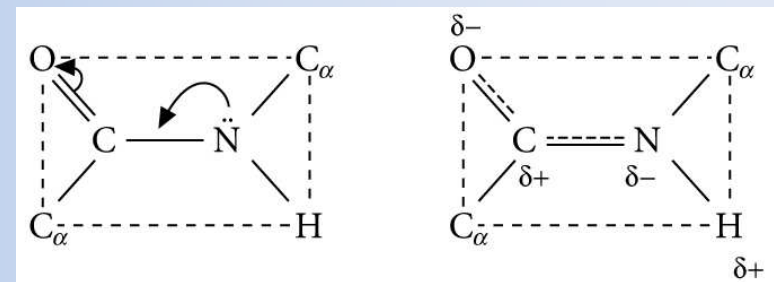
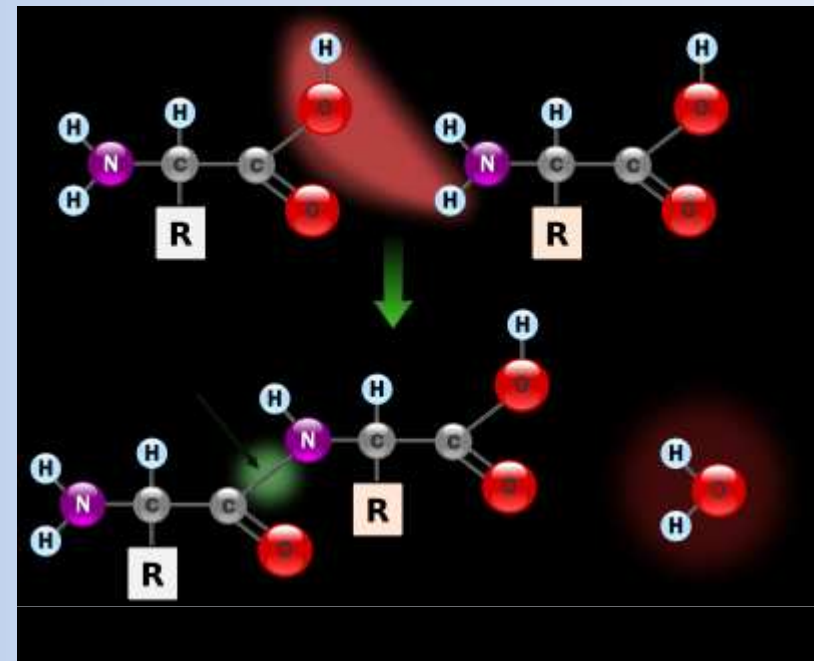
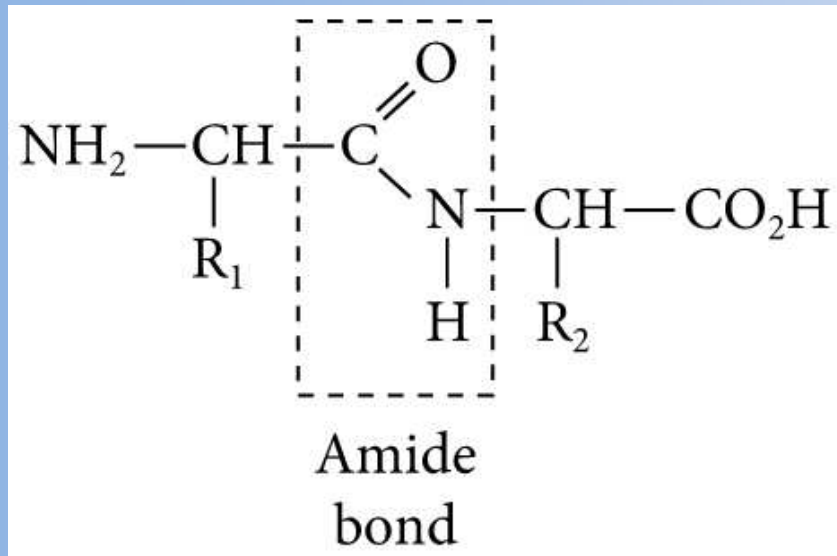
1.2. The amino acids

1.2.4. Chemical characteristic of the amino acids

Amino Acid	Symbol	One-letter	Mass wo H ₂ O	pK _a R Group	Frequency of occurrence %
Alanine	Ala	A	71,08		7,83
Arginine	Arg	R	156,19	12,5	5,35
Asparagine	Asn	N	114,1		4,18
Aspartic Acid	Asp	D	115,09	3,9	5,32
Cysteine	Cys	C	103,14	8,3	1,52
Glutamic Acid	Glu	E	128,13	4,1	3,95
Glutamine	Gln	Q	129,12		6,64
Glycine	Gly	G	57,05		6,93
Histidine	His	H	137,14	6	2,29
Isoleucine	Ile	I	113,16		5,91
Leucine	Leu	L	113,16		9,64
Lysine	Lys	K	128,17	10,8	5,93
Methionine	Met	M	131,2		2,38
Phenylalanine	Phe	F	147,18		4
Proline	Pro	P	97,12		4,83
Serine	Ser	S	87,08		6,86
Threonine	Thr	T	101,11		5,42
Tryptophan	Trp	W	186,21		1,15
Tyrosine	Tyr	Y	163,18	10,1	3,06
Valine	Val	V	99,13		6,71

1.3. The primary structure of proteins

1.3.1. The peptide bond



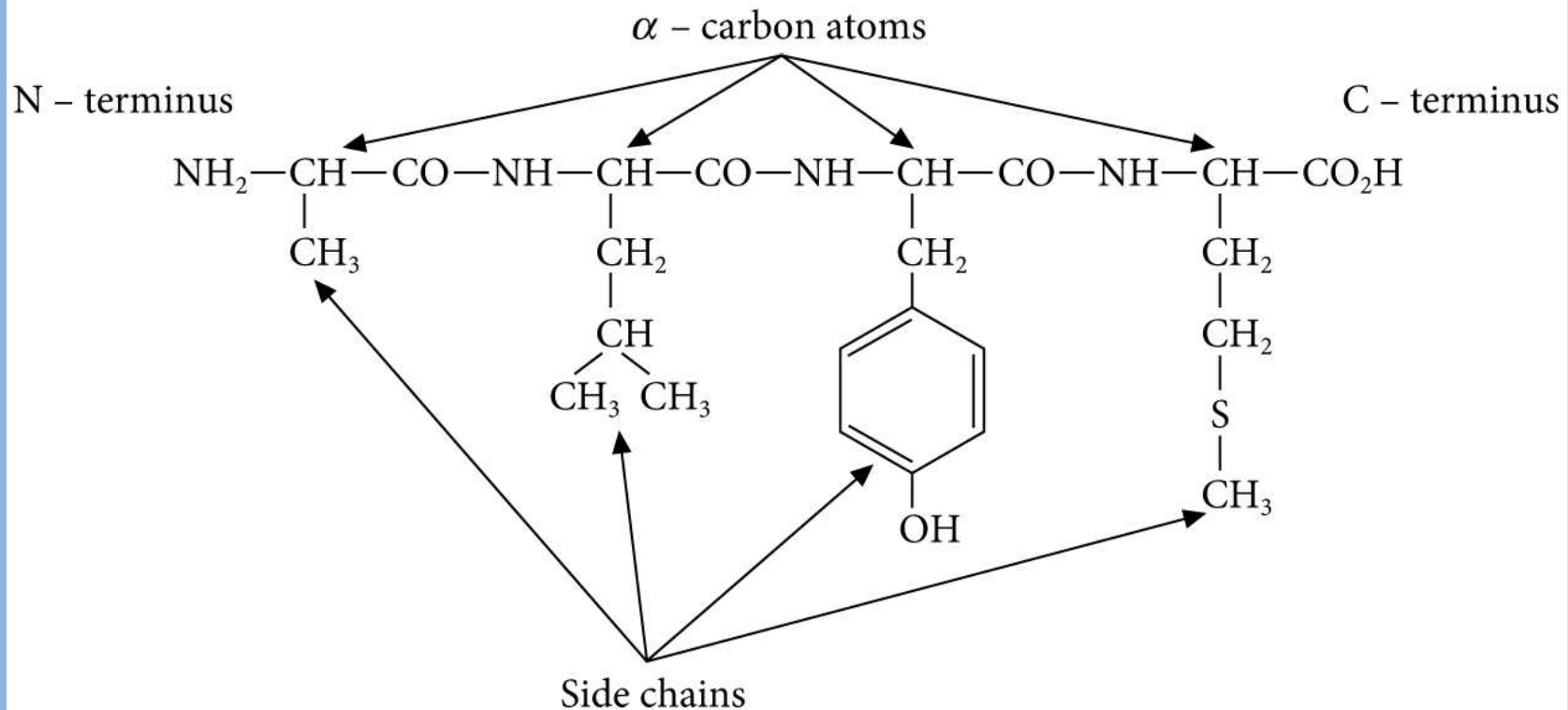
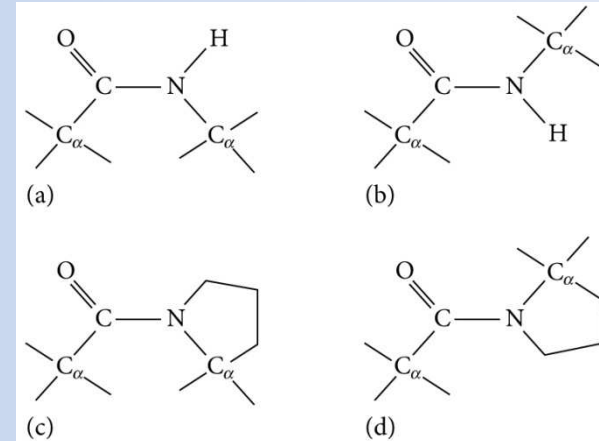
Resonance stabilization of the peptide bond

The *cis* and *trans* forms of the peptide bond

1.3. The primary structure of proteins

1.3.1. The peptide bond

Structure of a tetrapeptide showing the alpha-carbon atoms, side chains, and N and C-termini.



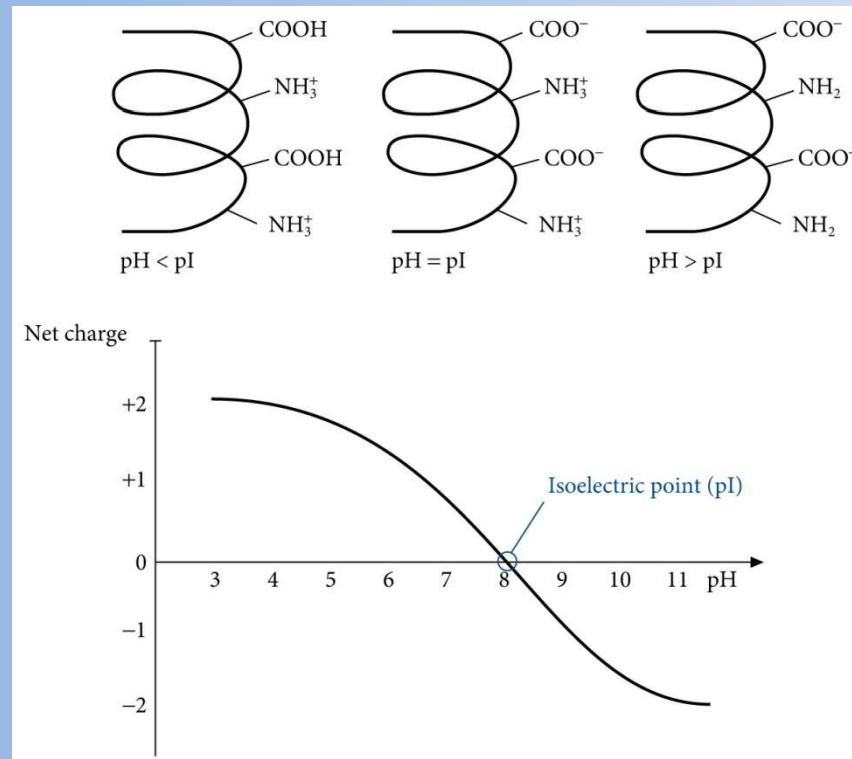
1.3. The primary structure of proteins

1.3.2. Information available from the amino acid sequence of a protein

1.3.2.1. Exact molecular mass

1.3.2.2. Isoelectric point

http://www.expasy.ch/tools/pi_tool.html



1.3. The primary structure of proteins

1.3.2. Information available from the amino acid sequence of a protein

1.3.2.3. Absorption coefficient

1.3.2.4. Hydrofobicity

1.3.2.5. Post-translational modifications

Type of modification

Possible effect on function

Proteolysis

Removal of targeting sequences.

Generation of several new products (hormones).

Activation of proteins (enzymes).

Disulphide bond formation

Stabilization of structure of secreted proteins.

Hydroxylation

Formation of hydroxy -Lys or -Pro increases the stability of the triple helix of collagen.

Glycosylation

Many cell surface proteins are involved in cell - cell recognition.

Attachment of glycosyl-phosphatidylinositol groups anchors proteins to membrane.

The polar nature of proteins can be enhanced.

Phosphorylation

Phosphorylation of Ser, Thr, Tyr, His, Asp side chains can regulate the activity of proteins, especially in signalling pathways.

N-terminal acylation

Attachment of C₁₄ (myristoylation) or C₁₆ (palmitoylation) chains will enhance the association of the protein with membranes.

1.3. The primary structure of proteins

1.3.2. Information available from the amino acid sequence of a protein

1.3.2.6. Structural and functional motifs

• Transmembrane domains

• Targeting sequences

-S-K-L peroxisomes

-K-D-E-L endoplasmic
reticulum

• Metal binding

-C-X₄-C-X₂-C- Fe binding

• Glycosylation sites

-N-X-S/T-

• Phosphorylation sites

-R-X₁₋₂-S/T- Protein kinase A
-R-R-X-S/T-

• Nitrosylation

-(G,S,T,C,Y,N,Q)-(K,R,H,D,E)-C-(D,E)

>AHK4/CRE1 Receptor histidine kinase

MNWALNNHQEEEEPRRIEISDSELENLKSSDFYQLGGGGALNSSEKPRKID
FWRSGLMGF AKMQQQQQLQHSVAVKMNNNNNDLMGNKKGSTFIQEHRALLPK
ALILWIIIVGFISSGIYQWMDANKIRREEVLVSMCDQRARMLQDQFSVSVNH
VHALAILVSTFHYHKNPSAIDQETFAEYTARTAFERPLLSGVAYA EKVVNFER
EMFERQHNWVIKTMDRGEPSPVRDEYAPVIFSQDSVSYLES LDMMSGEEDREN
ILRARETGKAVLTSPFRLLETHHLGVVLTFFVYKSSLPENPTVEERIAATAGY
LGGAFDVESLVENLLGQLAGNQAIVVHVYDITNASDPLVMYGNQDEEADRSL
HESKLD FGD PFRKHKMICRYHQKAP IPLNVLTTPVPLFFAIGFLVGYILYGAAM
HIVKVEDDFHEMQELKVRAEAADVAK

1.3. The primary structure of proteins

1.3.2. Information available from the amino acid sequence of a protein

1.3.2.7. Sequence relationships between proteins

Orthologous protein	different species, same function (beta-glucosidase, <i>Zea mays</i> and <i>Brassica napus</i>)
Paralogous protein	same species, different function histidine kinase – ETR1 (ethylene signalling pathway) histidine kinase – AHK4 (cytokinin signalling pathway)
Segmented proteins	independent folded and functional unit AHK4 - extracellular CHASE domain - histidine kinase domain - receiver domain

Please solve a problem.

Question 1: I am an amino acid.

My name in Greek means „sweet“. 5 points

I belong to the group of amino acids with non-polar side chains. 3 points

I am frequently found in secondary structures called loops. 2 points

I am the smallest amino acid. 1 point

Glycin

Please solve a problem.

Question 2: I am an amino acid.

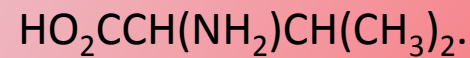
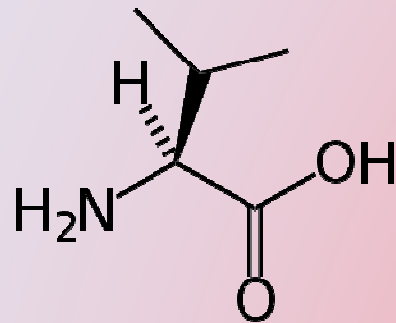
I belong to the group of five amino acids with aliphatic side chains.

My molecular mass is in the middle of this group.

5 points

My structure contains five carbon atoms.

3 points



2 points

I am not Gly, Ala, Ile or Leu.

1 point

Valine

Please solve a problem.

Question 3: I am an amino acid.

My name in Greek means "silk". 5 points

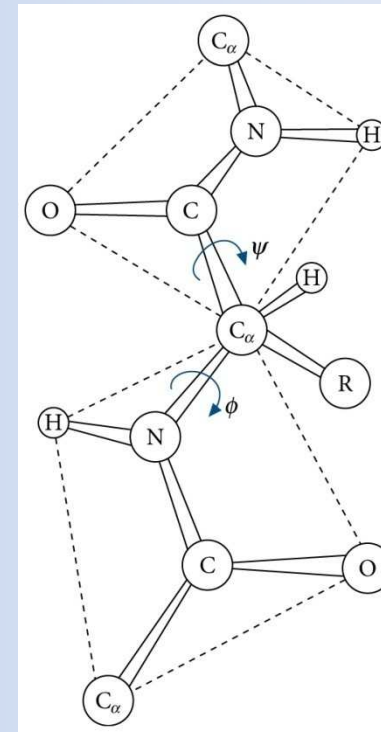
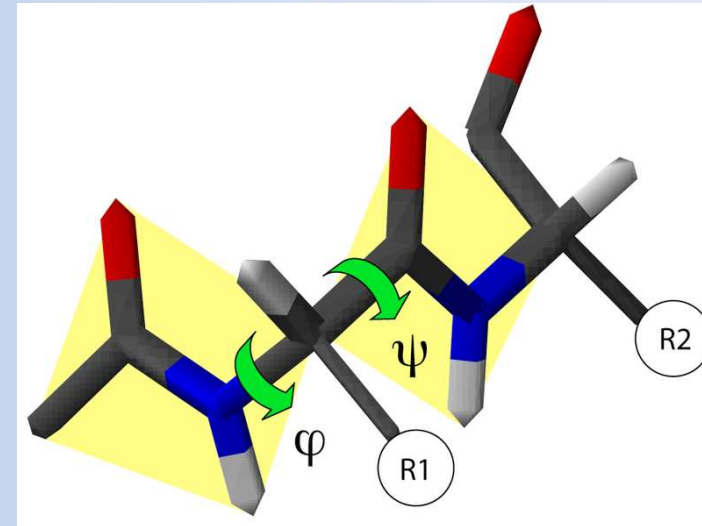
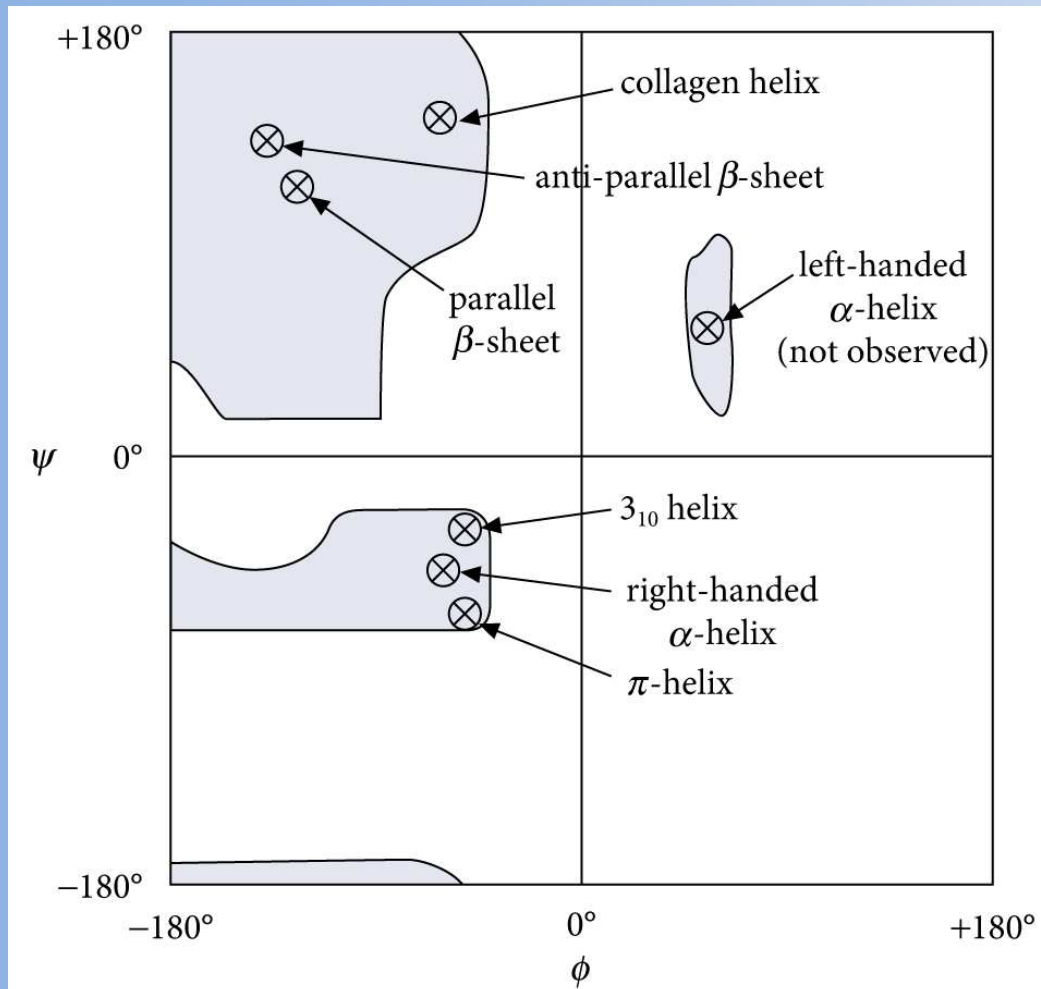
I belong to the group of polar, uncharged side chain amino acids.
I am not an aromatic amino acid. I contain a hydroxyl group and
my MW is less than 100 Da. 3 points

I am very often phosphorylated 2 points

I am related to threonine. 1 points

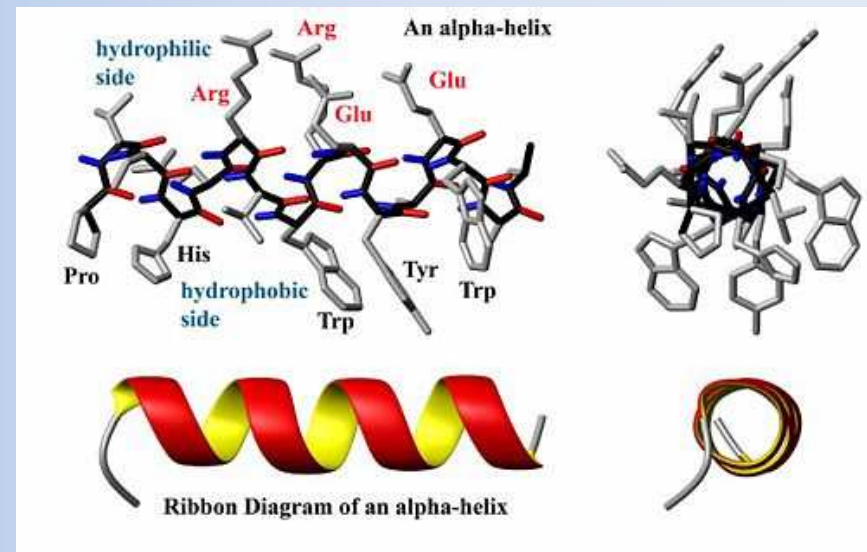
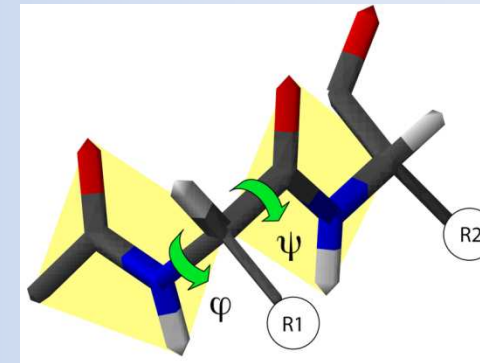
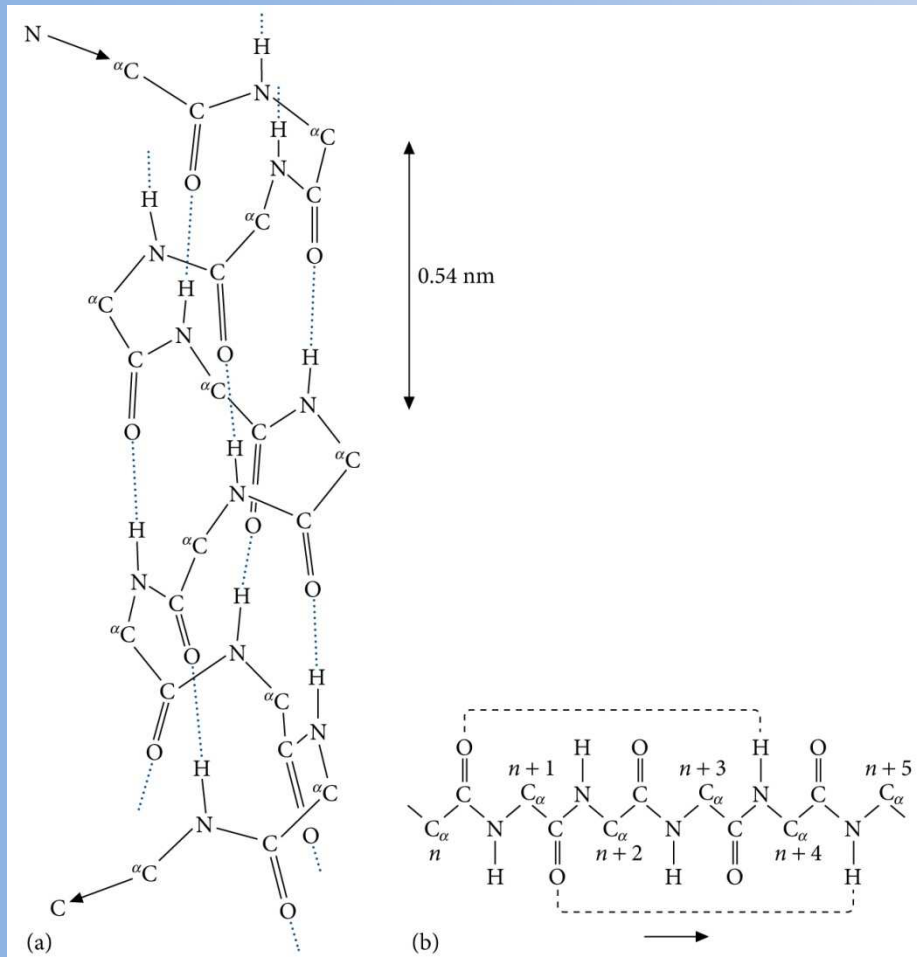
Serine

1.4. The secondary structure of proteins



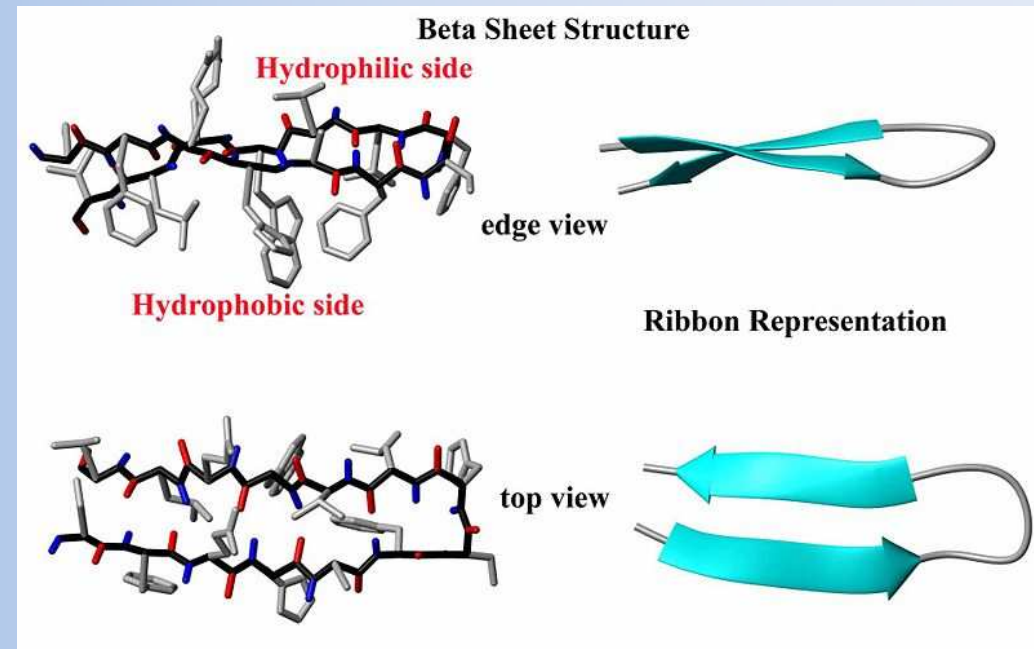
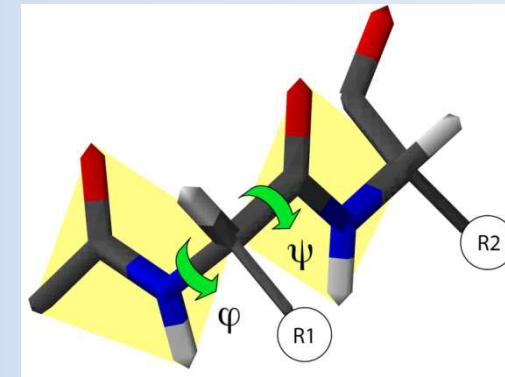
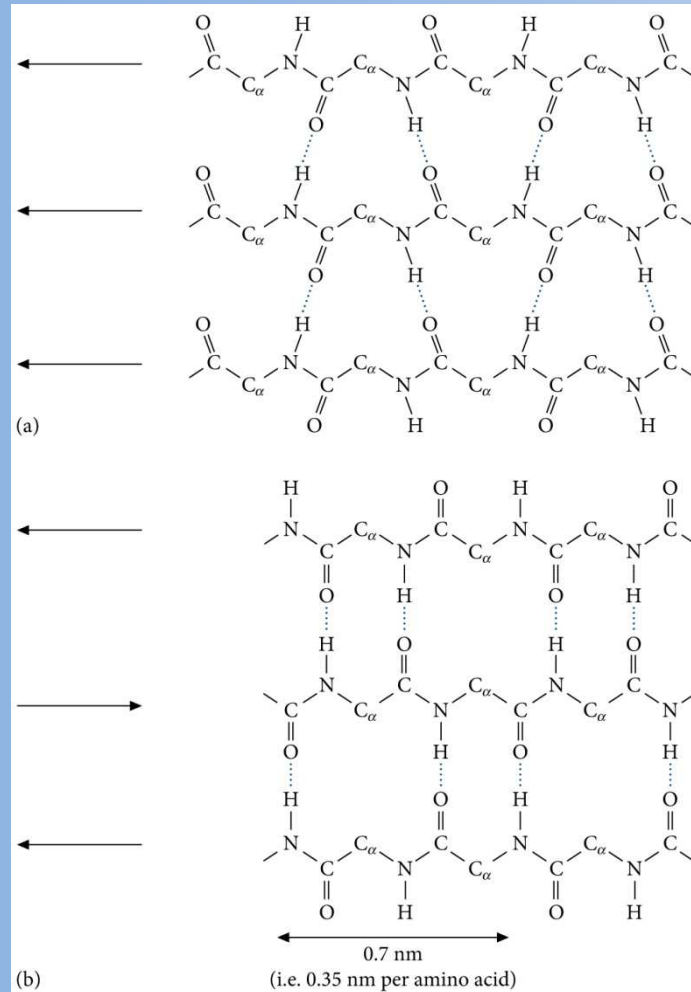
1.4. The secondary structure of proteins

1.4.1. The alpha-helix



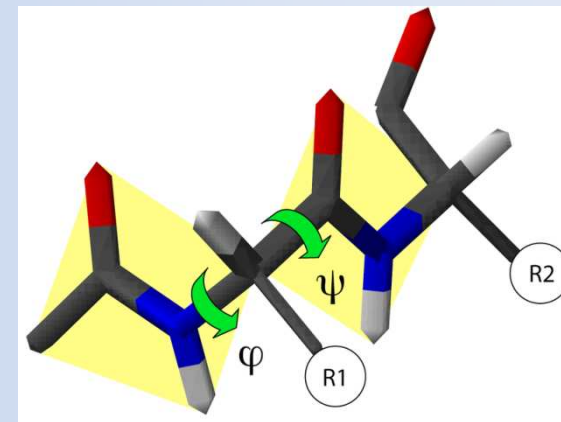
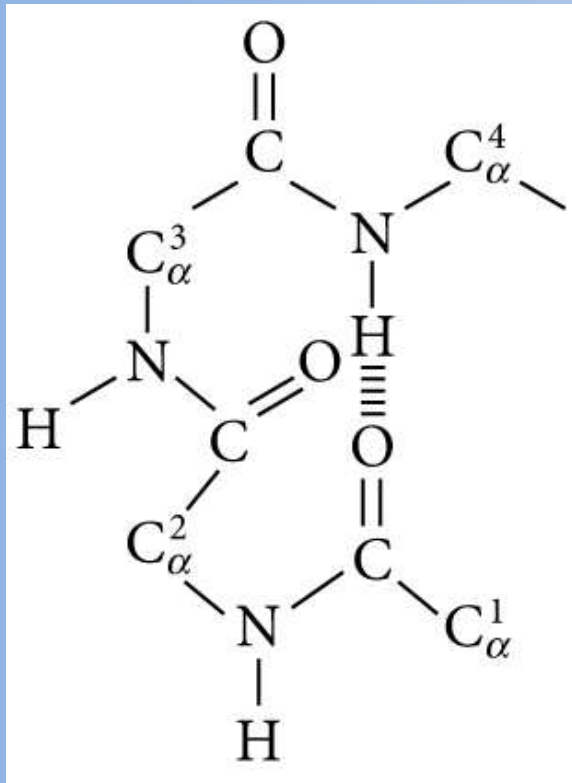
1.4. The secondary structure of proteins

1.4.2. The beta-strand



1.4. The secondary structure of proteins

1.4.3. Other structural features in proteins



Structural preferences of the different amino acids

Met, Glu, Leu, Ala	+	alpha - helices
Pro, Gly, Tyr	-	alpha helices
Val, Ile, Phe	+	beta sheets
Pro, Asp	-	beta sheets
Pro, Gly, Asp	+	beta turn
Met, Val, Ile	-	beta turn

1.5. The tertiary structure of proteins

1.5.1. General principles

Close packing

Elements of secondary structure

Distribution of side chains

Pairing of polar group

Formation of domains

Average Conformational Parameters of Helical Elements

Conformation	Phi	Psi	Omega	Residues per turn	Translation per residue
Alpha helix	-57	-47	180	3.6	1.5
3-10 helix	-49	-26	180	3.0	2.0
Pi-helix	57	-70	180	4.4	1.15
Polyproline I	-83	+158	0	3.33	1.9
Polyproline II	-78	+149	180	3.0	3.12
Polyproline III	-80	+150	180	3.0	3.1

1.5. The tertiary structure of proteins

1.5.2. Classification of protein structures

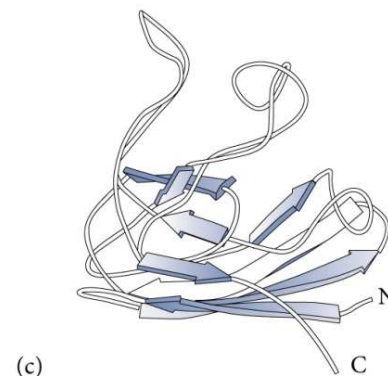
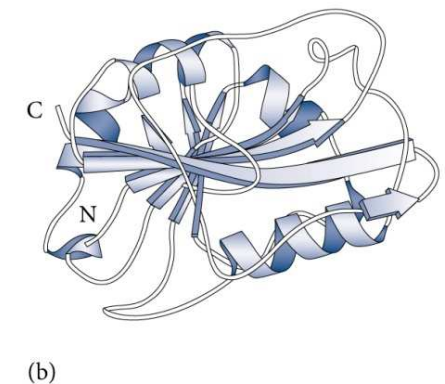
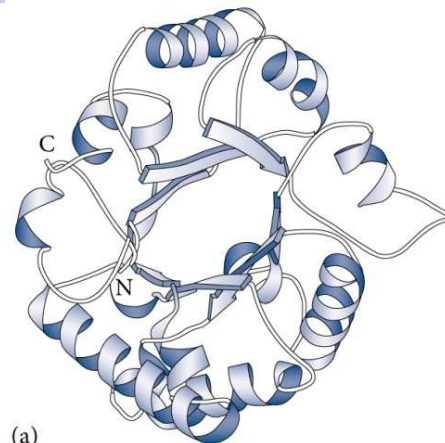
1.5.3. Forces involved in stabilizing tertiary structure

SCOP Classification Statistics

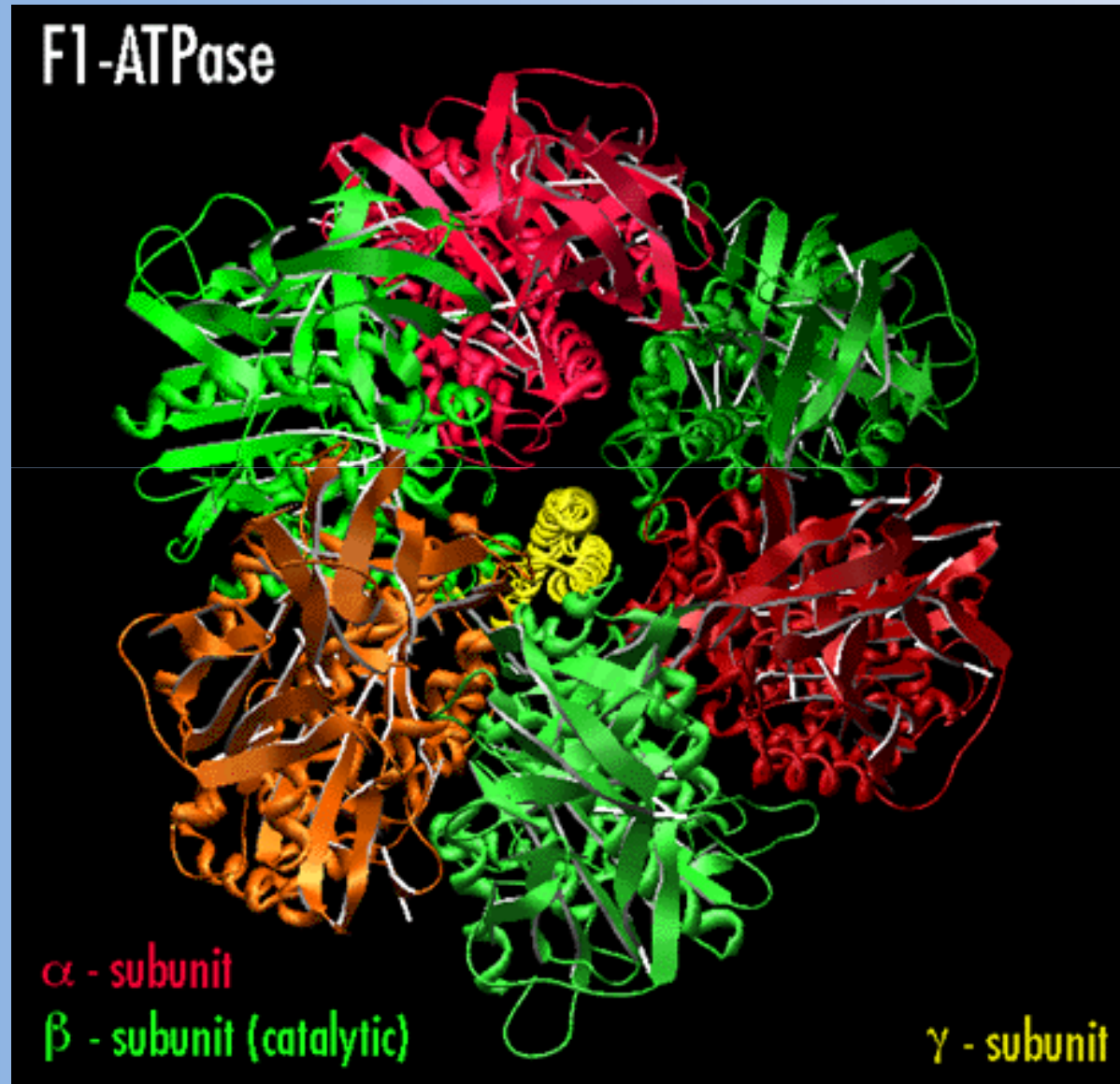
SCOP: Structural Classification of Proteins. **1.75** release

38221 PDB Entries (23 Feb 2009). **110800** Domains. 1 Literature Reference
(excluding nucleic acids and theoretical models)

Class	Number of folds	Number of superfamilies	Number of families
All alpha proteins	284	507	871
All beta proteins	174	354	742
Alpha and beta proteins (a/b)	147	244	803
Alpha and beta proteins (a+b)	376	552	1055
Multi-domain proteins	66	66	89
Membrane and cell surface proteins	58	110	123
Small proteins	90	129	219
Total	1195	1962	3902

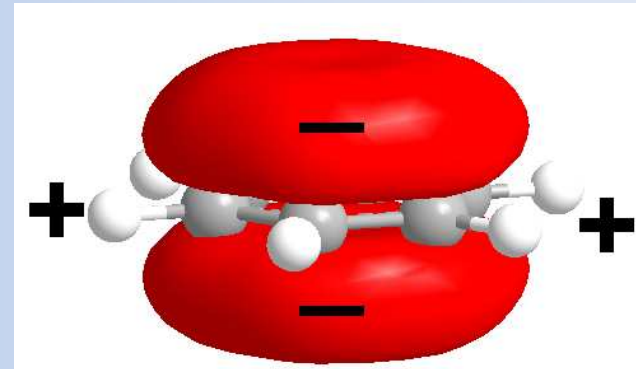
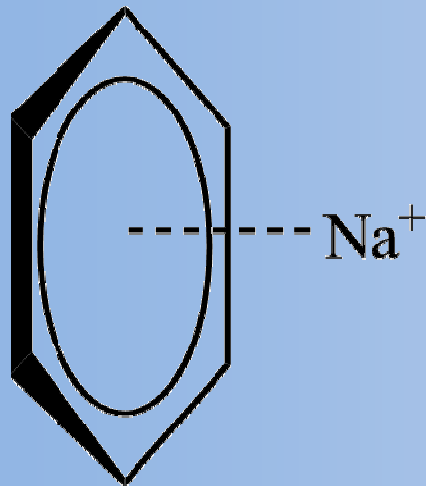


1.6. The quaternary structure of proteins

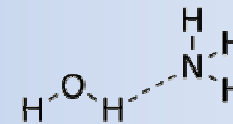
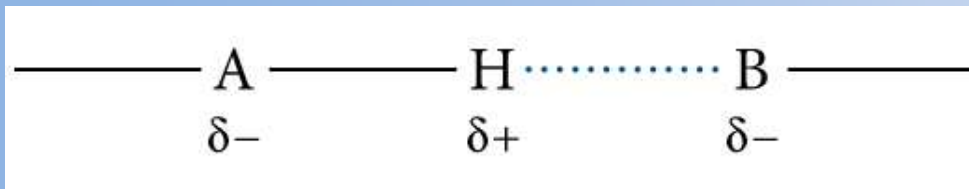


1.7. Forces contributing to the structures and interactions of proteins

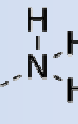
1.7.1. Ionic (electrostatic) interactions



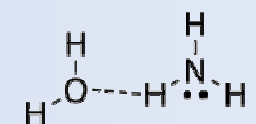
1.7.2. Hydrogen bonds



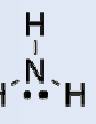
hydrogen
bond
donor



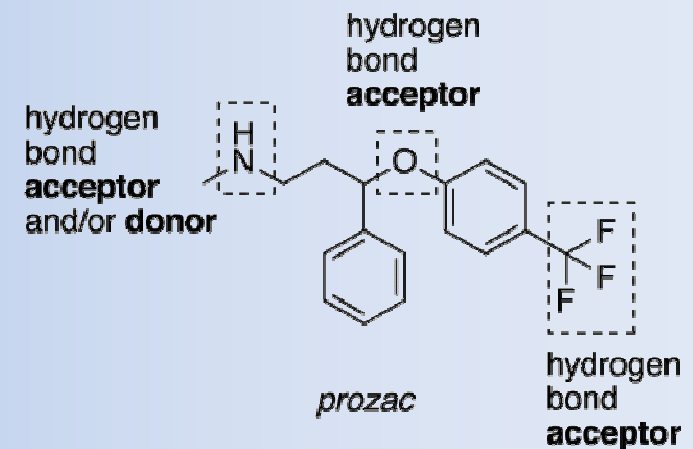
hydrogen
bond
acceptor



hydrogen
bond
acceptor



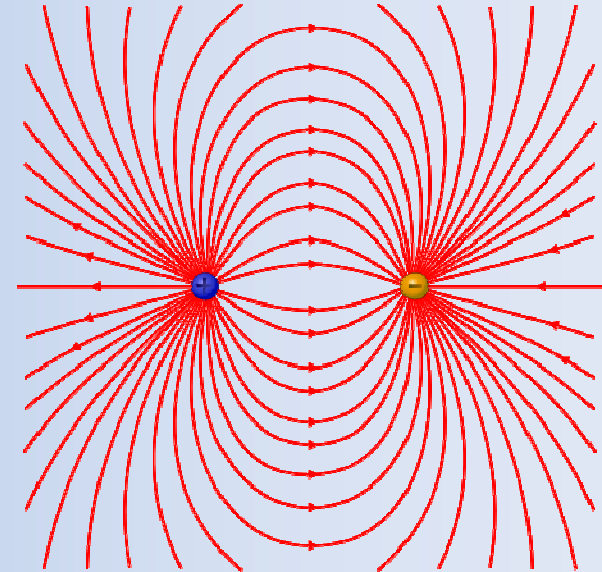
hydrogen
bond
donor



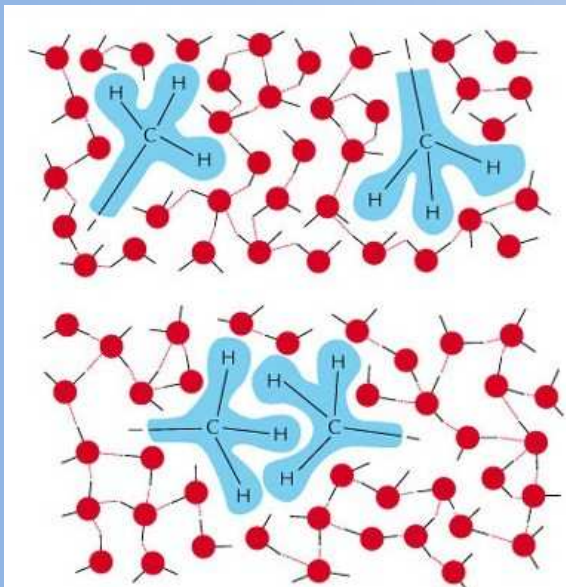
1.7. Forces contributing to the structures and interactions of proteins

1.7.3. Van der Waals' interactions

Many molecules have such dipole moments due to non-uniform distributions of positive and negative charges on the various atoms. Such is the case with polar compounds like hydroxide (OH^-), where electron density is shared unequally between atoms.



1.7.4. Hydrophobic interactions



1.7. Forces contributing to the structures and interactions of proteins

1.7.5. Balance of energy contributions

Balance between an unfavourable enthalpy term but a favourable entropy.

In proteins, the folded state is in the range 20-60 kJ/mol.

Entropy and enthalpy are in the range of several hundred kJ/mol.

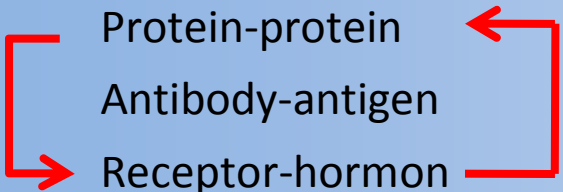
GnHCl and urea weaken hydrophobic interaction and promote the unfolding of proteins.

<http://fikus.omsk.cz/~bojkovsm/termodynamika/vdws.html>

1.7.6. The range of energies involved in protein interactions

$$-\Delta G^\circ = RT \ln K_{eq}$$

Interaction	Typical dissociation constant (Kd) (M)	ΔG° (kJ/mol)
Avidin-biotin	10^{-15}	89
Protein-protein	10^{-10}	59
Antibody-antigen	10^{-9}	53
Receptor-hormon	10^{-7}	42
Enzyme-substrate	10^{-5}	30



Please solve a problem.

Question 1: I am a secondary structure.

LSFAAAMIGLA

5 points

The average length of the structure is 11-12 amino acids.

3 points

In general I need 3.6 amino acids to be in the same position

2 points

This structure is not a beta strand

1 point

alpha-helices

Please solve a problem.

Question 2: I am a weak forces.

I need to have two atoms – one donor and one acceptor. 5 points

This interaction is found between particular side chain and main chain atoms (N-H or C=O groups of the peptide bond). 3 points

I very often interact with water. 2 points

This interaction is also found between water molecules, where five atoms are arranged in a tetrahedral structure. 1 point

Hydrogen bond

Please solve a problem.

Question 3: I am a 3D structure protein.

According to structural classification of proteins, I belong to the smallest family with a homogenous secondary structure. 5 points

„Greek key“ structure represents a typical structure. 3 points

My structure most frequently contains Val, Ile and Phe. 2 points

The words "sheet" and "strand" are used with this Greek character. 1 point

Beta proteins