Sample in structural biology

Josef Houser

Autumn 2023

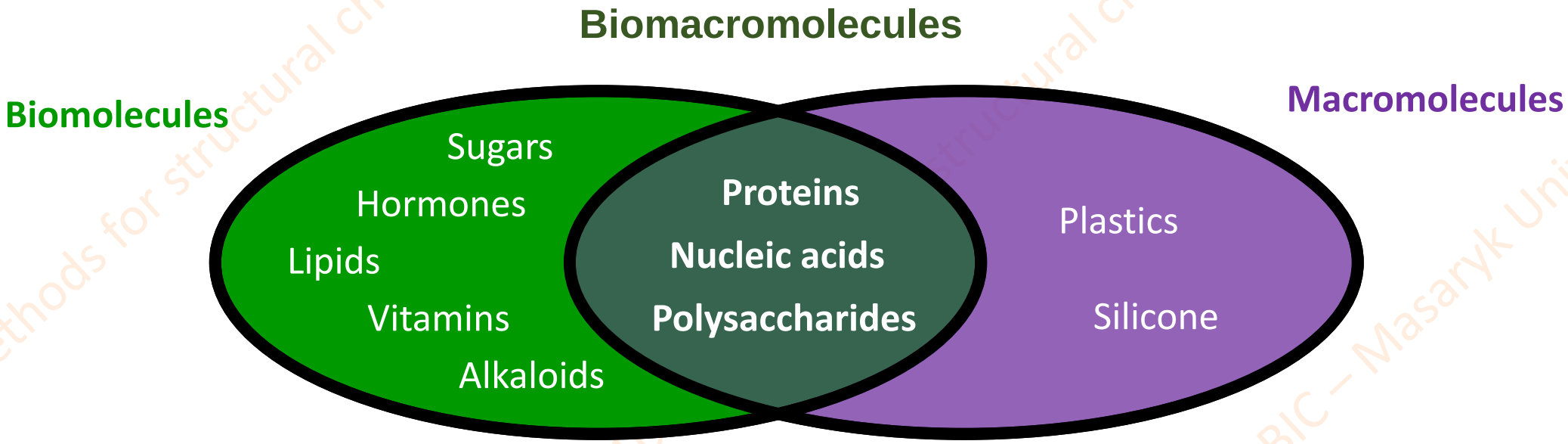
S1004 Methods for structural characterization of biomolecules

Biomacromolecules

Biomolecules are natural parts of living organisms.

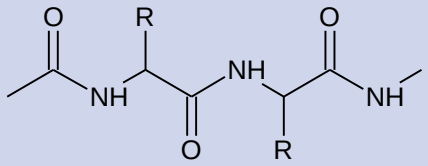
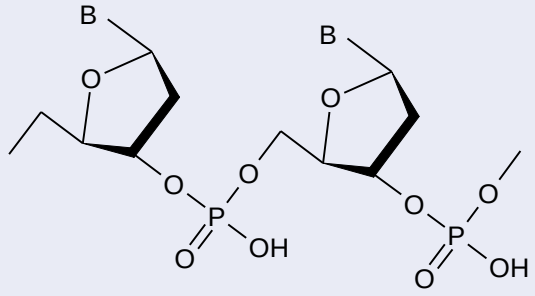
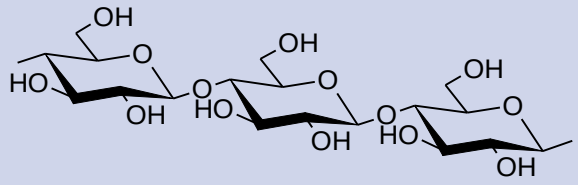
Macromolecules typically compose of thousands to millions of atoms. Small molecules compose of hundreds of atoms or less.

Molecules are essential parts of matter. They consist of atoms that are linked through covalent bonds.

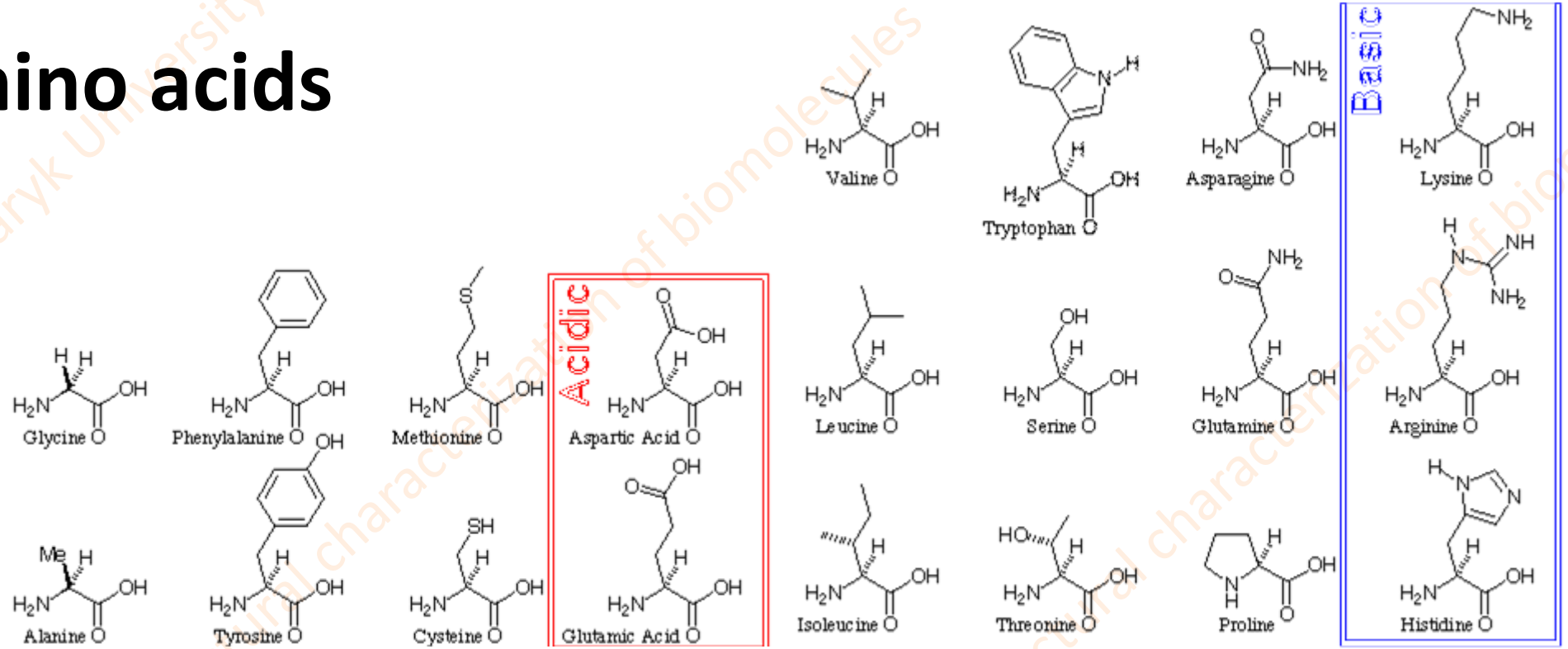


Basic chemical composition of biomacromolecules

- Heteropolymers consisting of various subunits

Macromolecule	Building blocks	Type of bond	Scheme
Protein	Amino acids	Peptidic	
Nucleic acid	Nukleotides	Esteric	
Polysaccharide	Monosaccharides	Glycosidic	

Amino acids



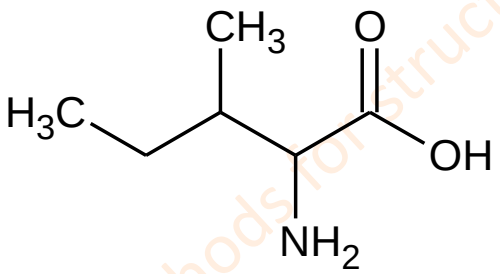
Glycine	Alanine	Valine	Leucin	Isoleucine	Aspartic acid	Asparagine	Glutamic acid	Glutamine	Arginine	Lysine	Histidine	Phenylalanine	Serine	Threonine	Tyrosine	Tryptophan	Methionine	Cysteine	Proline	Selenocysteine	Pyrolysine
Gly	Ala	Val	Leu	Ile	Asp	Asn	Glu	Gln	Arg	Lys	His	Phe	Ser	Thr	Tyr	Trp	Met	Cys	Pro	Sec	Pyr
G	**A**	**V**	**L**	**I**	**D**	**N**	**E**	**Q**	**R**	**K**	**H**	**F**	**S**	**T**	**Y**	**W**	**M**	**C**	**P**	**U**	**O**

J	**B**	**Z**

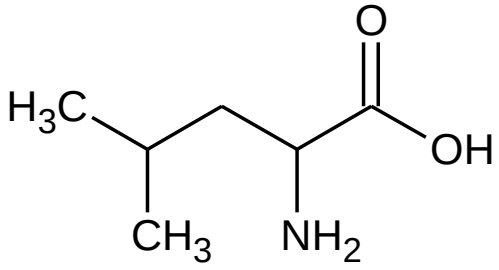
X - Any

Types of amino acids

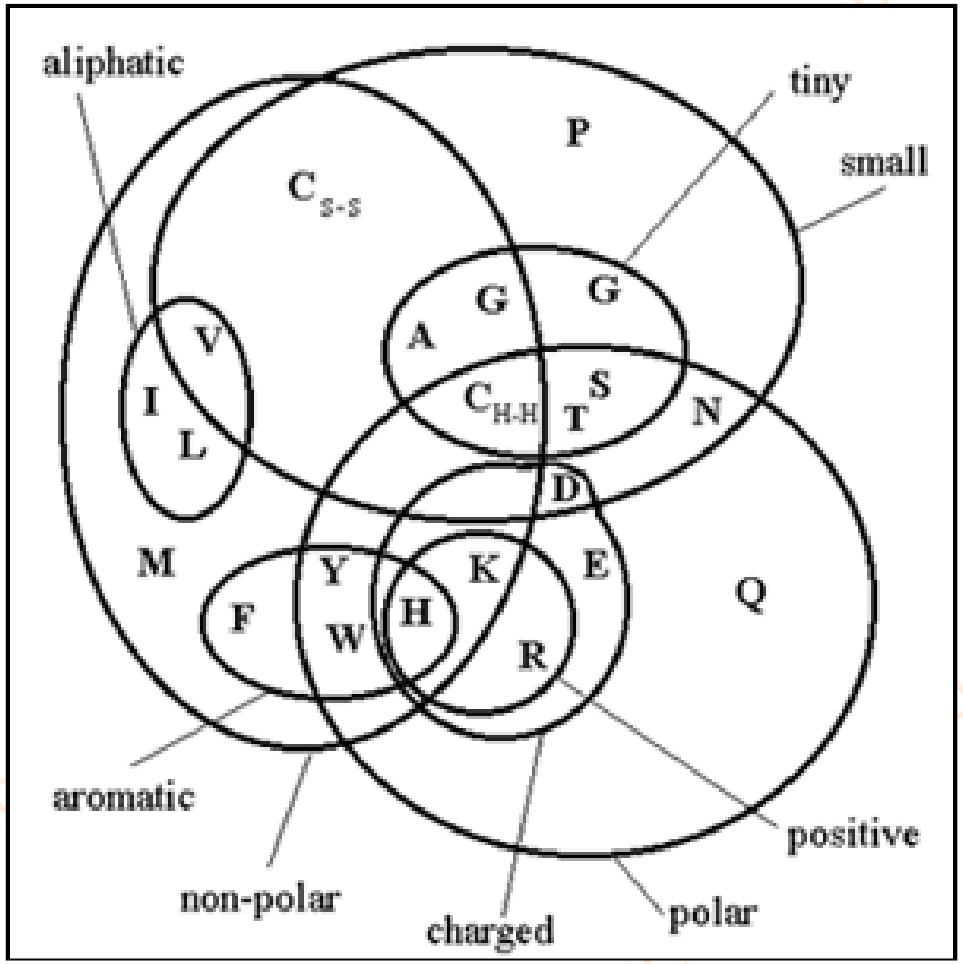
Amino acids with similar properties may substitute each other in protein



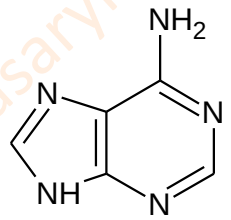
Isoleucine



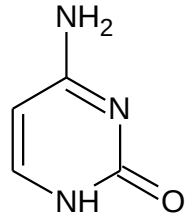
Leucine



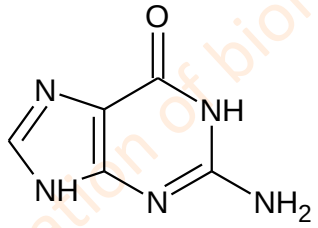
Nucleic bases



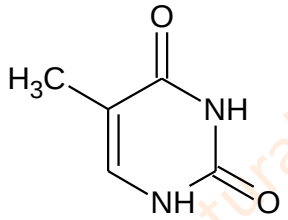
Adenine



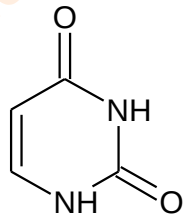
Cytosine



Guanine



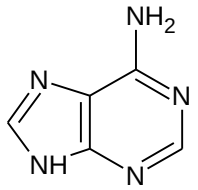
Thymine



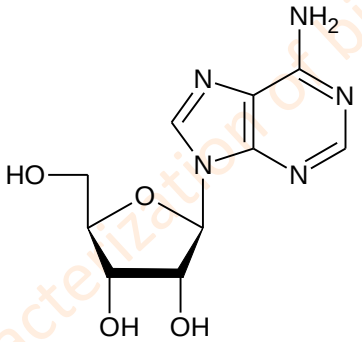
Uracil

adenine	cytosine	guanine	thymine	uracil
A	C	G	T	U

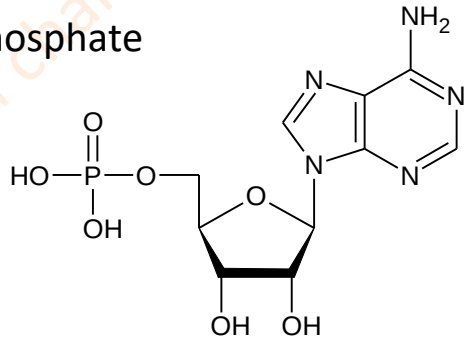
Nucleic base
Adenine



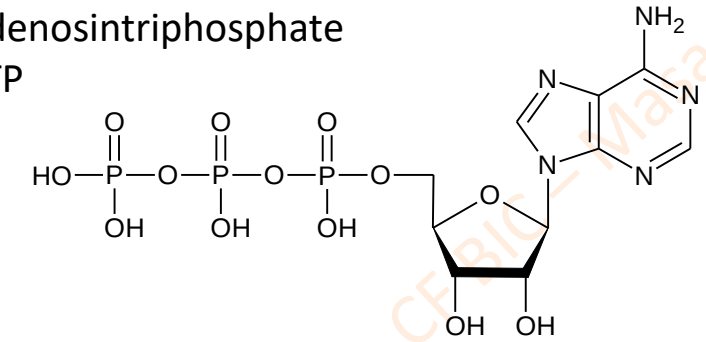
Nucleoside
Adenosine



Nucleotide
Adenosinmonophosphate
AMP

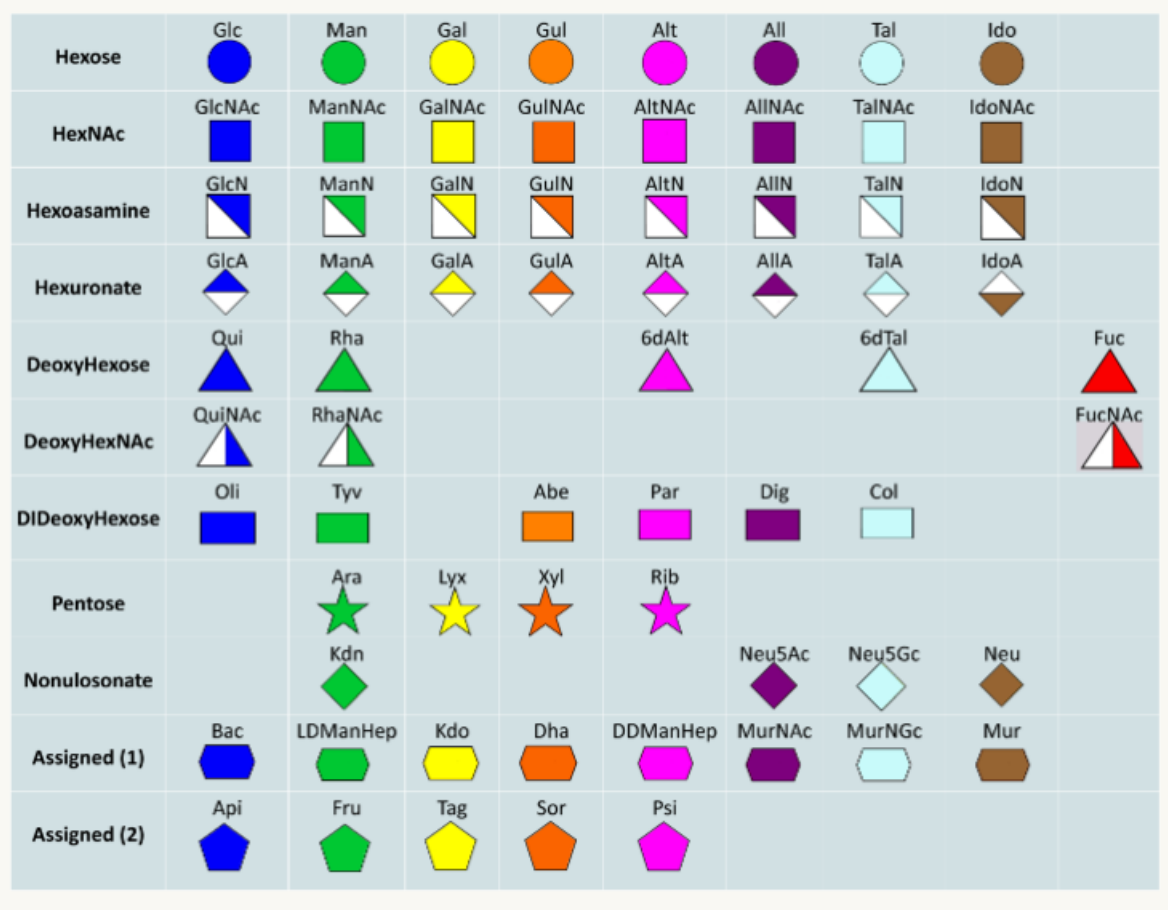


Nucleotide
Adenosintriphosphate
ATP



Polysaccharides

App. 150 different saccharides identified as polysaccharide building units – graphical code



Combinatorics

Structural variability reflects length, variability of units and variability of bonds

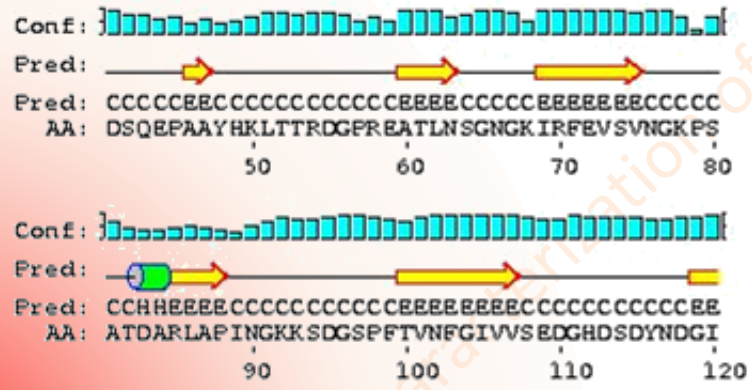
Polymer	Protein	Nucleic acid	Polysaccharide
Number of various basic units	20 (22)	4 (DNA) 4 (RNA)	150 (identified)
Number of possible bond types	1	1	2 x 4 (for hexose)
Theoretical number of possible molecules consisting of 2 units	22 x 22 = 484	4 x 4 = 16	150 x 150 x 8 = 180 000

Structural hierarchy

1D

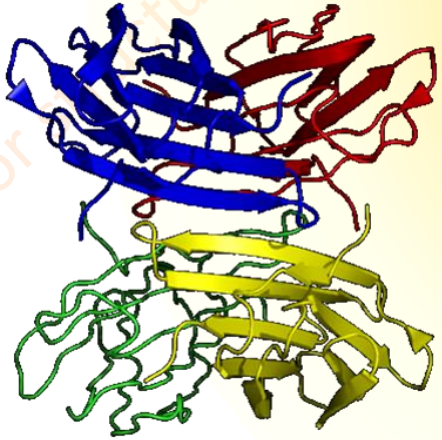
ADSQTSSNRAGEFSIPPNTDFRAIFFANAAE
QQHIKLFIGDSQEPAAYHKLTTRDGPREATL
NSGNGKIRFEVSVNGKPSATDARLAPINGK
KSDGSPFTVNFGIVVSEDGHDSDYNDGIVV
LQWPIG

primary (sequence)

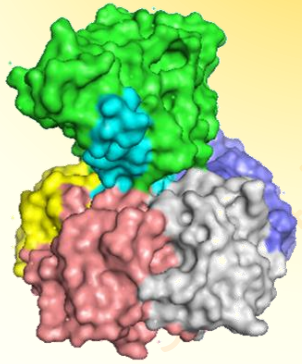


secondary

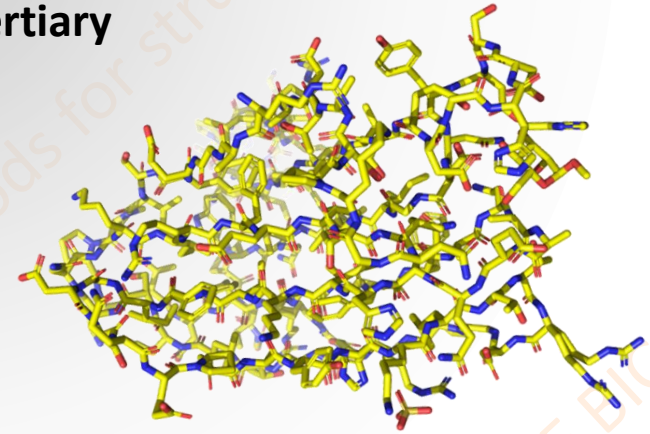
2D



quaternary



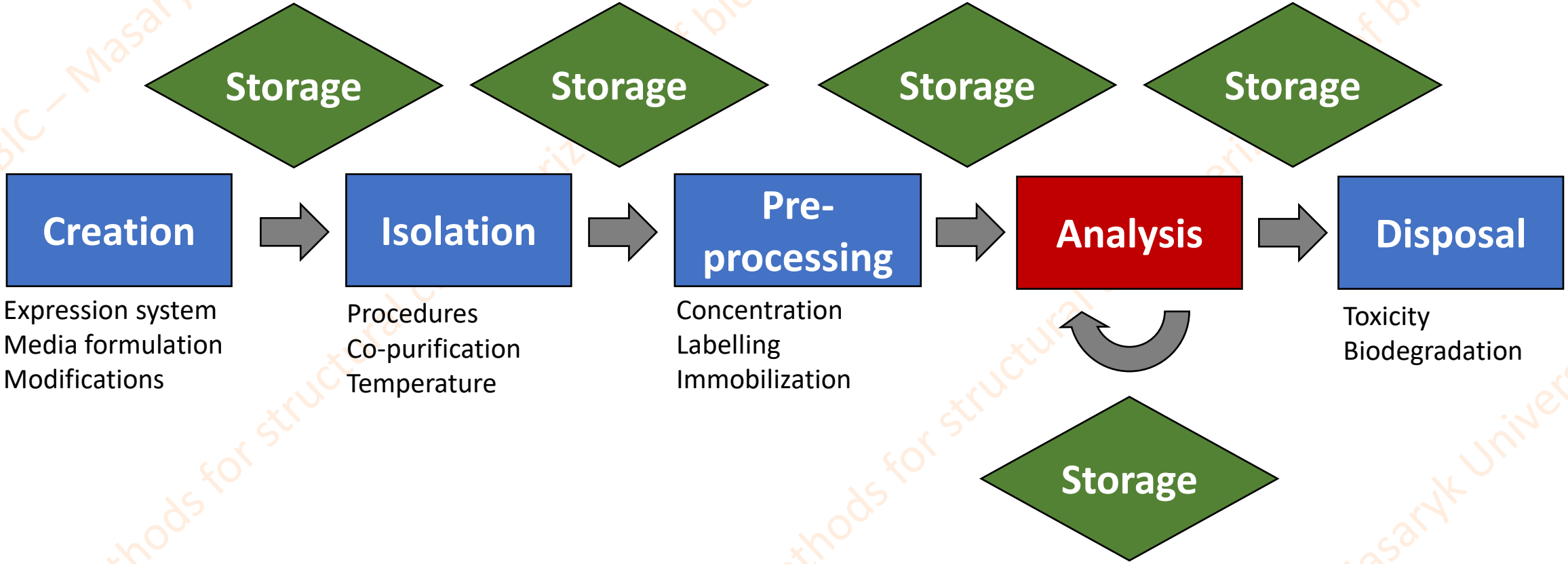
tertiary



4D

3D

Sample's life



- Forms of sample – solution, powder, crystal, surface-bound

Quality of sample

- All properties that relates to sample state and determining its behavior

Sample requirements

- Minimal requirements:
  - Conditions that sample has to meet in order to give some results
  - Differ heavily for individual techniques
  - Minimal requirements for specific technique do not ensure good sample !!!

Example

Minimal requirements:
0.5 mg/ml sample
450 ul volume
No DTT in buffer

Concentration

- Mass concentration (ρ_i, γ_i): [mg ml^{-1}] = [ug ul^{-1}]
- Molar concentration (c_i): [M] = [mol l^{-1}], [mM], [uM]
- Conversion:

$$c_i = \frac{\rho_i}{M_i}$$

Molar mass – inaccurate knowledge cause errors!

Concentration determination

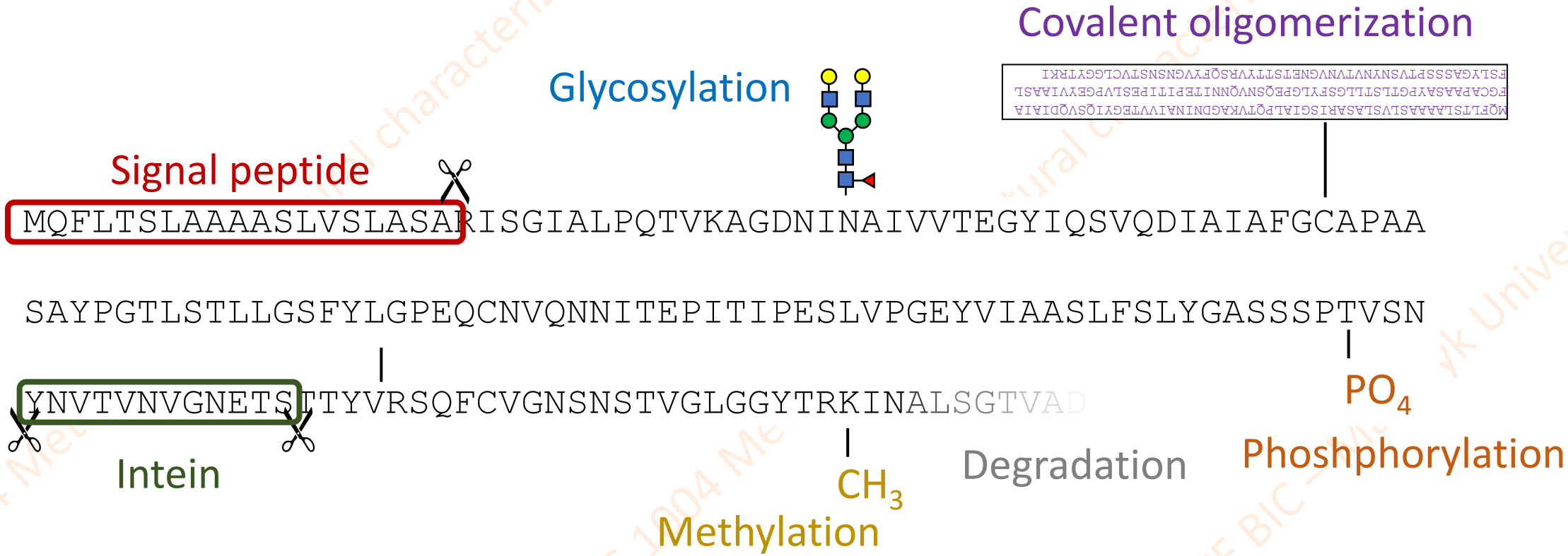
Method	+	–
Nitrogen content (e.g. Kjeldahl)	Absolute (golden standard)	Time, sample and equipment demanding
UV absorbance at 280 nm	Fast, easy, low sample consumption, no calibration	Sequence dependent, buffer influence, (inaccuracy in l, ε)
Bradford (Coomassie Brilliant Blue)	Easy, fast	Standard dependent (calibration), sequence dependent, buffer influence
Bicinchoninic acid	Less buffer dependent	Standard dependent (calibration), more time demanding
UV absorbance at 205 nm	Less sequence dependent, + the same as A_{280}	Buffer absorbance

Ideal sample properties

- **Defined** (chemically, biologically, conformationally)
- **Pure** (contamination by small molecules, macromolecules)
- **Homogeneous** (micro-/macro- heterogeneity)
- **Stable** (storage, time-demanding analysis)

Sample identity

- Exact composition of sample (sequence, modifications, cleavage)
- Influence on MW, pI, interactions

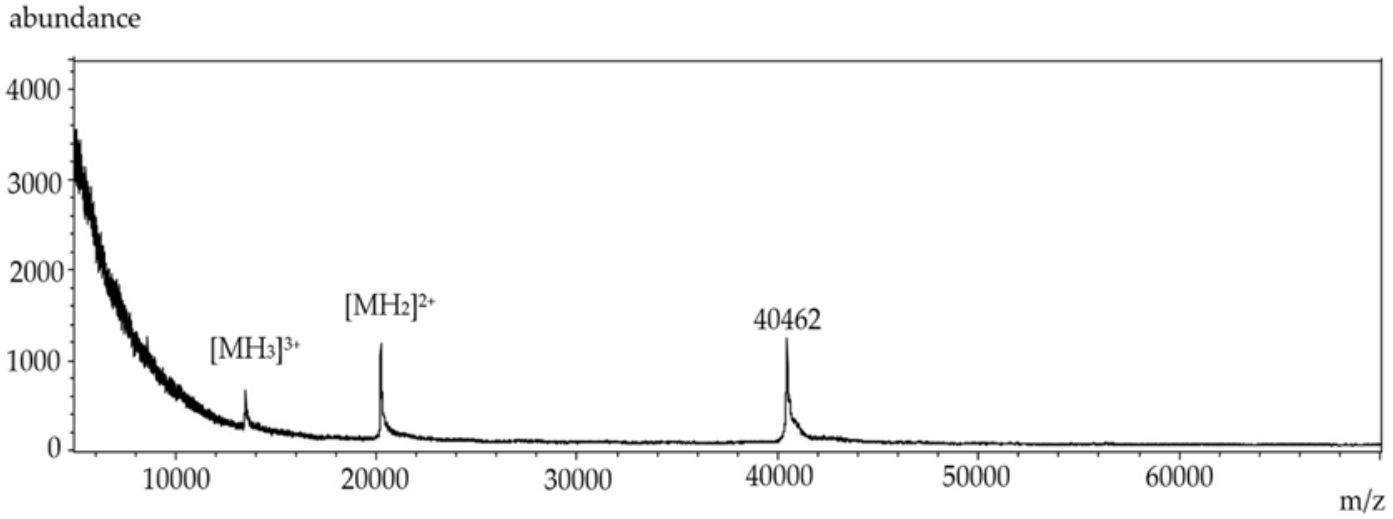


Sample identity

- MS identification

```
  1 MKKESINTSG PDNTKSSISD EIEISNEISW TALSGVISAA NNADGRLEVF
 51 GVGTNNAVWH NWQTVPNTGS SWSGWHSLNE GATSKPAVHI NSDGRLEVFV
101 RGTDNALWHN WQTVPGAGWS GWQSLGGQIT SNPVVYINSD GRLEVFARGA
151 DNALWHIWQT APHAGPWSNW QSLNGVLTSD PTVYVNASGR PEVFARSNDY
201 SLWYIKQTAS HTYPWTNWQS LSGVITSNPV VISNSDGRLE VFARGSDNAL
251 WHIWQVAPNA GWTNWRSLSG IITSDPAVHI NADGRLEVFA RGPDNALWHI
301 WQTATSDAWS EWTSLSGVIT SAPTVAKNSD GWLEVFARGA NNALCHIQQT
351 TSSWSTWTSL GGNLIDASAI K
```

- MS intact mass analysis



Post-translational modifications
Isotope labeling
Matrix adducts

Sample purity

Contaminants – co-purified molecules

- Small molecules
  - Co-factors
  - Ligands
  - Salts, imidazole
  - Lipids
  - Saccharides
- Macromolecules
  - Protein isoforms
  - Proteins
  - Nucleic acids
  - Polysaccharides
  - Binding partners

Sample purity – methods

- **SDS-PAGE**
- UV-VIS spectroscopy
- SEC (SEC-MALS)
- FFF (FFF-MALS)
- Mass spectrometry

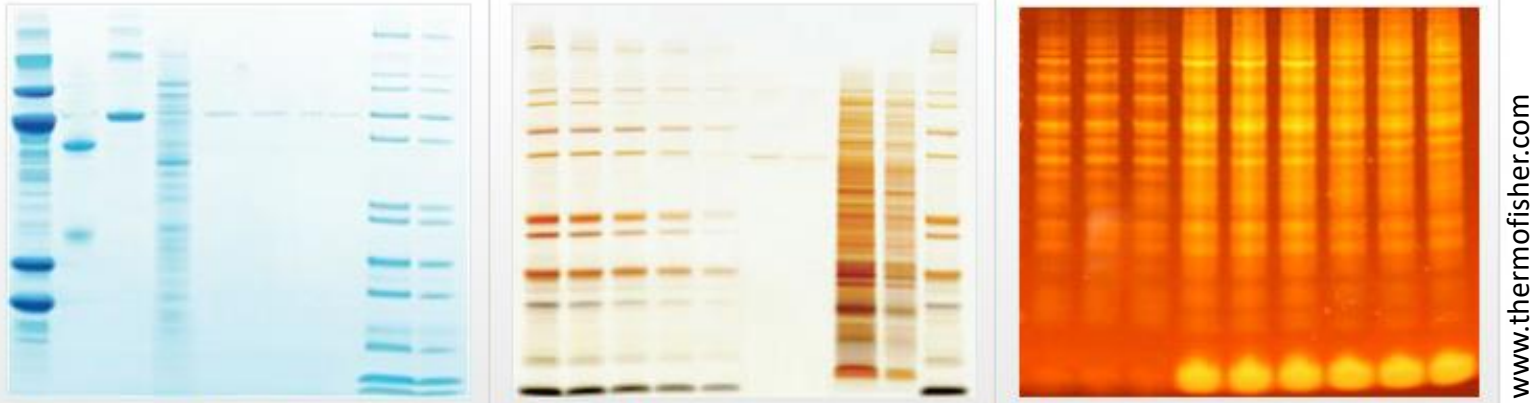
small molecules
- Co-factors
- Ligands
- Salts, imidazole
- Lipids
- Saccharides

macromolecules
- Protein isoforms
- Proteins
- Nucleic acids
- Polysaccharides
- Binding partners

SDS-PAGE

- **Polyacrylamide** gel (8 – 20 %)
- **SDS** – uniform (?) protein charge (composition dependent)
- **Reducing agent** (optional) – βME
- **Staining** – CBB, Silver, Fluorescent, Radiological

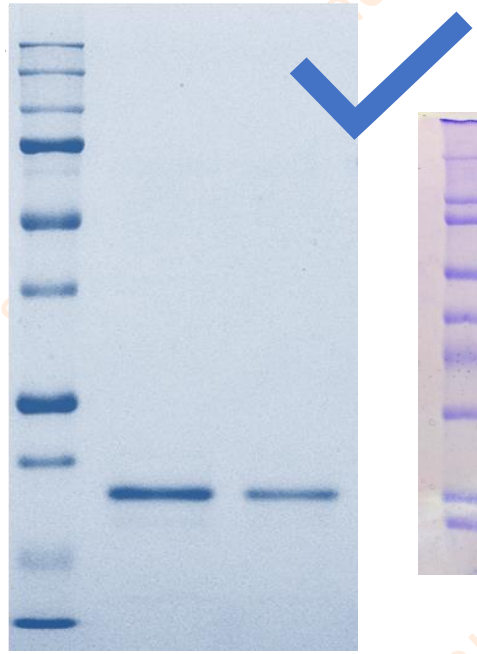
	Coomassie staining	Silver staining	Fluorescent protein staining
Sensitivity	5-25 ng	0.25-0.5 ng	0.25-0.5 ng

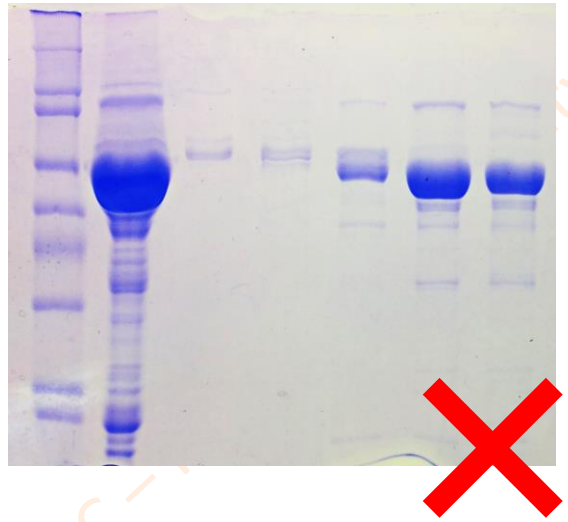


www.thermofisher.com

SDS-PAGE

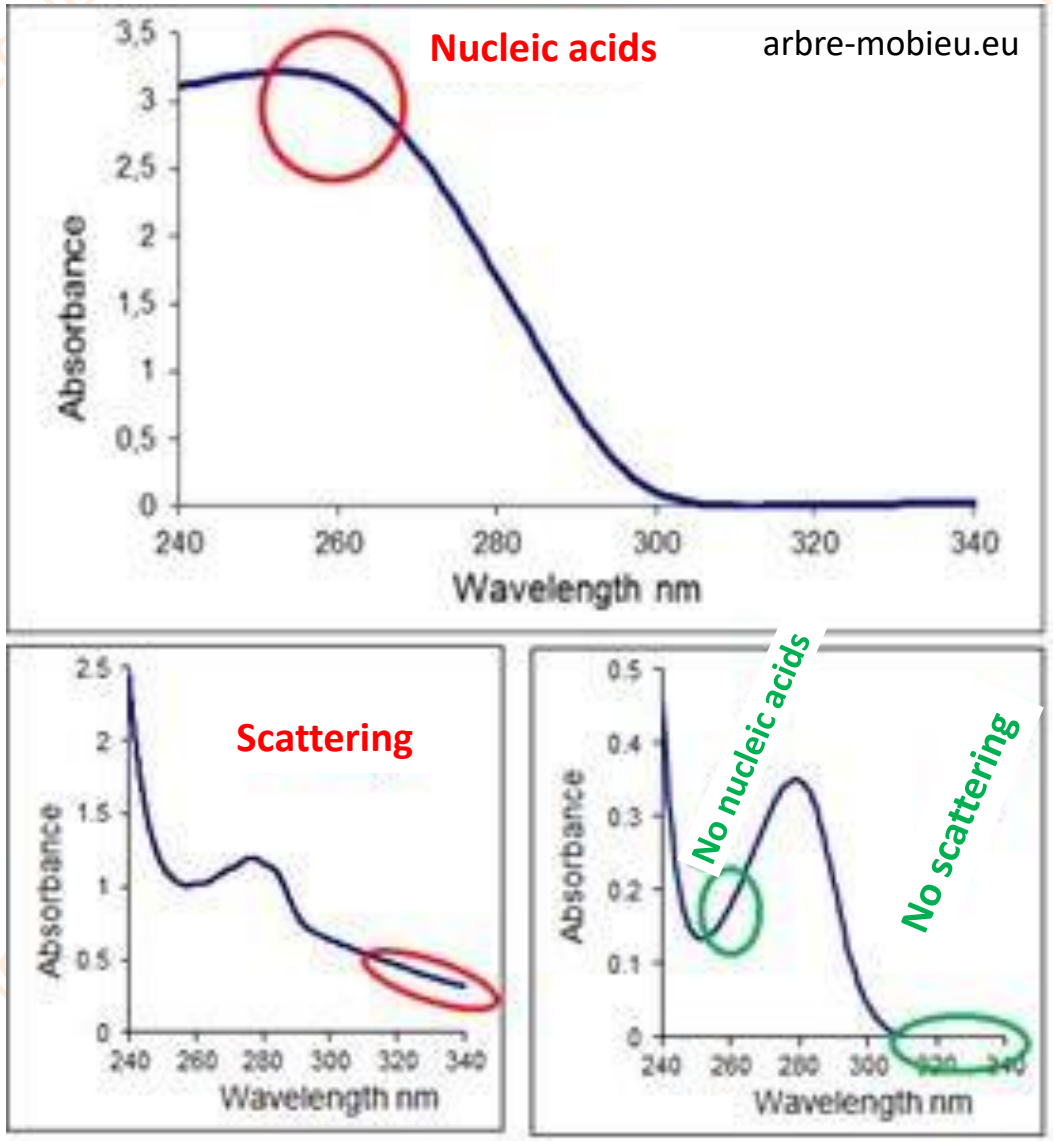
- Check overloaded as well as underloaded sample





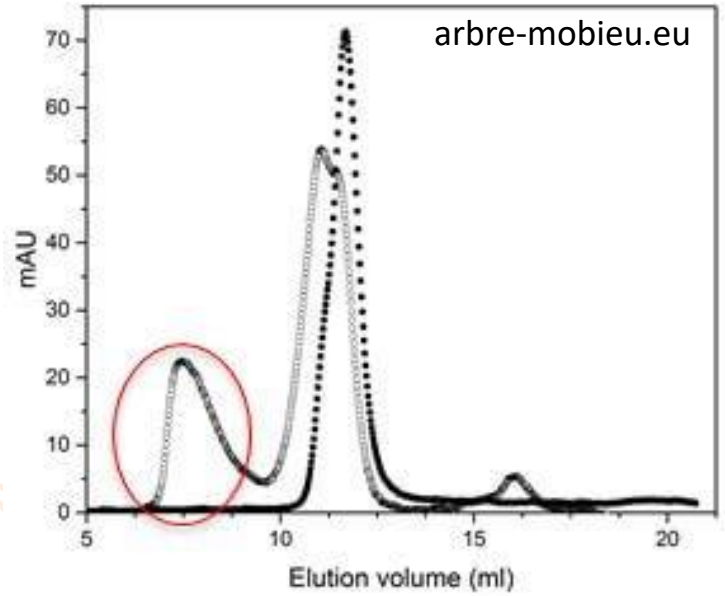
UV-VIS spectroscopy

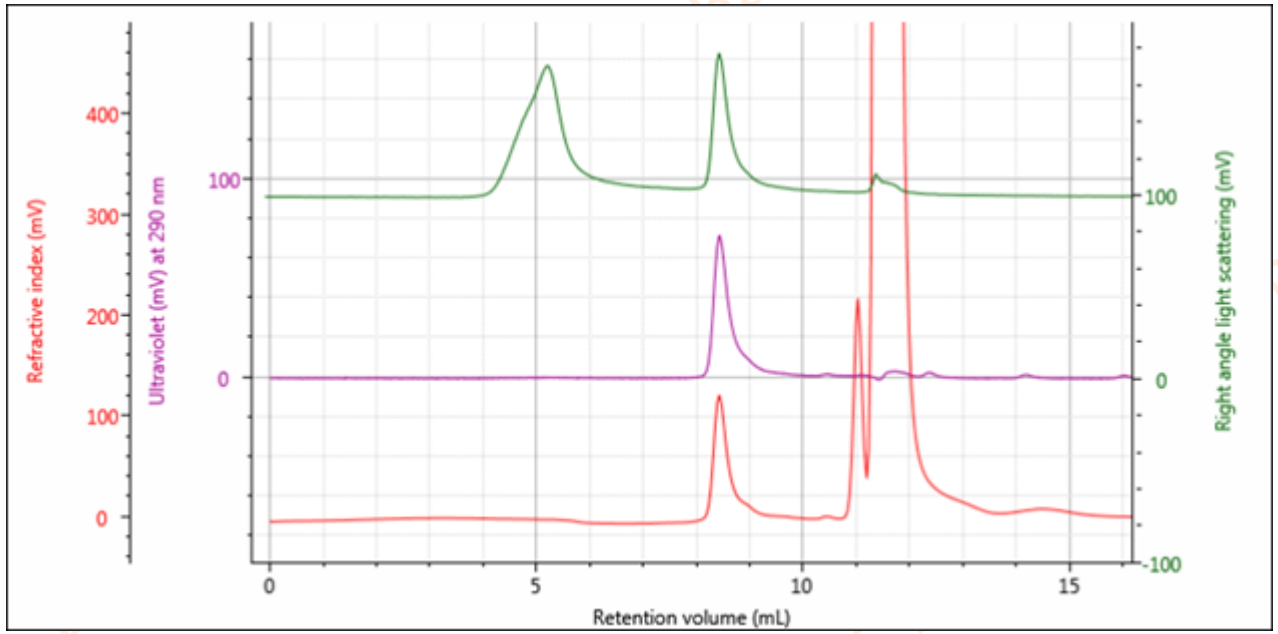
- (200-) 240 – 340 nm
- Trp (and Tyr) has absorption peak around 280 nm
- Detection of:
  - **Nucleic acid** contamination
  - **Aggregation** (scattering)
  - **UV-absorbing contaminants**



Size exclusion chromatography

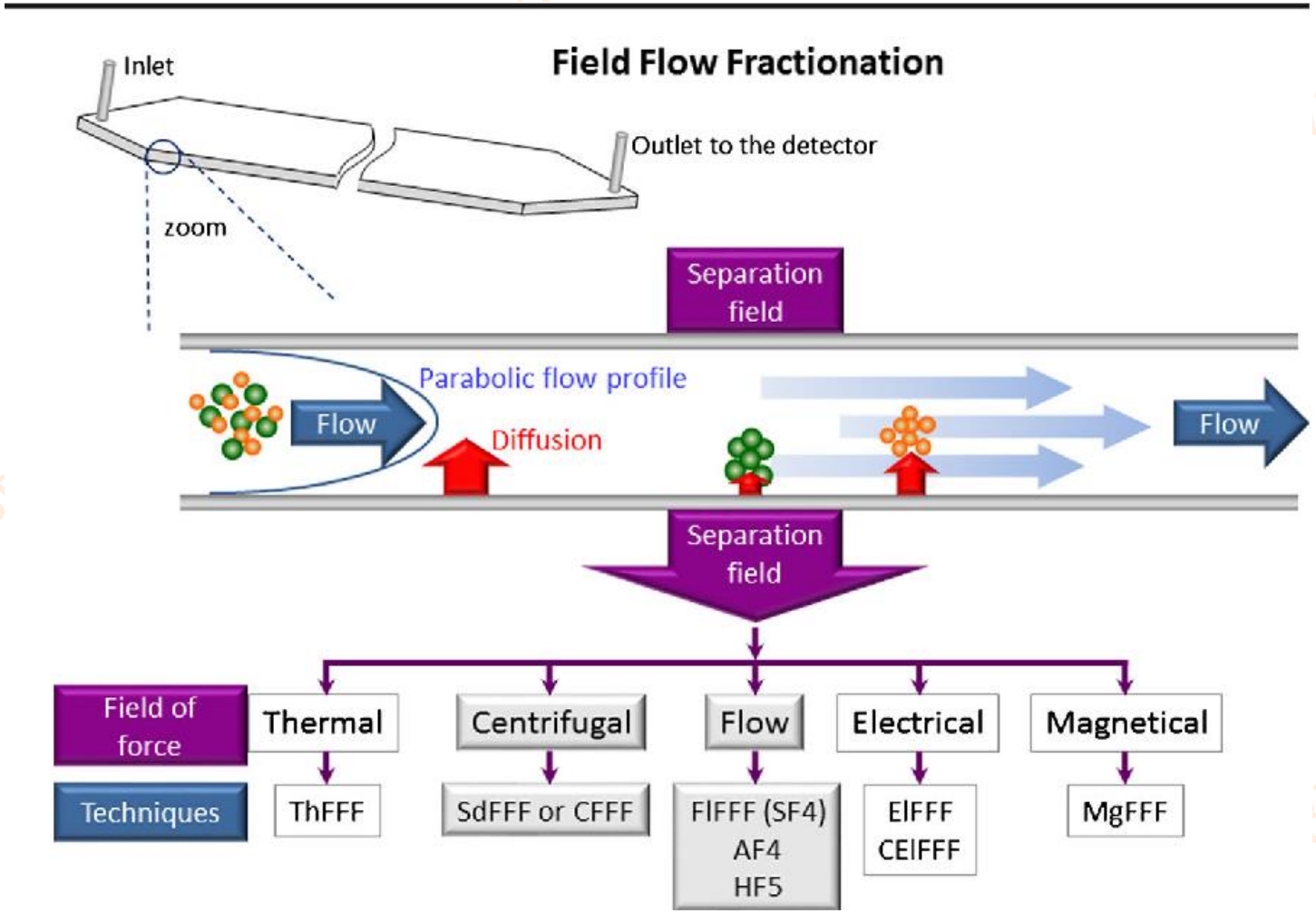
- Separation of particles based on “size”
- Interaction with matrix possible (!)
- Possibility to couple to multiple detectors (UV, RI, MALS, viscosity)





Field flow fractionation

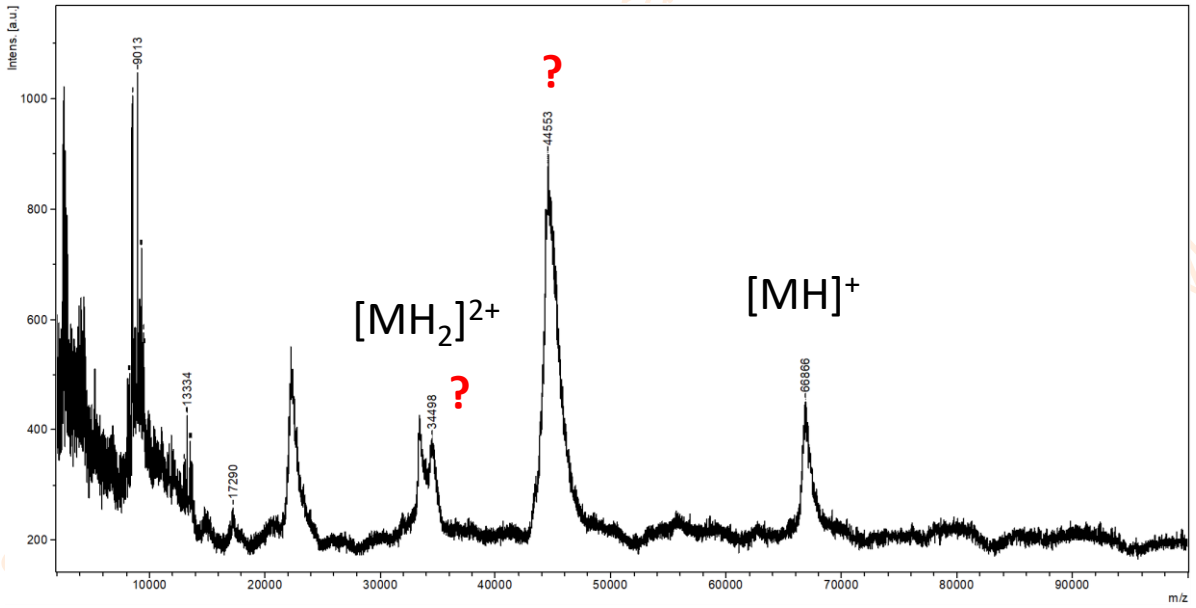
- Separation of particles in solution by external force



Mass spectrometry

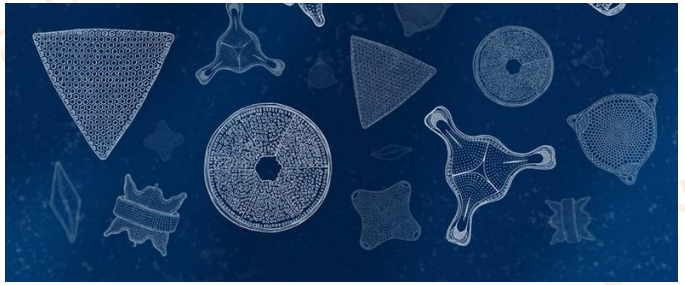
- Detecting of exact mass of particles
- Various applications based on set-up

- **Intact mass analysis** – protein and non-protein contaminants



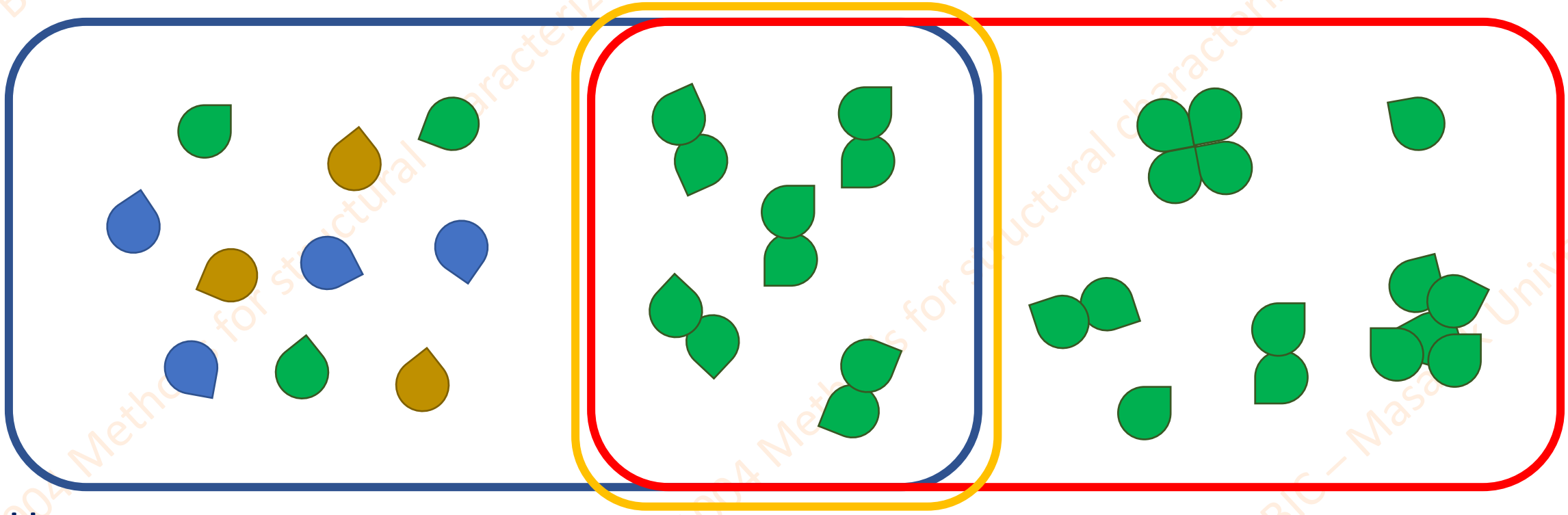
Sample homogeneity

- **Macroscopic** – precipitation – visual detection
- **Microscopic** – oligomeric states, folding states, microheterogeneity – biophysical methods



Sample homogeneity vs. purity

- Various methods may evaluate sample in different way

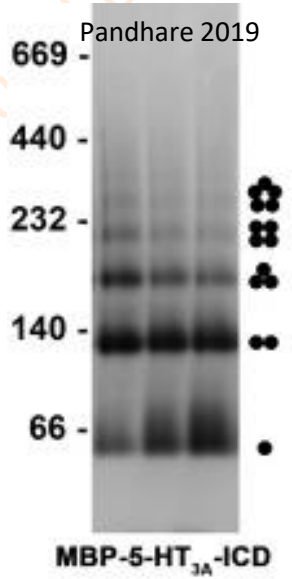


Sample homogeneity – methods

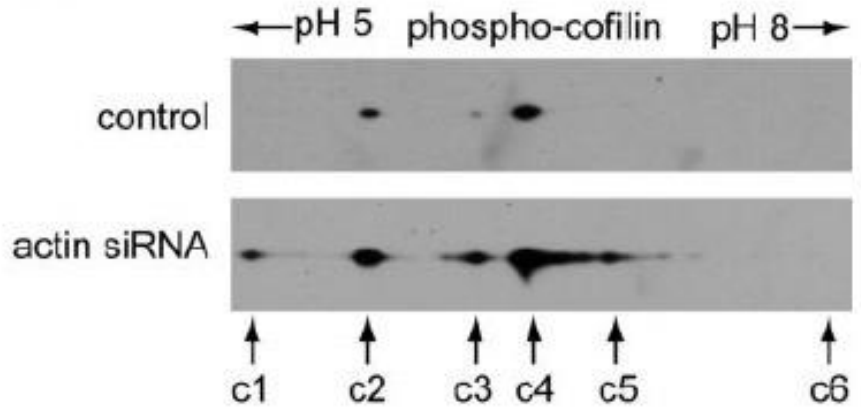
- **SEC-MALS, FFF**
- **Native electrophoresis**
- **Light scattering**
- **Analytical ultracentrifuge**

Native electrophoresis

- Possibility to observe various **oligomers** (relatively imprecise and unreliable) and **isoforms** (2D PAGE preferred)
- Not efficient for **aggregation** detection

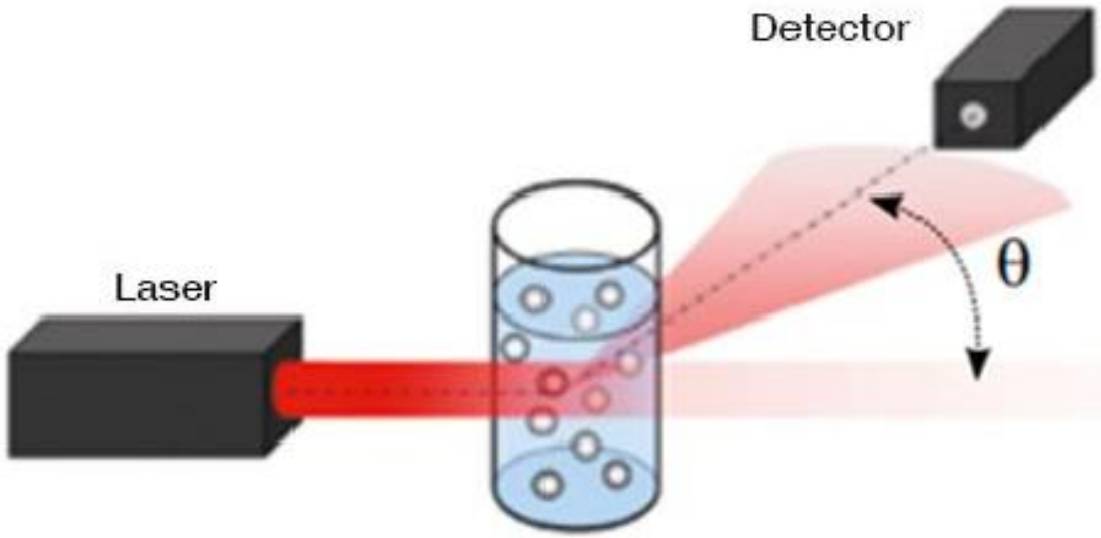


Pandhare 2019



Light scattering

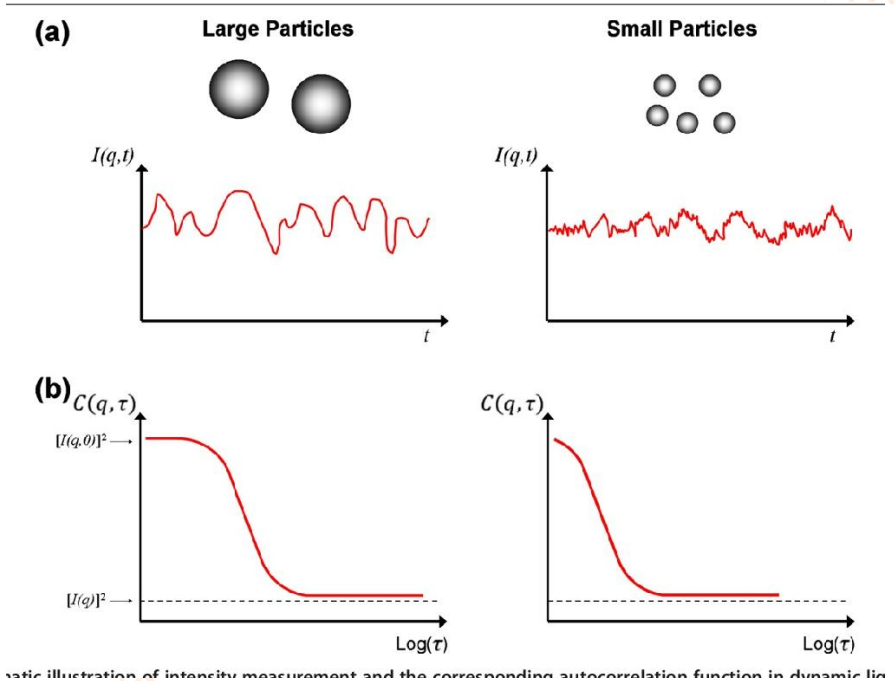
- Interaction of incident light with particles in solution
- Intensity of light at given
- Typically red/infrared light

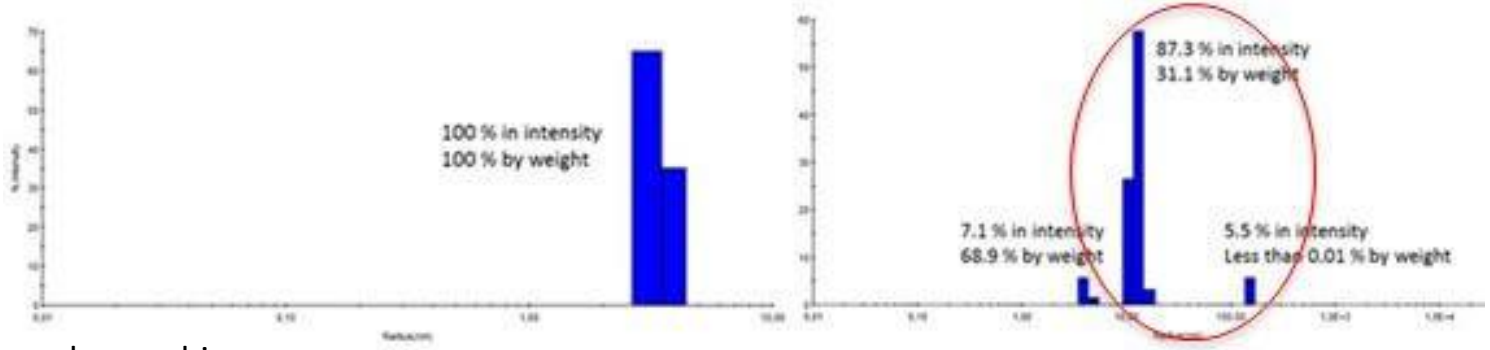


Light scattering

- **Dynamic light scattering**
  – size of particles
  – sensitive to **aggregation**

- **Static light scattering**
  – mass of particles
  – averaged value, separation required



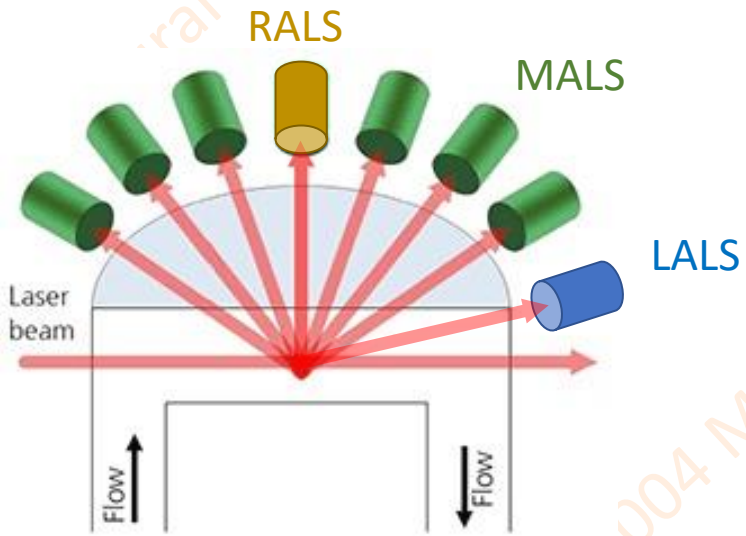


arbre-mobieu.eu

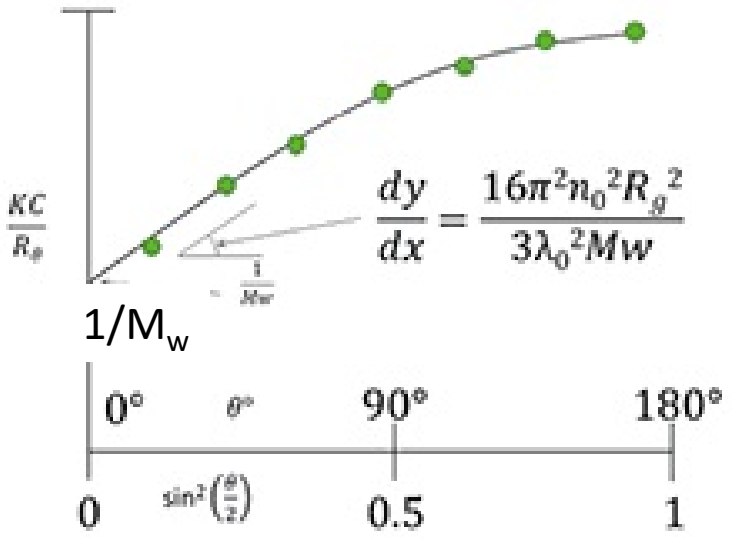
Static light scattering (SLS)

Low-angle light scattering (LALS) – big molecules
Right-angle light scattering (RALS) – small molecules
Multi-angle light scattering (MALS) – M_w and R_g

- Intensity of scattered light
- **Mass** of the particle (molecular weight)

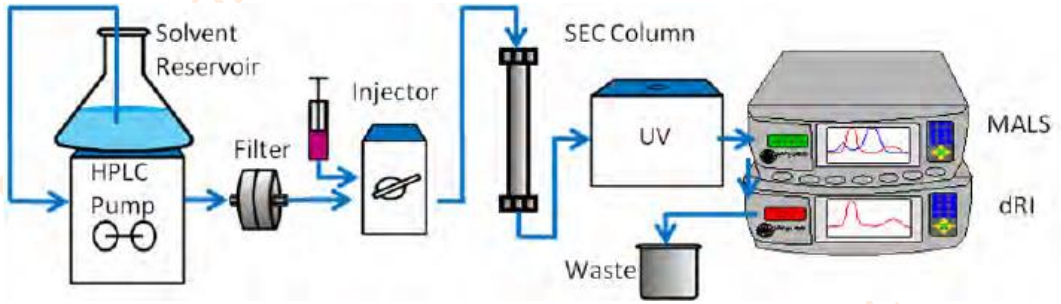


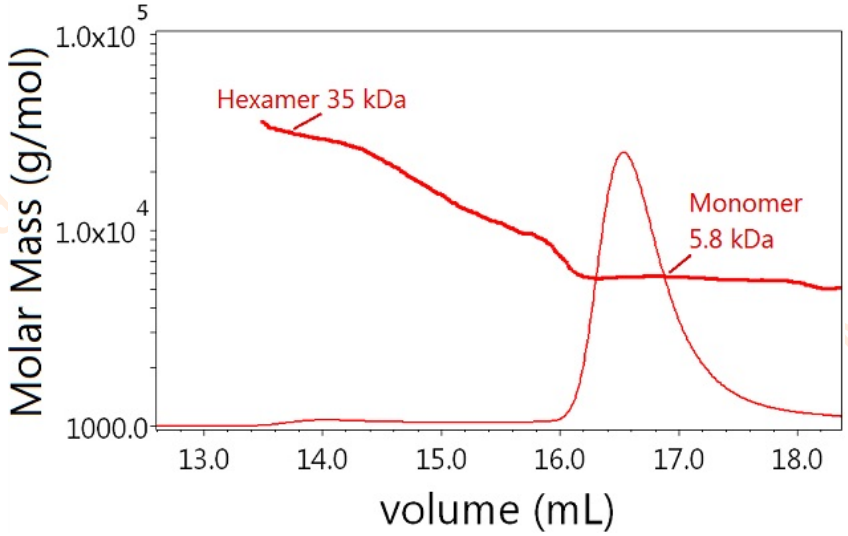
www.malvernpanalytical.com

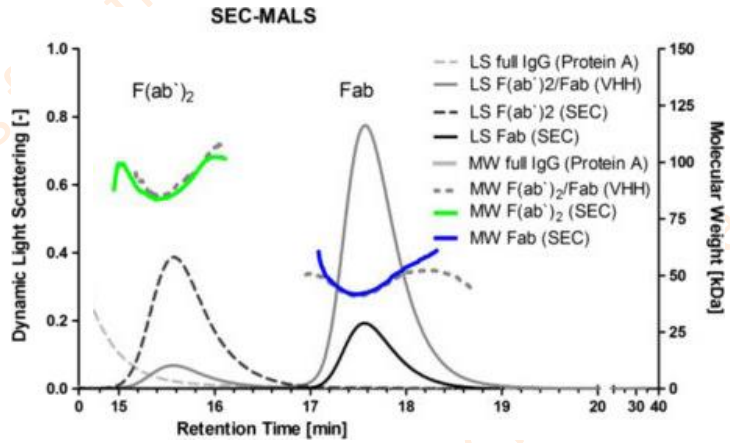


Static light scattering

- Average of all sample particles !
- Typically coupled to separation (SEC, FFF)

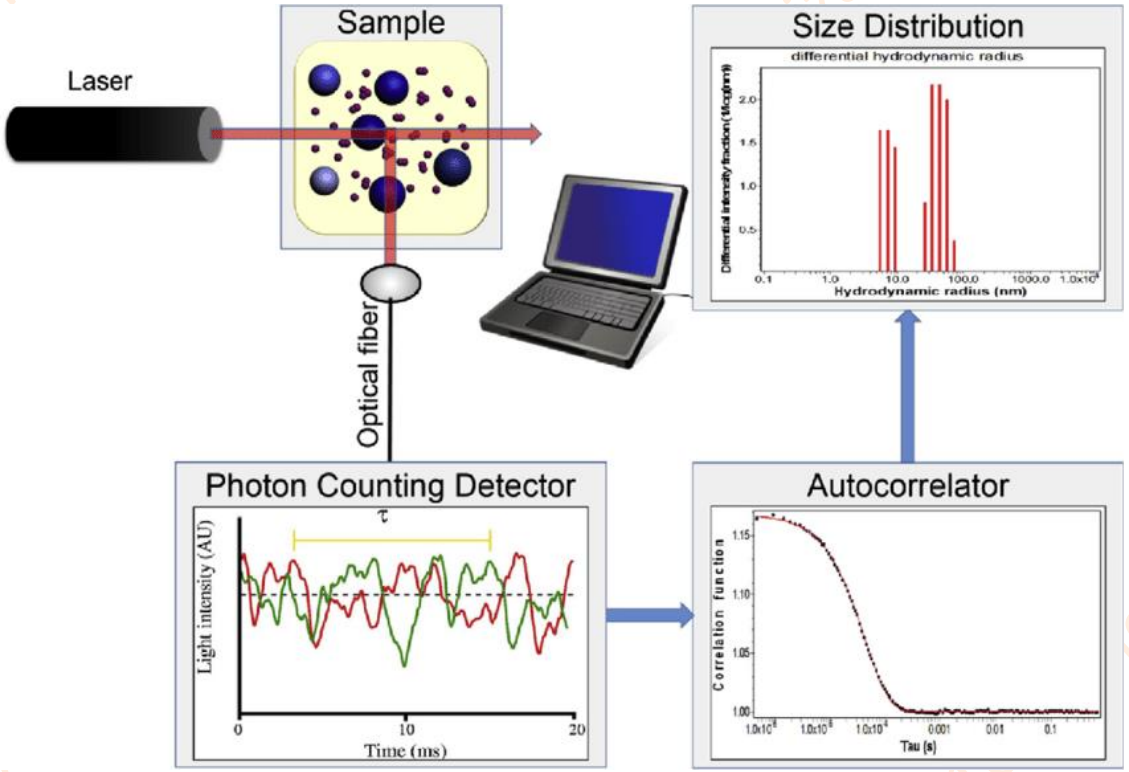


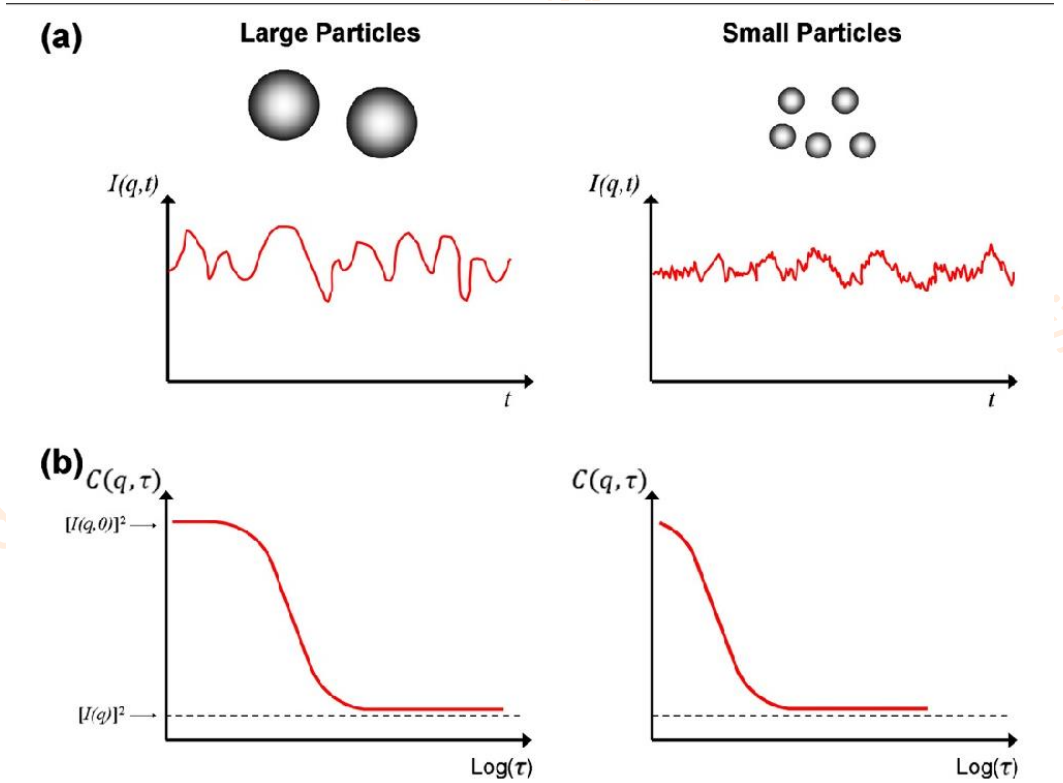




Dynamic light scattering (DLS)

- Time-dependent fluctuations in scattered light
- **Size** of the particle (hydrodynamic radius)

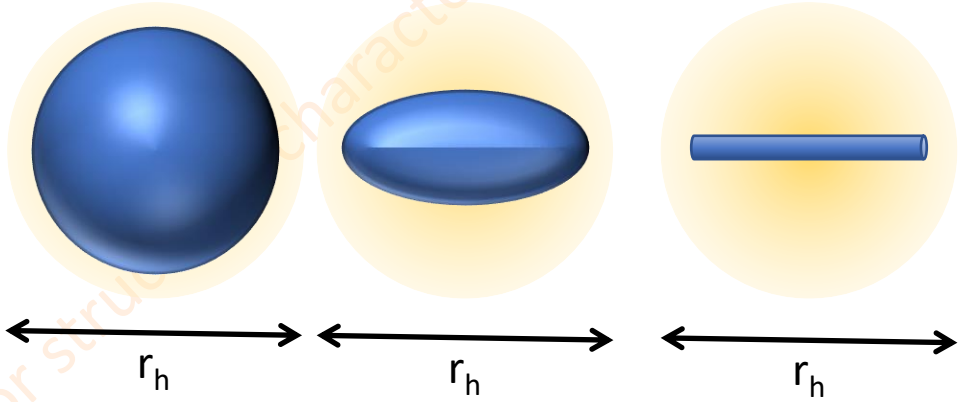


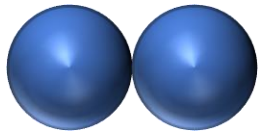


Allen P Minton,
DOI: 10.1016/j.ab.2016.02.007

Dynamic light scattering (DLS)

- Shape dependent
- Low resolution





r_h(dimer) ~ 2 × r_h(monomer)

r_h(dimer) ~ r_h(monomer)

For ideal sphere:

$$M \sim V = 4/3\ \pi\ r^3$$

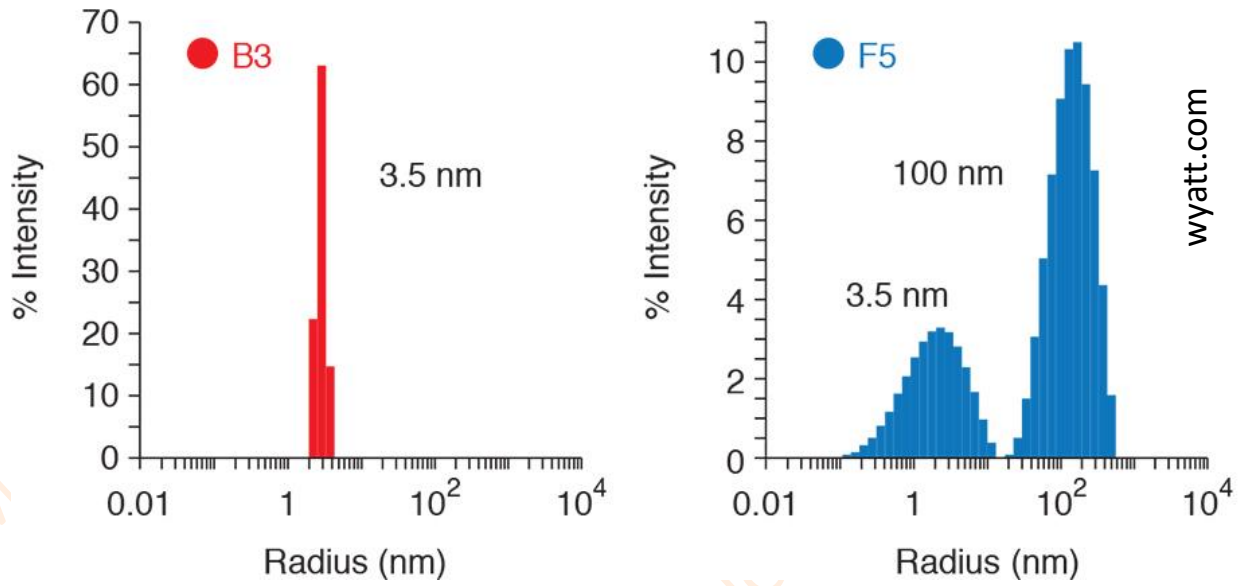
$$M_2 = 2 \times M_1$$

$$r_2 = \sqrt[3]{2} \times r_1 = 1.26\ r_1$$

Dynamic light scattering (DLS)

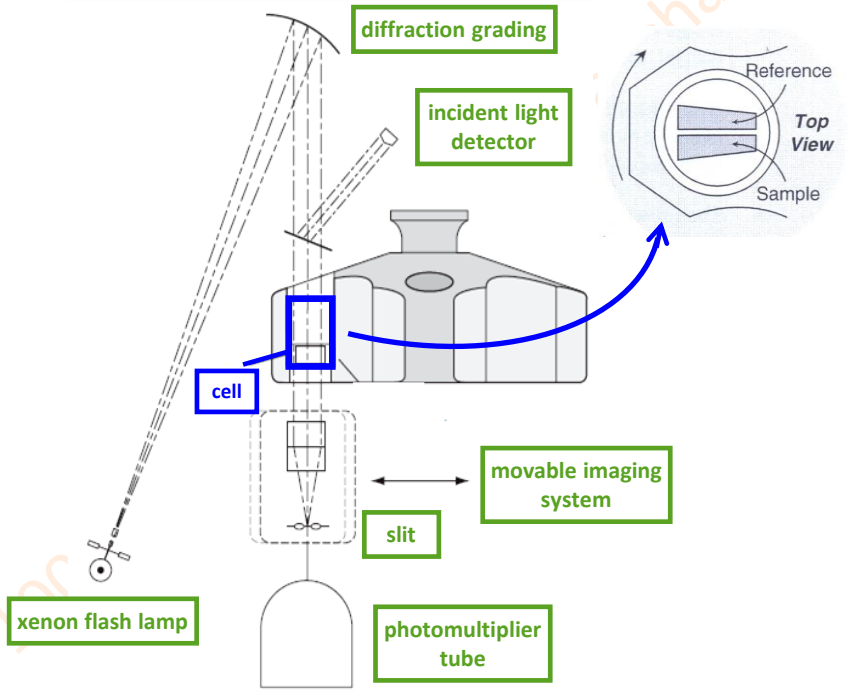
– Microheterogeneity reflects in **polydispersity** – peak width

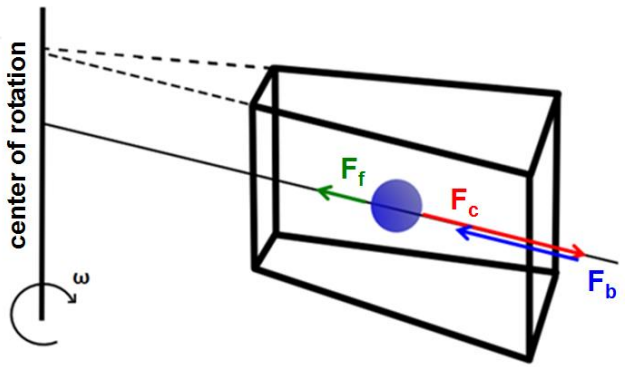
– Large particles scatter light with much higher intensity – sensitive to **aggregation**

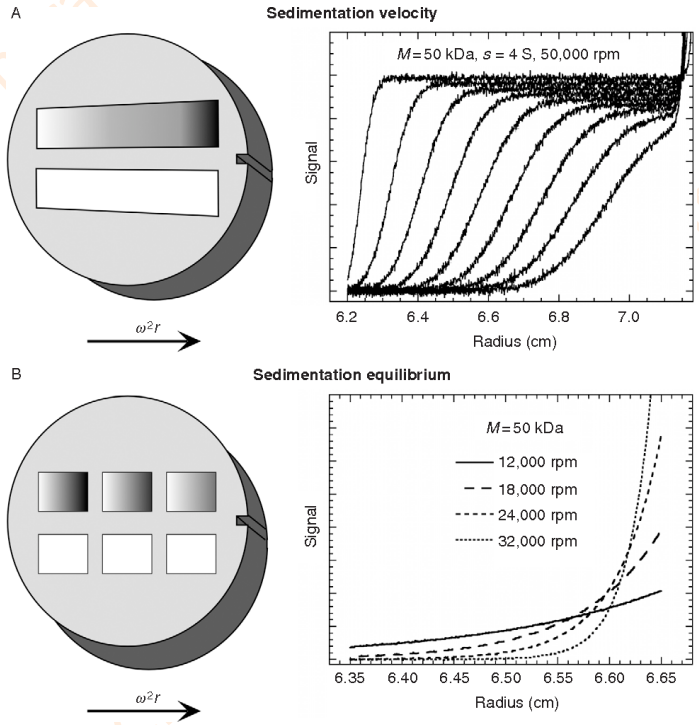


Analytical ultracentrifugation (AUC)

- Sedimentation of particles in centrifugal field by **hydrodynamic properties**
- Two modes:
  - Sedimentation equilibrium – mass determination
  - Sedimentation velocity – size distribution

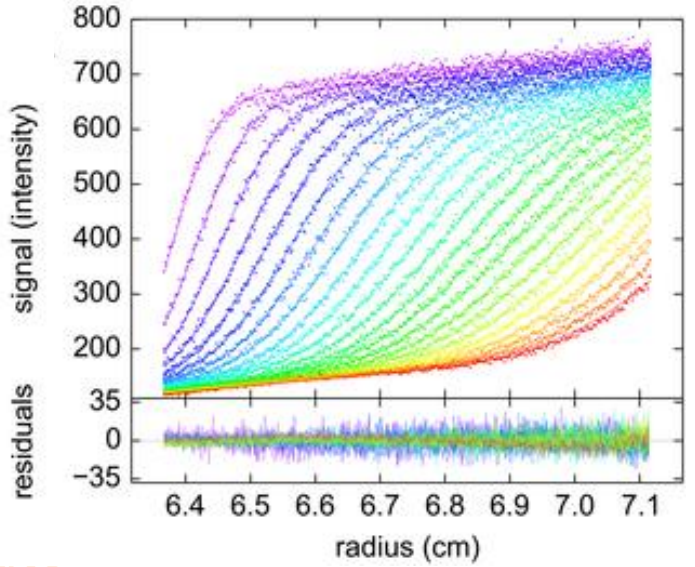


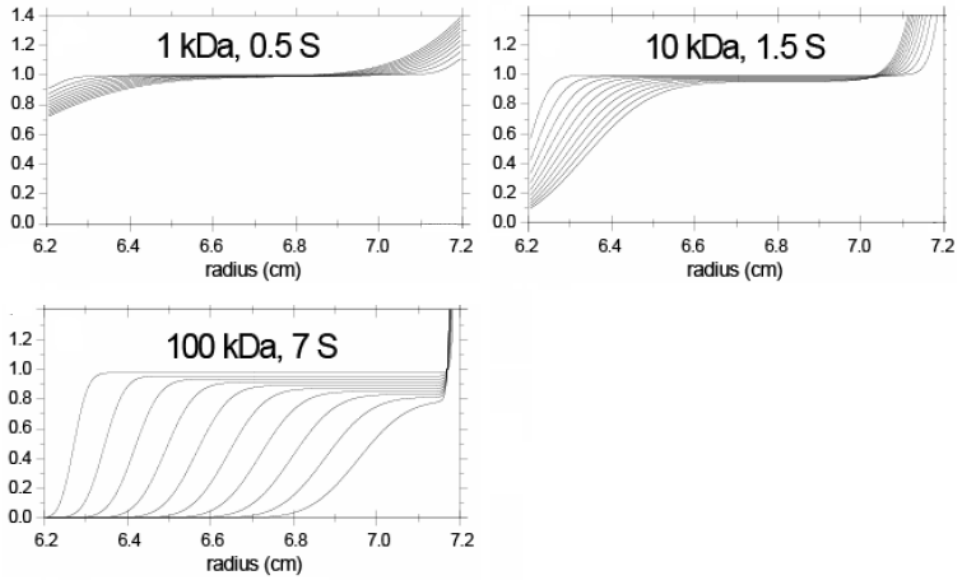




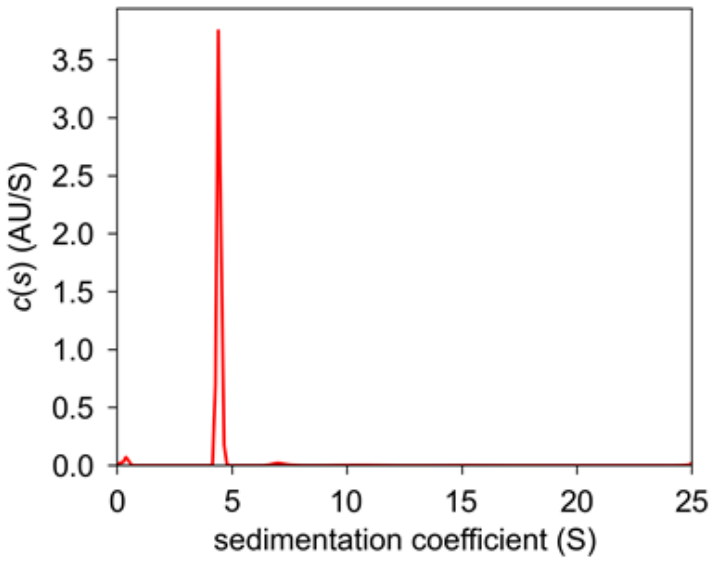
AUC – Sedimentation velocity

- Sedimentation of particles over time observed
- Suitable to detect and quantify **aggregates**
- **Size** of the particle (hydrodynamic radius)
- Sensitive to shape (and density)

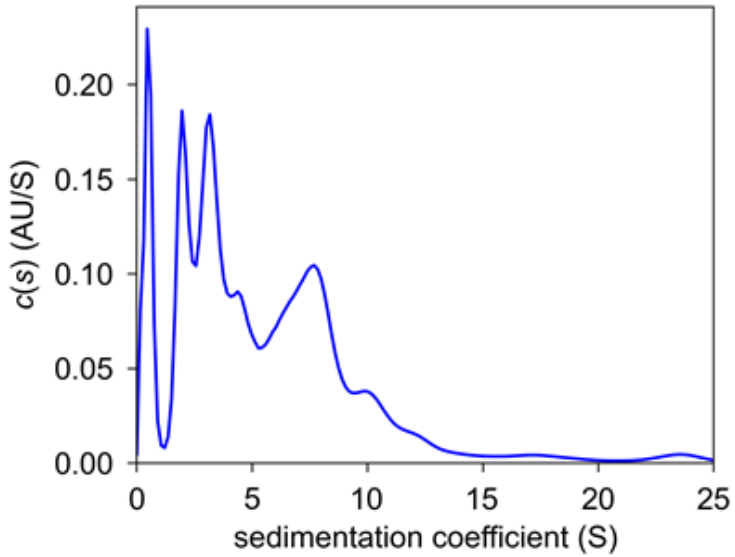




monodisperse sample



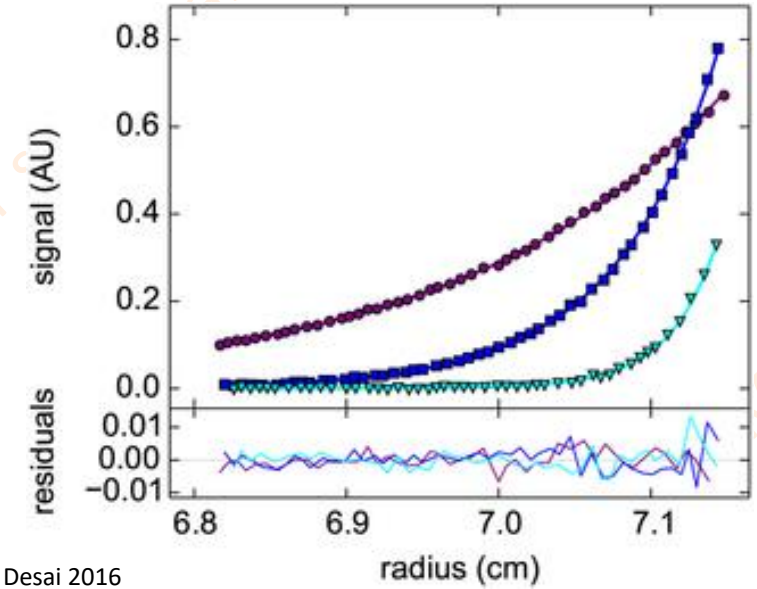
polydisperse sample



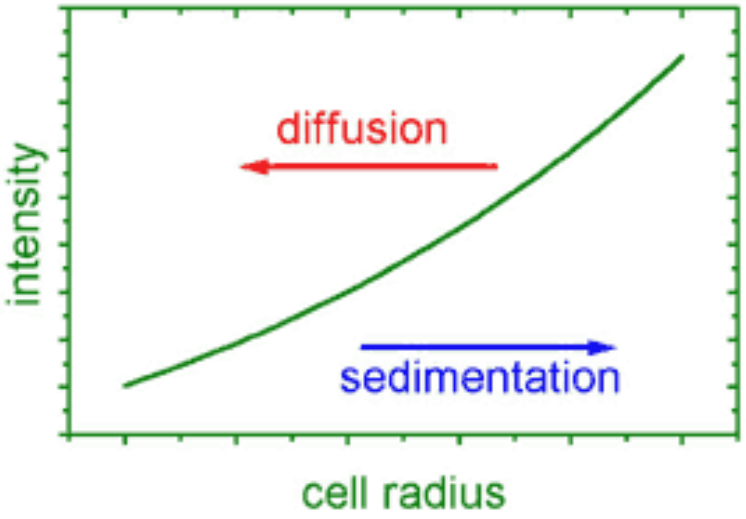
AUC – Sedimentation equilibrium

- Distribution of particles in cell

- **Molecular mass** of particle
- Problematic for mixtures



Desai 2016

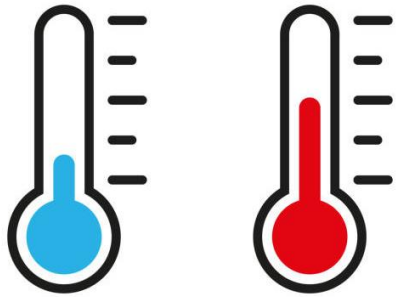


Comparison

	Light scattering	Analytical ultracentrifugation
Sample volume	0.5-30 ul (DLS) 1-50 ul (SLS, SEC-MALS)	150 – 450 ul
Sample concentration	0.1 – 200 mg/ml	0.1 – 1 mg/ml
Particle size	1 nm – 10 µm	1 – 300 nm
Resolution and accuracy	Low – Average	Average – High
Speed of analysis	1 min (DLS, SLS) 30 mins (SEC-MALS)	4 hrs (SV) 3-4 days (SE)

Sample stability

- **Temperature** stability



- **Chemical** stability
  - pH
  - Ionic strength
  - Oxidizing agents
  - Protein-specific compounds

- Long-term stability – **storage**

Temperature

- Affects stability and interaction parameters

$$\ln K_A = -\frac{\Delta G_0}{R\,T} \qquad k = A\,e^{\frac{-E_a}{R\,T}}$$

Arrhenius equation

- Typical temperatures:

  –80 °C, –20 °C, 4 °C, 20 °C, 25 °C, 37 °C

- **Room temperature** (RT) – vaguely defined

  mostly 20 – 25 °C, but varies from 15 – 30 °C

  usually means that temperature was not set (!)

pH

$$pH = -\log [H_3O^+]$$

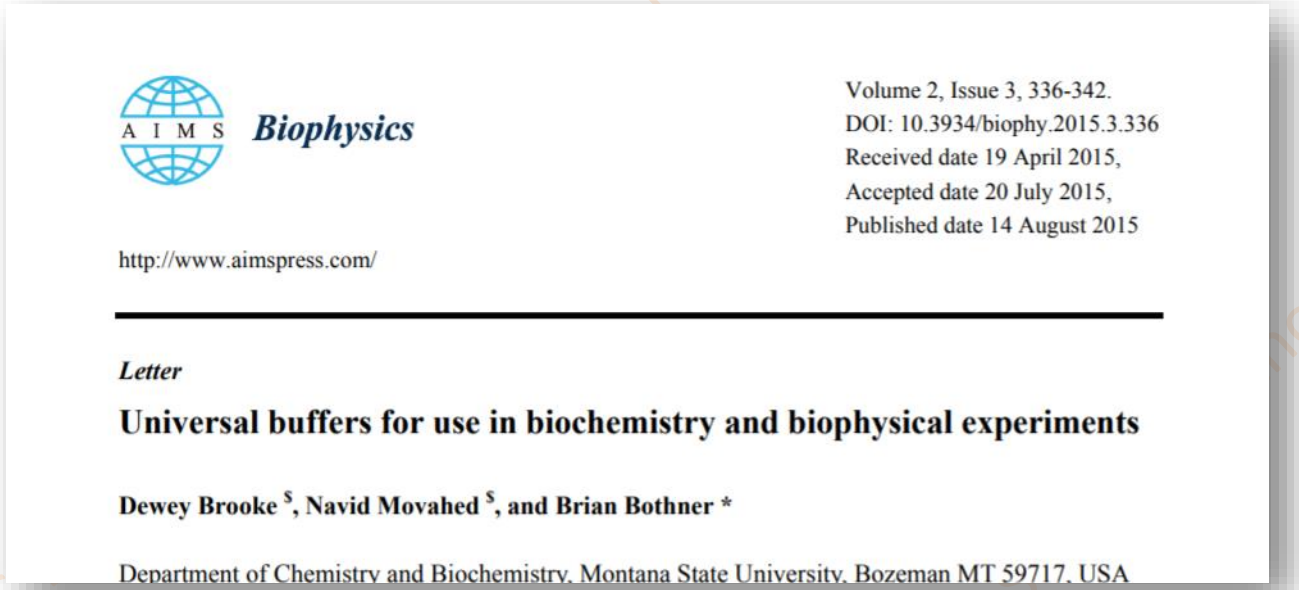
Typical range: 4 – 9, specific proteins 1 – 12

pH of **pure water**: 7 (theor.), 5.8 (due CO_2 absorption)

Buffers: dissociable compounds with defined pK_a
various pH ranges – typically (pK_a–1) – (pK_a+1)

pH – buffers

- Organic/Inorganic
- Universal buffers – mixtures with broad pH range



AIMS Biophysics

Volume 2, Issue 3, 336-342.
DOI: 10.3934/biophy.2015.3.336
Received date 19 April 2015,
Accepted date 20 July 2015,
Published date 14 August 2015

http://www.aimspress.com/

Letter

Universal buffers for use in biochemistry and biophysical experiments

Dewey Brooke [$], Navid Movahed [$], and Brian Bothner *

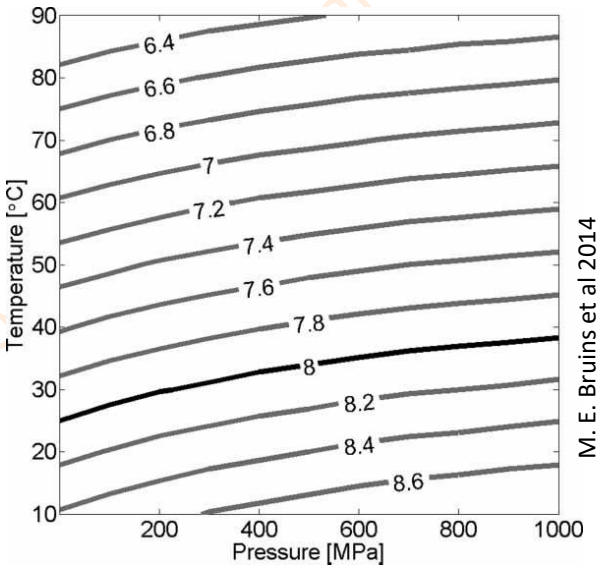
Department of Chemistry and Biochemistry, Montana State University, Bozeman MT 59717, USA

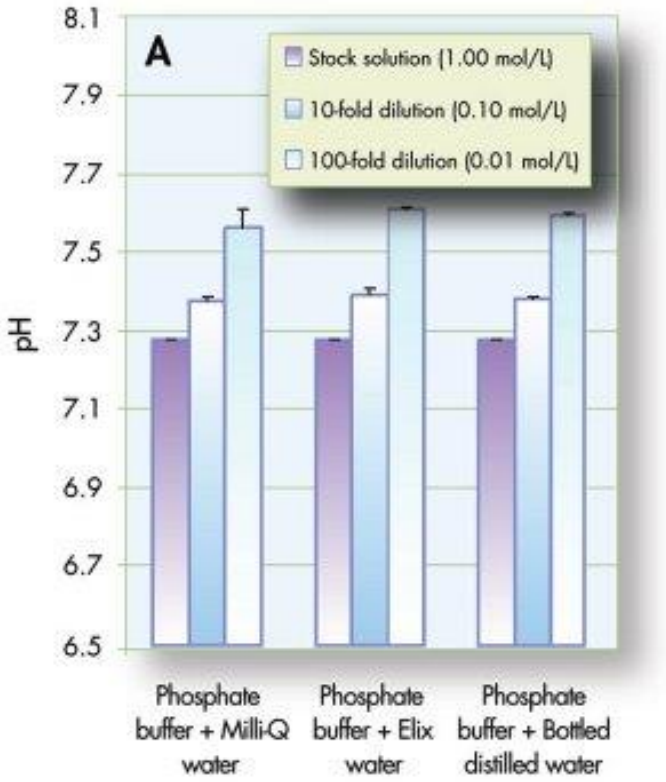
Good's Buffer	pKa (20 °C)	pH
MES	6.15	5.5-7.0
Bis-Tris	6.46	5.7-7.3
ADA	6.60	5.8-7.4
PIPES	6.80	6.1-7.5
ACES	6.90	6.0-7.5
MOPSO	6.95	6.2-7.4
BES	7.15	6.6-8.0
MOPS	7.20	6.5-7.9
TES	7.50	6.8-8.2
HEPES	7.55	6.8-8.2
TAPSO	7.70	7.0-8.2
POPSO	7.85	7.2-8.5
HEPPSO	7.90	7.4-8.6
EPPS	8.00	7.5-8.5
Tricine	8.15	7.8-8.8
Bicine	8.35	7.7-9.1
TAPS	8.40	7.7-9.1
CHES	9.50	8.6-10.0
CAPS	10.40	9.7-11.1

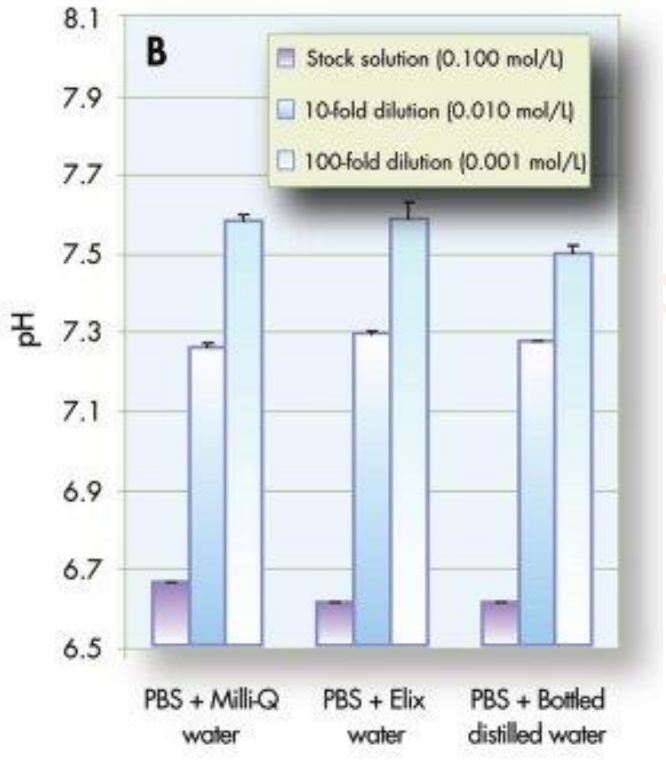
pH

- is **temperature** dependent
- changes with **dilution**
- changes in **time**

Tris buffer pH set to 8.0 at 25°C



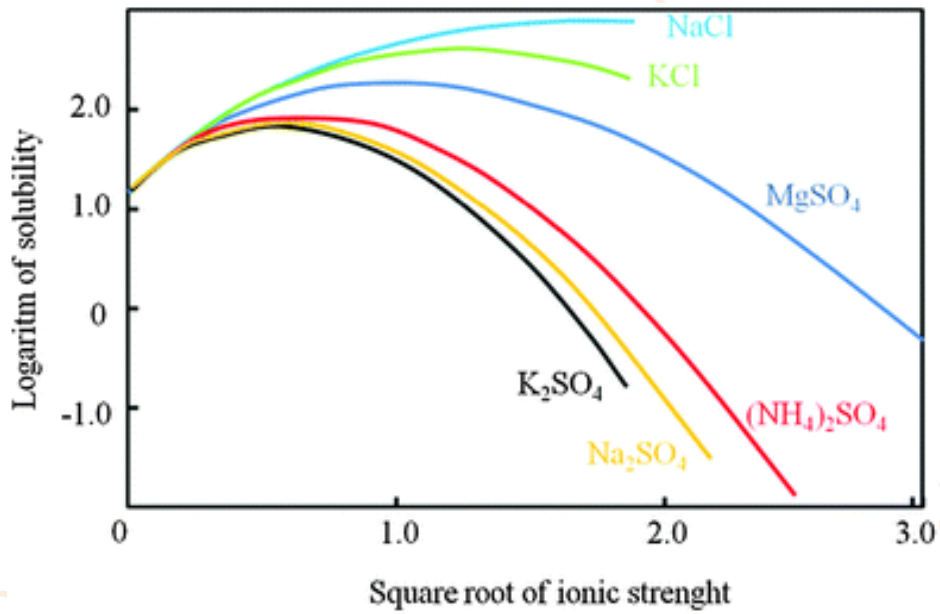




Ionic strength

Ionic strength, *I*, is a measure of the concentration of electrically charged species in solution

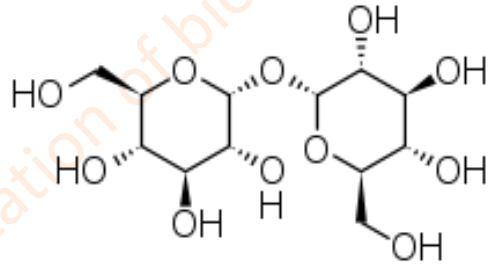
$$I = \frac{1}{2} \sum_i c_i Z_i^2$$



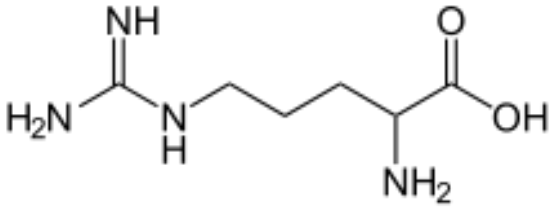
- Protein solubility changes with **ionic strength** as well as with **solute composition**

Impurities/Additives

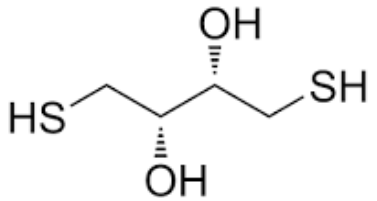
- Various compounds affect protein stability/solubility

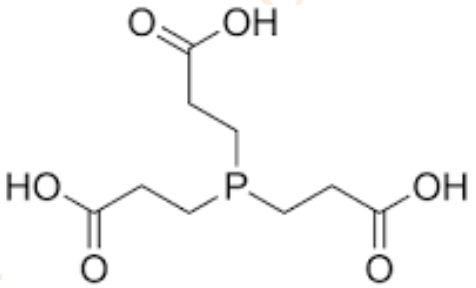


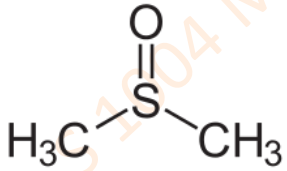
- **Saccharides** – saccharose, trehalose
- **Amino acids** – Arg, Glu, Pro



- **Reducing/oxidizing** agents – βME, DTT, TCEP
- **DMSO**
- Protein-specific compounds (ligands)







Buffer optimization

- Buffer affects:
  - Stability
  - Activity (interactions)
  - Storage

- Many buffers do not meet all requirements

Buffer optimization desired

Buffer optimization

- Various commercial screens available

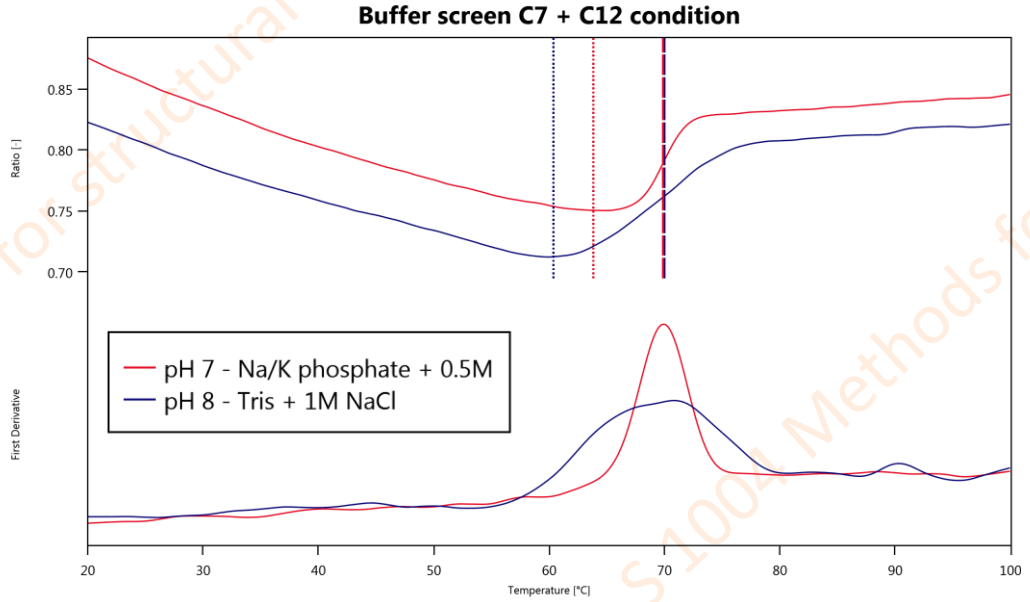
- Differences in composition, number of conditions

Example: buffer screen designed by CF BIC, CEITEC MU

	1	2	3	4	5	6	7	8	9	10	11	12
A	H_2O	pH 2-12										
B	pH 4-9.5 (different buffers to those in A row)											
C	Ionic strength (for pH 6-8)											
D	Pre-defined buffers					Additives						

Buffer optimization

	1	2	3	4	5	6	7	8	9	10	11	12
A	59.2°C	-	43.6°C	37.7°C	55.0°C	61.3°C	59.8°C	62.1°C	55.5°C	59.0°C	33.4°C	33.2°C
B	36.5°C	42.1°C	48.3°C	52.2°C	55.0°C	58.5°C	66.2°C	66.4°C	58.7°C	59.4°C	63.1°C	63.3°C
C	57.2°C	59.2°C	62.7°C	62.1°C	67.0°C	68.1°C	69.9°C	66.5°C	60.2°C	61.8°C	66.5°C	70.0°C
D	60.6°C	58.5°C	69.4°C	63.4°C	46.2°C	55.2°C	58.2°C	54.5°C	59.2°C	59.5°C	-	59.2°C



Original buffer **59.2**

vs.

Best buffer **69.9**

> 10°C difference !!!

Sample storage

- Depends on sample stability
- Freezing (phase transition) may decrease protein stability in solution

Avoid repeated freeze-thaw cycles !

- Fridge: 4 °C
- Freezer: – 20 °C, – 80 °C (cryo-protectants addition – glycerol)
- Lyophylization = Freeze-drying: water sublimation

Check sample quality BEFORE and AFTER storage !

Batch to batch quality check

- Enormous amount of variables in preparation process
- Two sample batches may not be the same
- Minimal tests desired to **verify sample quality**

Reproducibility crisis

- Based on 2016 poll with > 1500 scientists included:

  70 % were not able to repeat an experiment !

  50 % were not able to repeat at least one of their own experiments !!!

- Possible causes:
  - Selective choice of data (cherry picking)
  - Unsuitable experimental desing
  - Inappropriate data evalueation (statistics)

- It’s probable that partial problem is **insufficient characterization** of input material and procedures.

Source: nature.com

Summary

- **Sample quality** is crucial for downstream experiments

- Various sample **properties** to be checked
  - Identity
  - Purity
  - Homogeneity
  - Stability

- **Storage** and **buffer** optimization desired

Questions?

Biomolecular Interactions and Crystallography Core Facility

bic@ceitec.cz

bic.ceitec.cz

MUNI | CEITEC

CF Head

Josef Houser

- +420 549 492 527
- josef.houser@ceitec.cz