Purification of proteins for structural studies

Corso di Biocristallografia e Microscopia Elettronica

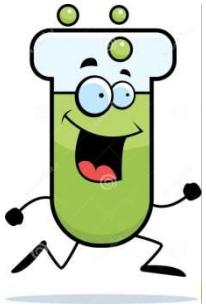
rdezorzi@units.it

Expression of recombinant proteins

Choice of the expression system

Choice of host and vector

Construct design



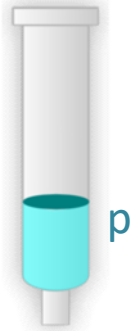
Cloning and vector ligation

Colony selection

Sequencing

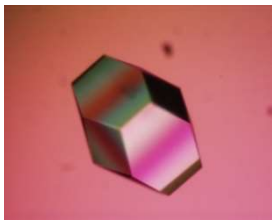
Expression protocol optimization

Protein purification



Sequence and activity check

Crystallization



Expression product

After expression of recombinant protein of interest,

CENTRIFUGATION to recover secreted proteins

or

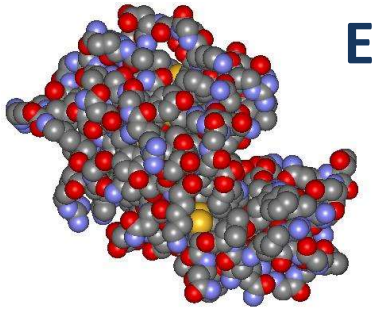
CELL LYSIS to recover cytoplasmic/membrane proteins or proteins in inclusion bodies

Lysate contains:
target protein,
together with soluble host proteins, membranes and membrane proteins, organelles, DNA , cytosolic matrix of the cell, **proteases**

Cell disruption

- Mammalian and insect cells: relatively easy to break, use of hypotonic solution
- Bacteria and yeast: harder to disrupt, especially yeast

Osmotic shock



Enzymatic method

Lysozyme used for bacterial hosts; zymolyase for yeast

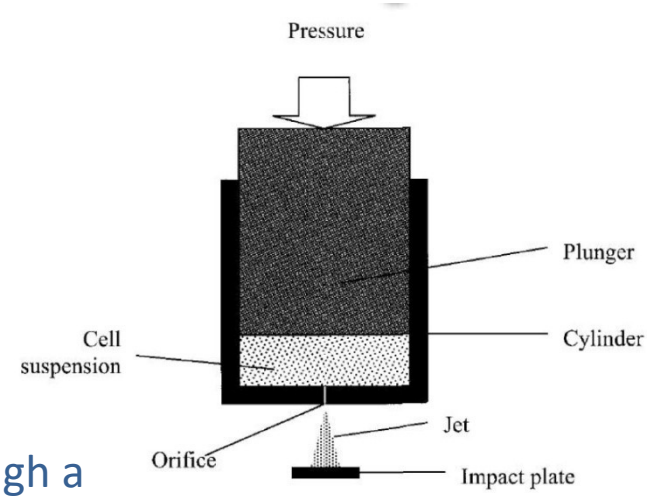
Sonication

Ultrasound used to create localized high pressure and break cell membranes

Bead mill

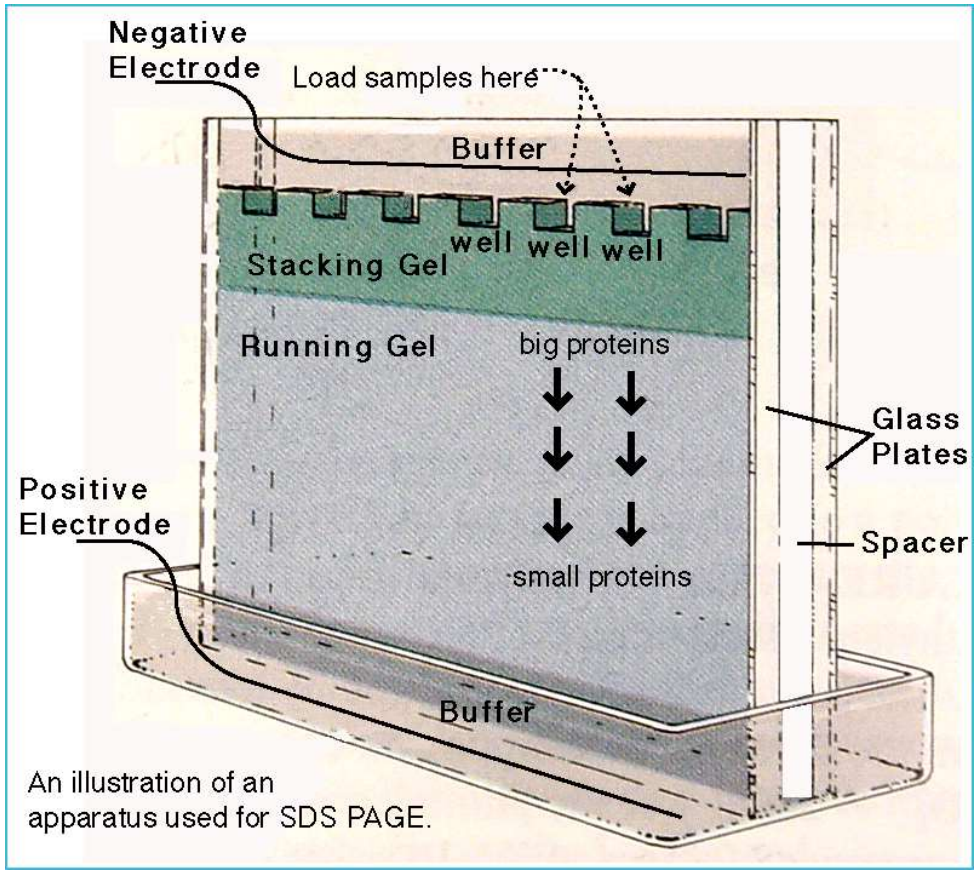
Mechanical disruption with beads and high frequency agitation

French press & Microfluidizer



Cell disruption by forcing media trough a narrow valve under high pressure

SDS-PAGE (PolyAcrylamide Gel Electrophoresis)

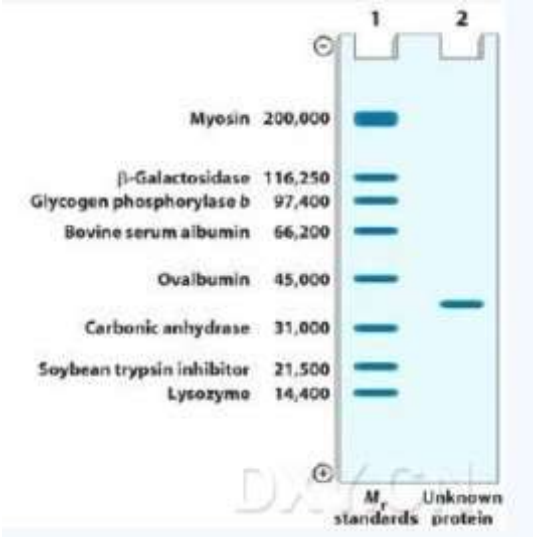


An illustration of an apparatus used for SDS PAGE.

Separation of proteins through an electric field.

Proteins are unfolded and covered by SDS, an anionic detergent.

Mobility of proteins depends on their dimension (≈ molecular weight)



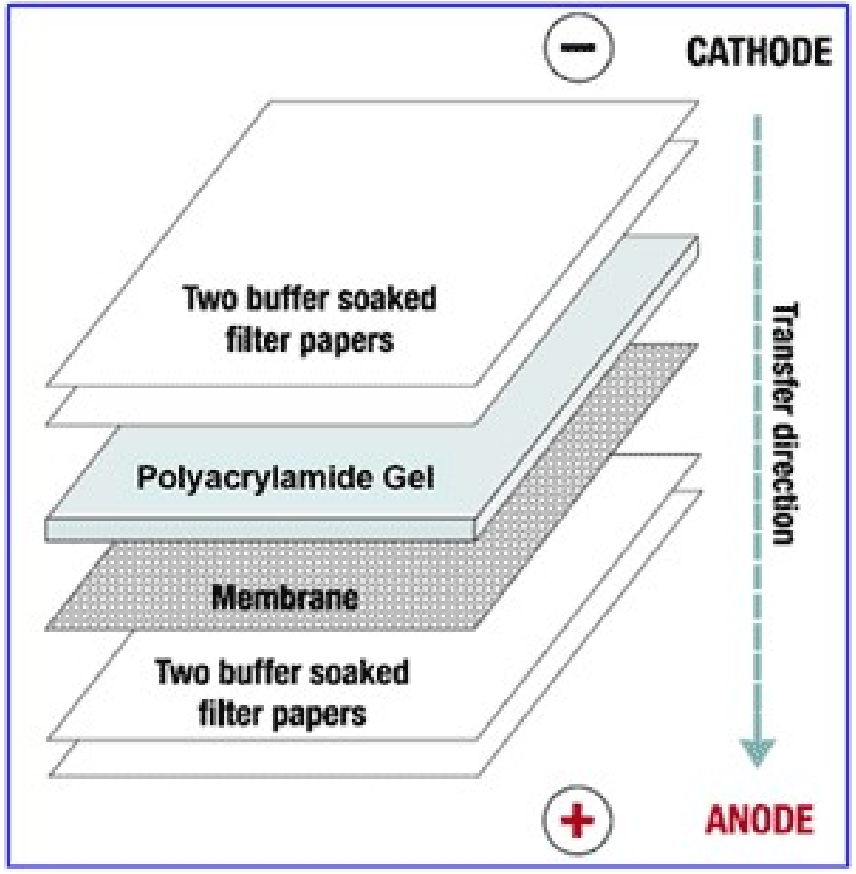
Protein detection:

Coomassie Blue staining

Silver staining

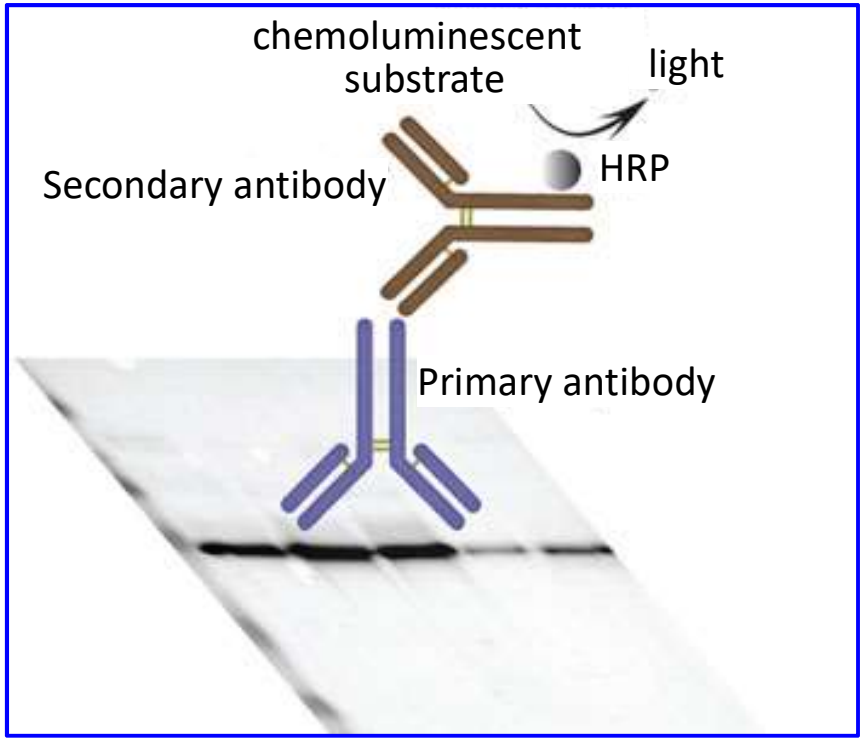
Western blotting

Western blotting



1. Proteins are transferred from polyacrylamide gel to nitrocellulose membrane

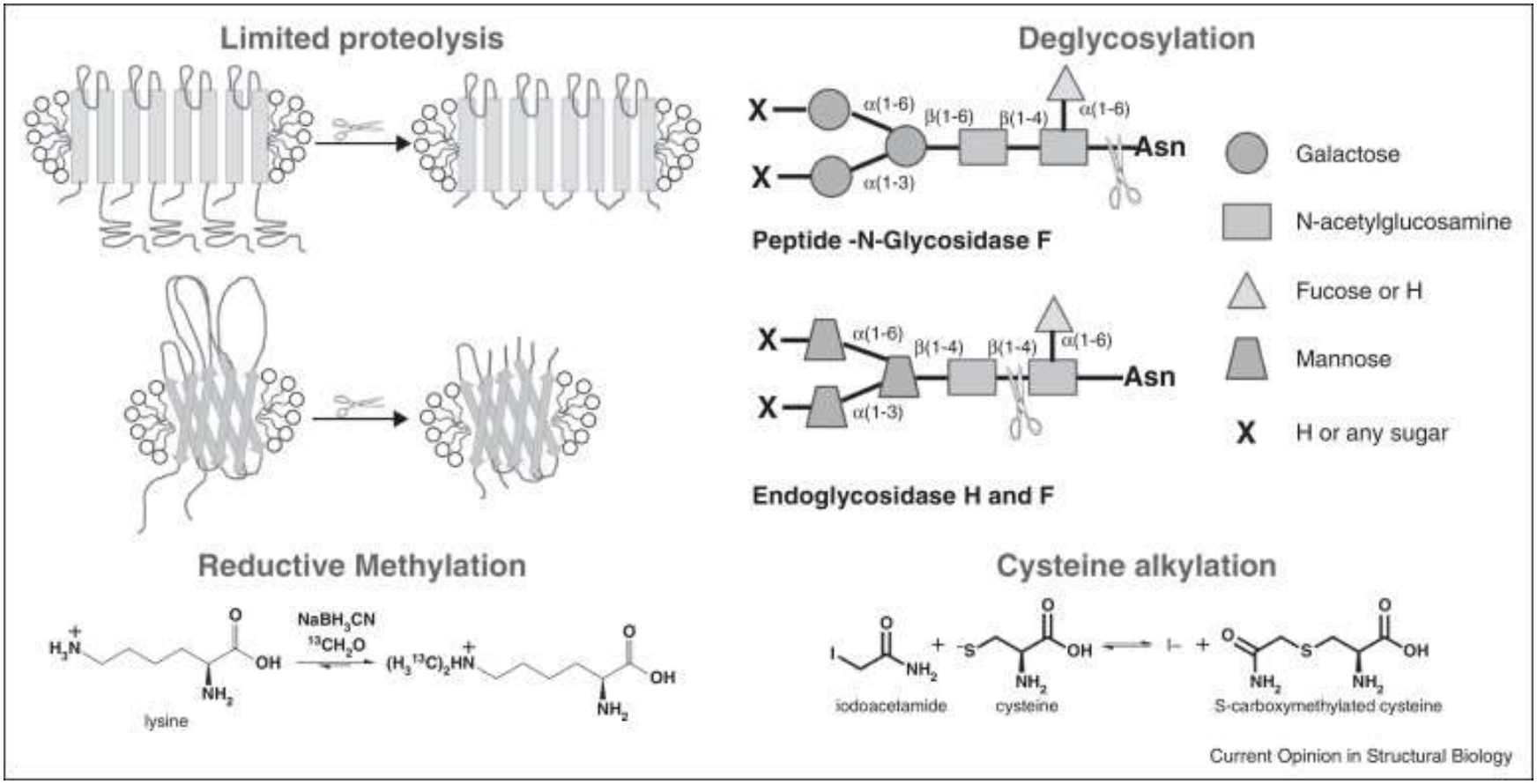
2. Binding of a primary antibody to the protein of interest (or its tag)
3. Binding of the secondary antibody, conjugated with enzyme (HRP)
4. Reaction producing chemoluminescence



Post-expression strategies

…recovery strategies for proteins difficult to crystallize…

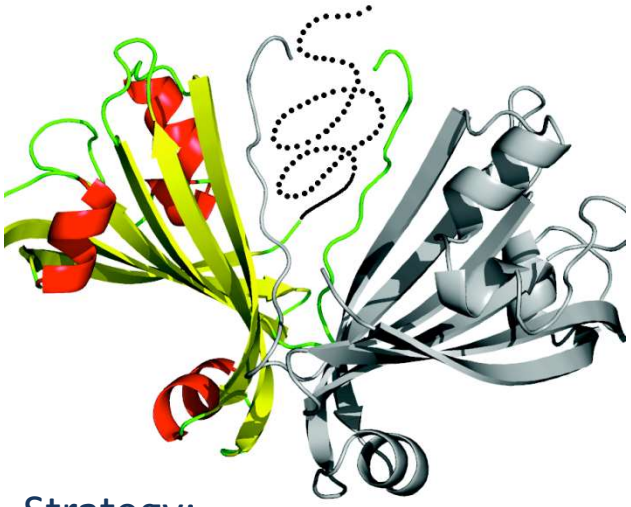
Modification of the sequence at the protein level (as opposed to modification of the construct) to improve stability, solubility, crystallizability, …



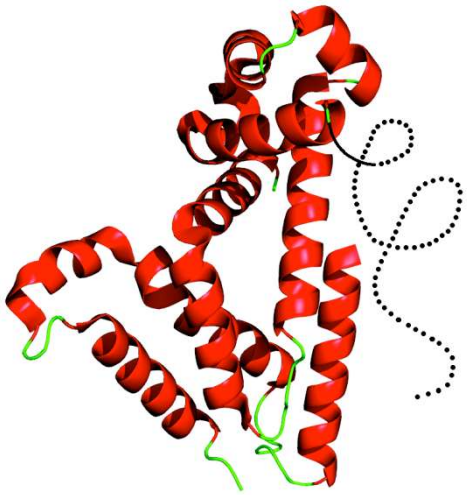
In addition… **Crosslinking** with glutaraldehyde or other chemical reagent

Limited proteolysis

Proteins are modular: they are composed of domains and flexible linkers…
Removal of flexible linkers/domains may improve crystallization.

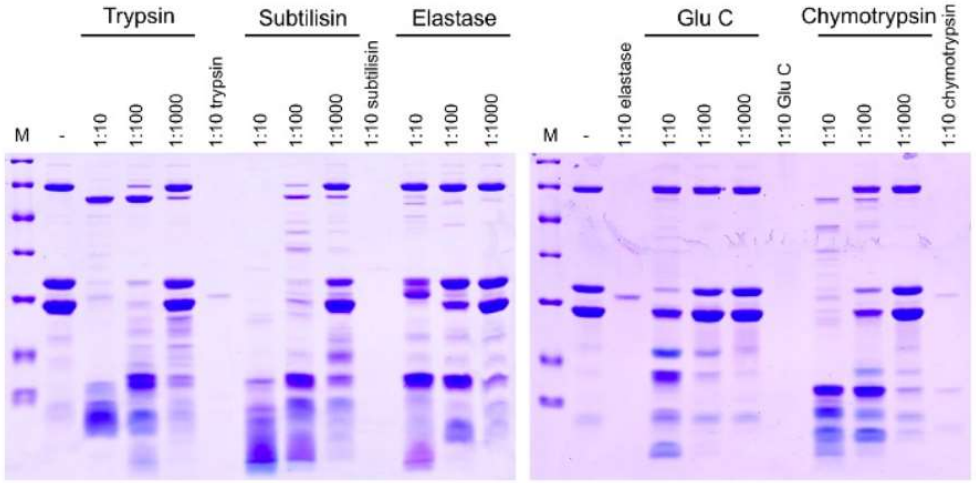


When the protein is folded, only exposed regions are cleaved by proteases, according to their selectivity. The remaining, fragmented protein remains folded.



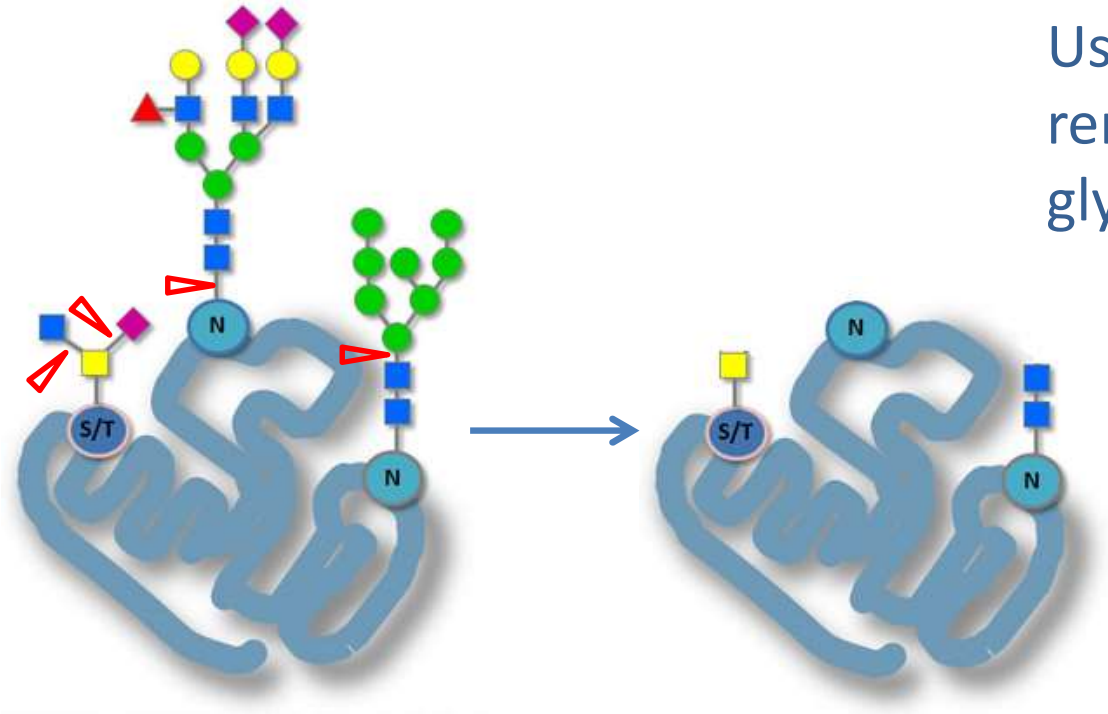
Strategy:

- Mix protein with proteolytic enzymes (trypsin, chymotrypsin, elastase, subtilisin, etc., possibly with tags for purification)
- Check proteolysis at different substrate:enzyme ratios and different incubation times by SDS-PAGE and/or Mass Spec



Deglycosylation by enzymatic digestion

Alternative to introduction of mutations that remove glycosylation sites



Use of glycosidases to remove all or part of the glycosydic chains

Glycosidases used:

- Peptide-*N*-Glicosidase F (cleaves on Asn residues, before first *N*-acetylglucosamine)
- Endoglycosidase H (cleaves after first *N*-acetylglucosamine in mannose-rich glycans)

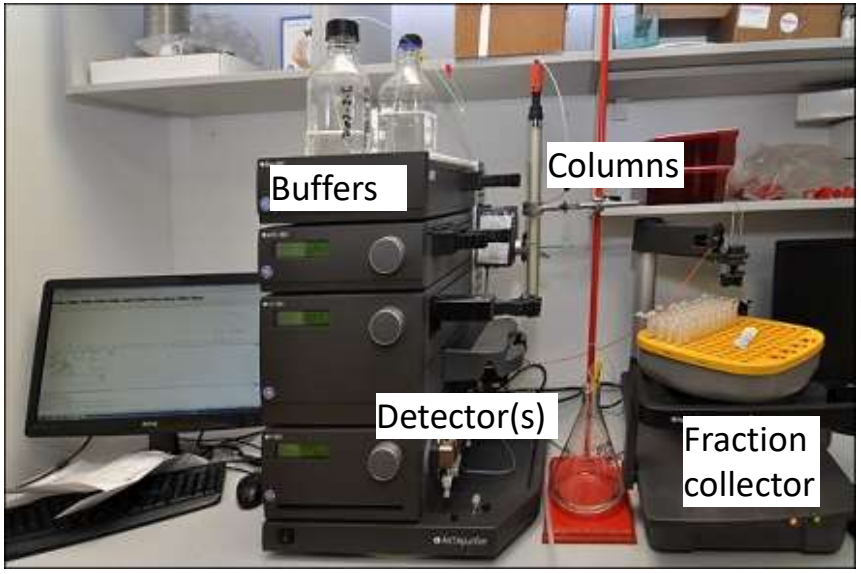
Purification: chromatographic methods

From cell lysate, containing many different proteins, separate target protein at high purity.

1) In batch methods:

Separation by gravity or spinning with centrifuge. Usually for small samples.

2) FPLC (Fast Protein Liquid Chromatography):



Usually more than one purification step is required, but every purification step decreases yield of pure protein.

1) Capture protein: isolate, stabilize and concentrate
2) Polish: to achieve high purity

Chromatographic methods

Tips to design optimal purification strategy:

1) Be aware of contaminants at each step of purification

2) Choose detection method (usually UV A_{280nm}, but also fluorescence, conductivity, light scattering, …)

3) Choose purity assessment method (SDS-PAGE, WB, MassSpec…)

4) Check stability of protein: pH, temperature, detergents, ionic strength, additives, organic solvents; sensitivity to proteases

5) Tags ⟶ Affinity chromatography

6) pI ⟶ Ion EXchange chromatography

} Concentrate protein

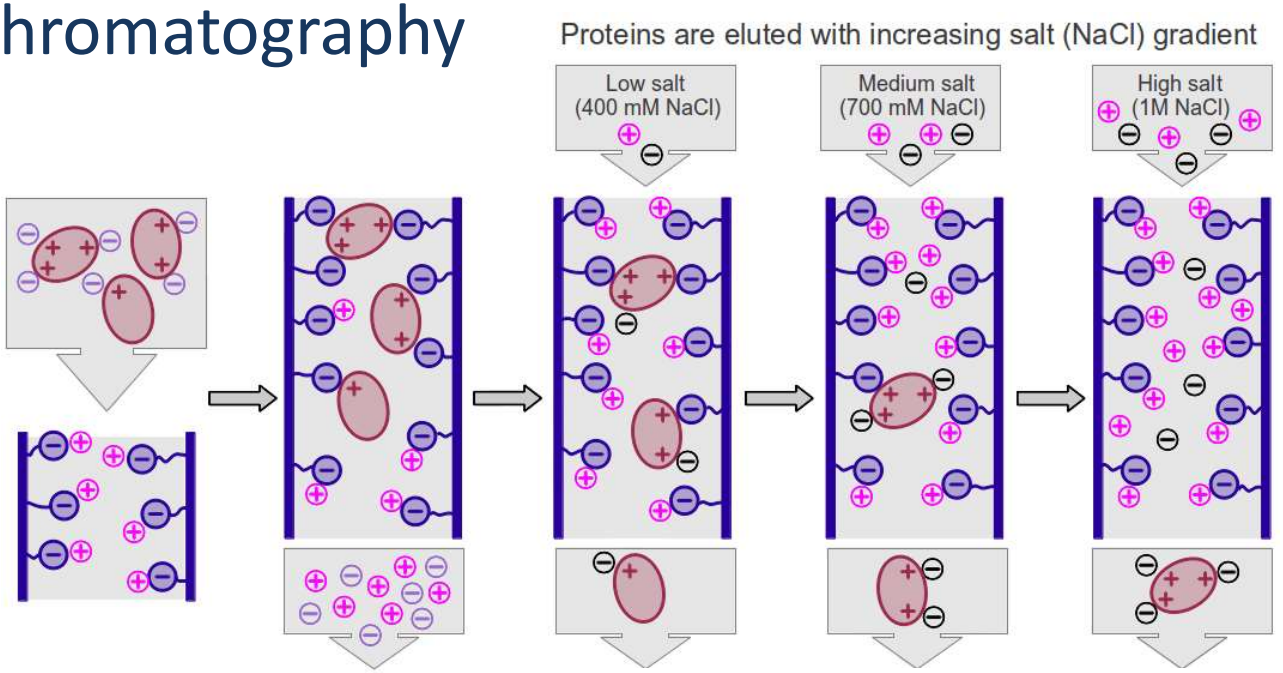
7) MW ⟶ Gel Filtration chromatography

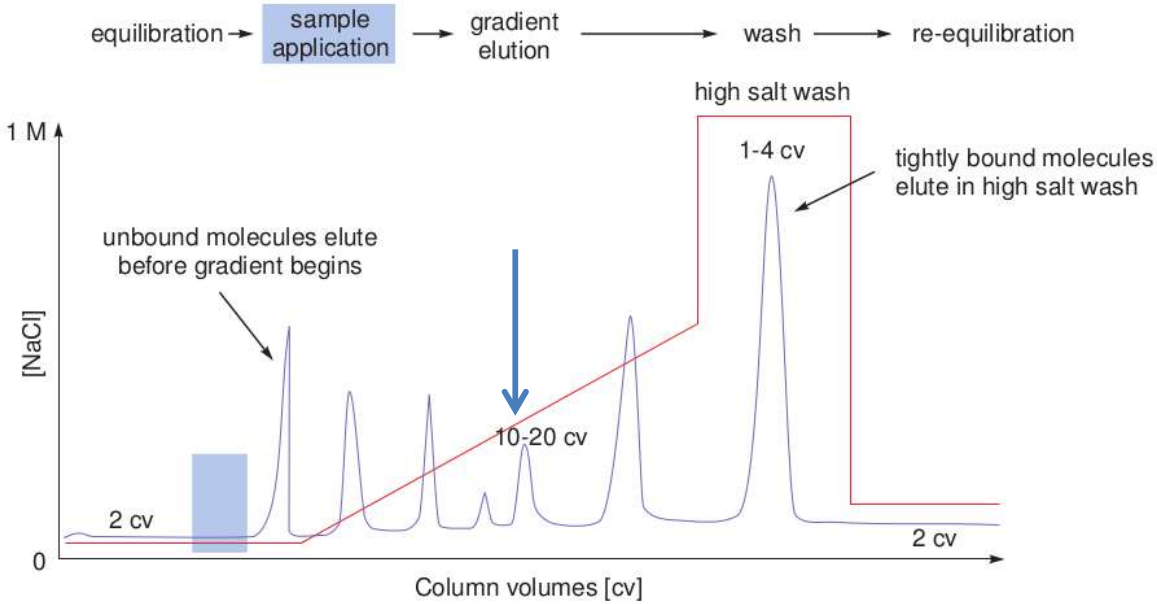
Ion exchange chromatography

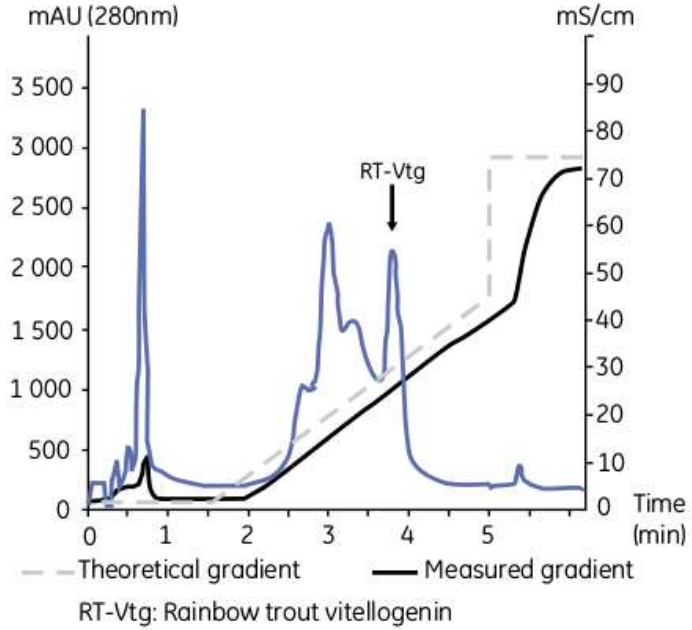
Anion (cation) exchange chromatography: separation by charge of the protein

Gradient elution allows separation of different proteins

Detector: Absorption @280nm

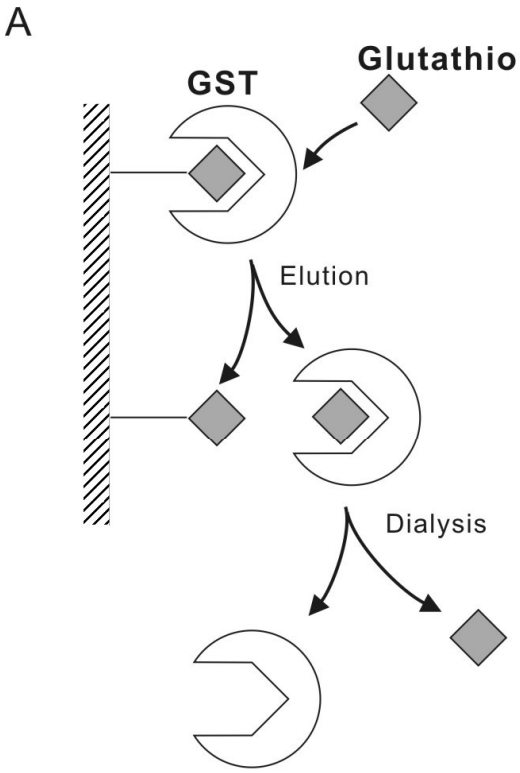




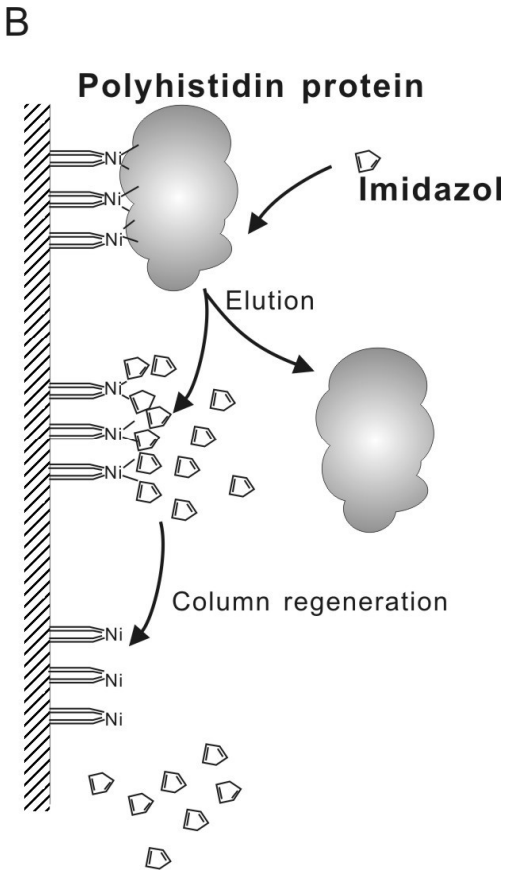


Affinity Chromatography

Affinity chromatography methods based on a wide range of biorecognition interactions



A. Purification of a protein with GST tag



B. Purification of a protein with His tag

(1) **enzymes and substrate** analogues, inhibitors, cofactors (e.g. for GST-tagged proteins)

(2) **antibodies and antigens** (e.g. FLAG-tagged proteins) CUSTOMIZABLE!!

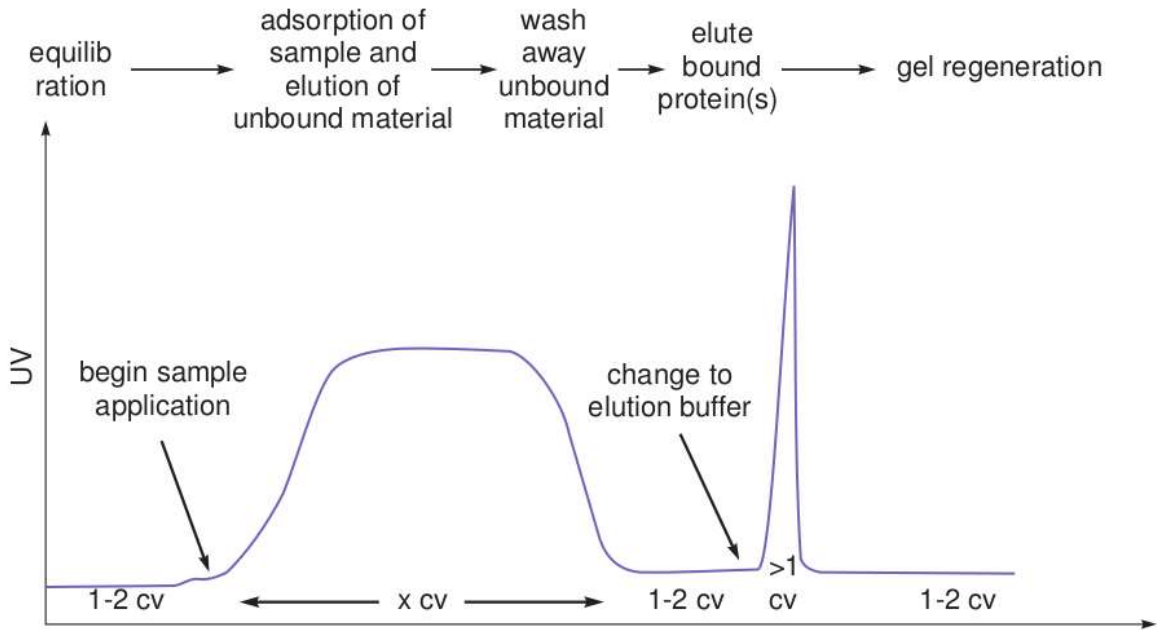
(3) **membrane receptors** and ligands

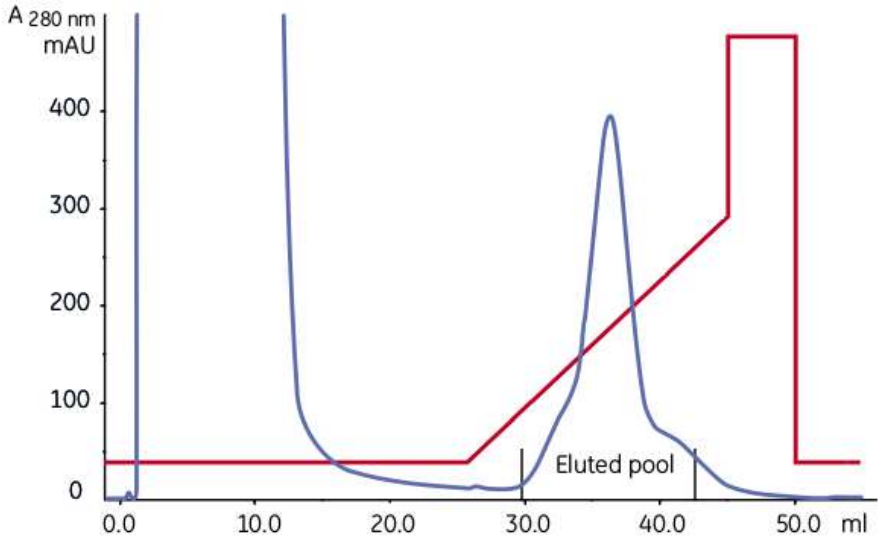
(4) **biological small molecules and their receptors** or carrier proteins (e.g. Strep-tagged proteins, MBP-tagged proteins)

(5) **metal ions** and proteins having **polyhistidine sequence** (His-tagged proteins)

IMAC

Immobilized Metal Affinity Chromatography

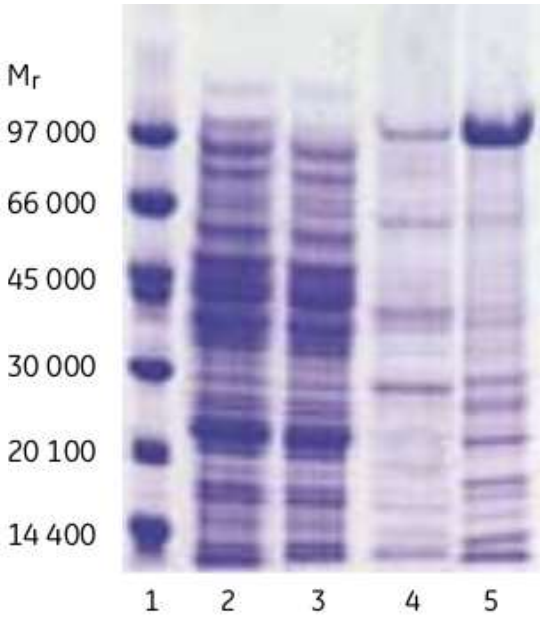




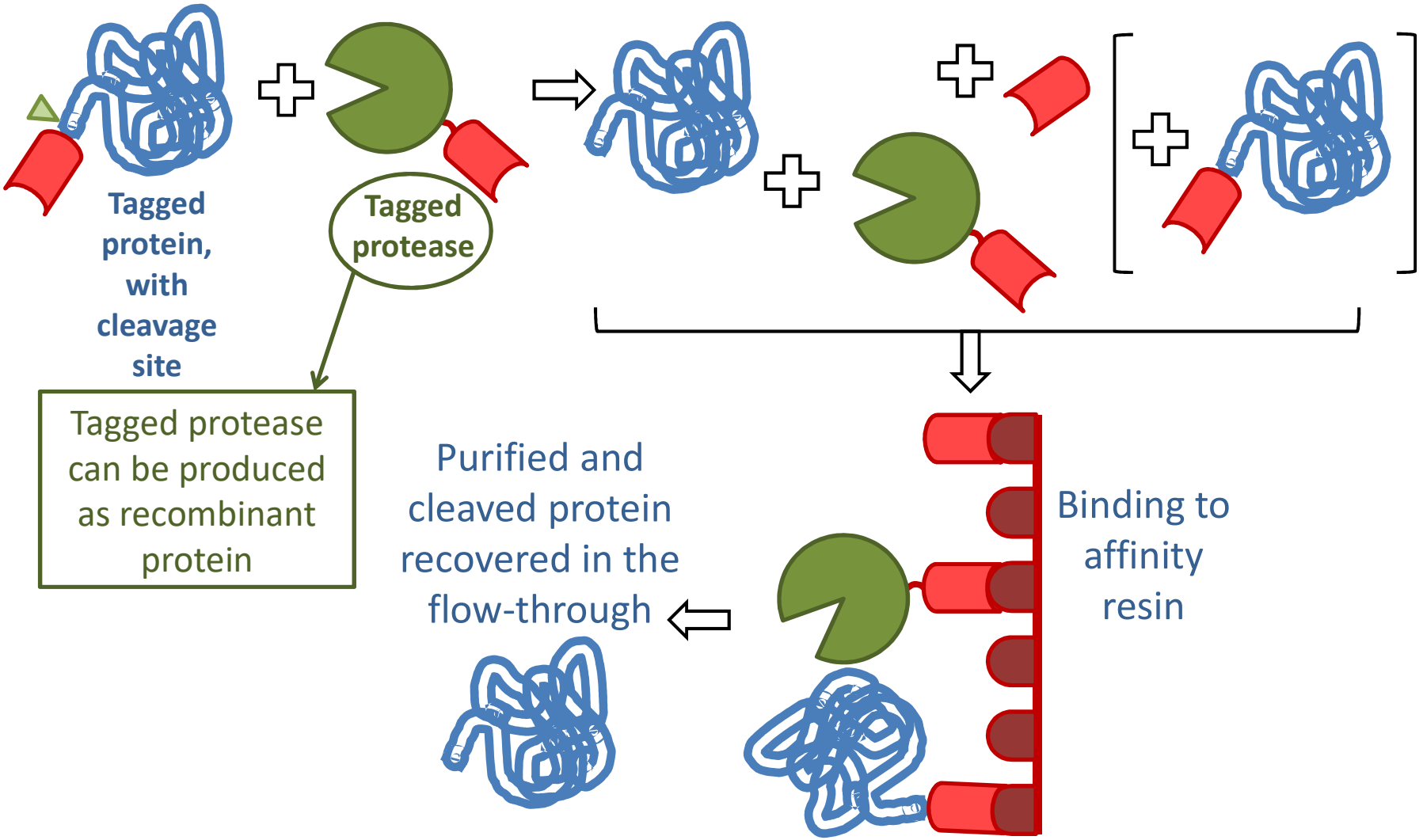
- Used for His tag purification
- Most used tag: simple and small
- Resin with immobilized Ni^{2+} or Co^{2+} metal ions
- Elution with Imidazole
- Tip: use imidazole in wash buffer to remove non specific binding
- Possible also in denaturing conditions

SDS-PAGE

Lane 1: LMW
Lane 2: *E. coli* extract
Lane 3: IMAC flow-through
Lane 4: Early IMAC fraction
Lane 5: IMAC eluted pool



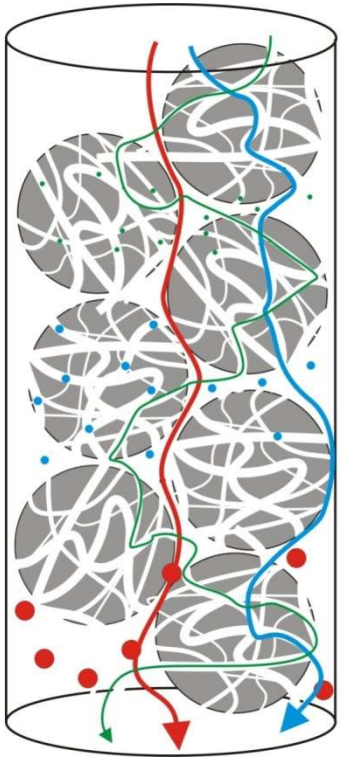
Reverse Affinity Chromatography

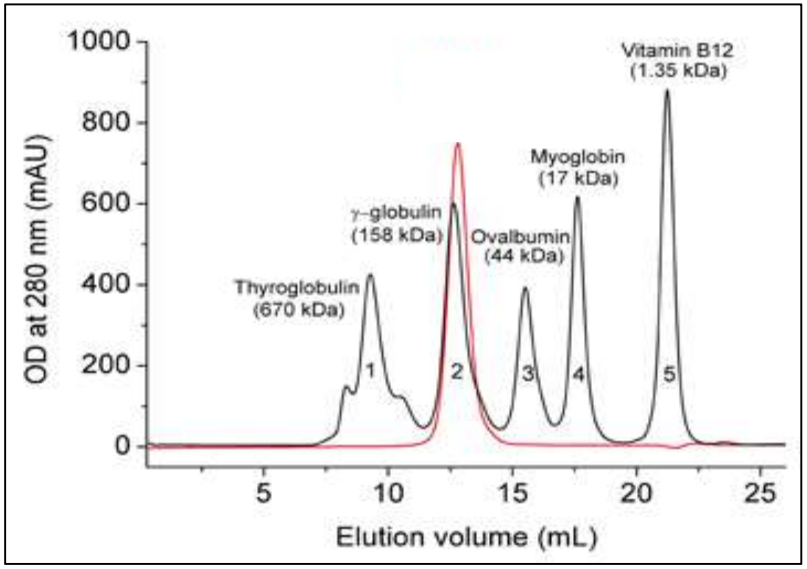


Gel Filtration Chromatography

Also known as Size Exclusion Chromatography (**SEC**)

Separation of proteins by **size**: larger proteins are eluted first, smaller proteins later

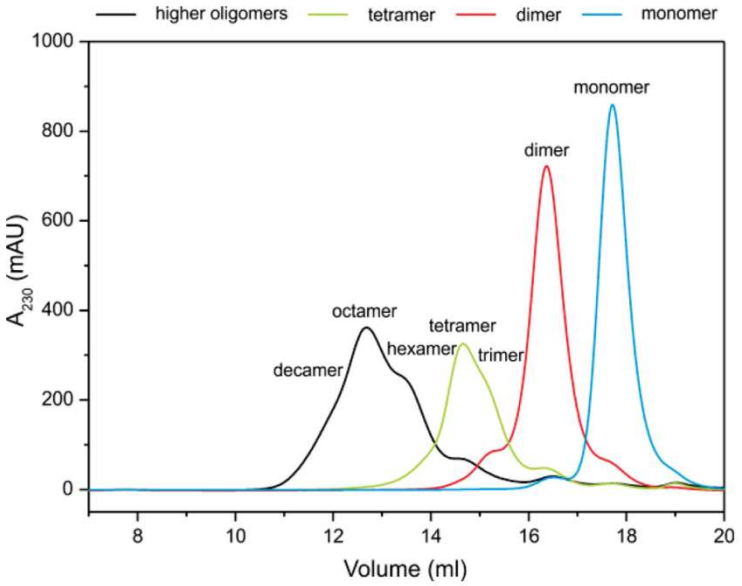




Usually, Gel Filtration is the **last step** of protein purification, before crystallization

Allows separation of aggregates or oligomers

Can also be performed for **qualitative determination** of the oligomeric state of a protein



Protein characterization

Protein parameter	Techniques
IDENTIFICATION & PURITY	SodiumDodecylSulfate PolyAcrylamide Gel Electrophoresis (SDS-PAGE) - Coomassie staining, Western Blot
CONCENTRATION	UV-VIS, Bradford assay, BCA assay
STABILITY	Thermo Stability Assay (TSA), Circular Dichroism (CD), Differential Scanning Calorimetry (DSC), UV-VIS
OLIGOMERIC STATE & AGGREGATION	Size Exclusion Chromatography (SEC) and Analytical SEC, Dynamic Light Scattering (DLS), Transmission Electron Microscopy (TEM), Sucrose Gradient Ultracentrifugation
CHEMICAL HETEROGENEITY	Mass Spectrometry, SDS-PAGE
CONFORMATIONAL HETEROGENEITY	TEM, NMR
FOLDING, BINDING TO LIGANDS	CD, DSC, Isothermal Titration Chalorimetry (ITC), Spectrofluorimetry, Deutorium Exchange Mass Spectrometry (DXMS)
ACTIVITY	Biochemical assays

Concentration: UV-VIS Spectrometry

Proteins absorb UV light with maximum within 275-280 nm

Quantification is obtained measuring absorbance at **280 nm** and is based on Lambert-Beer's law:

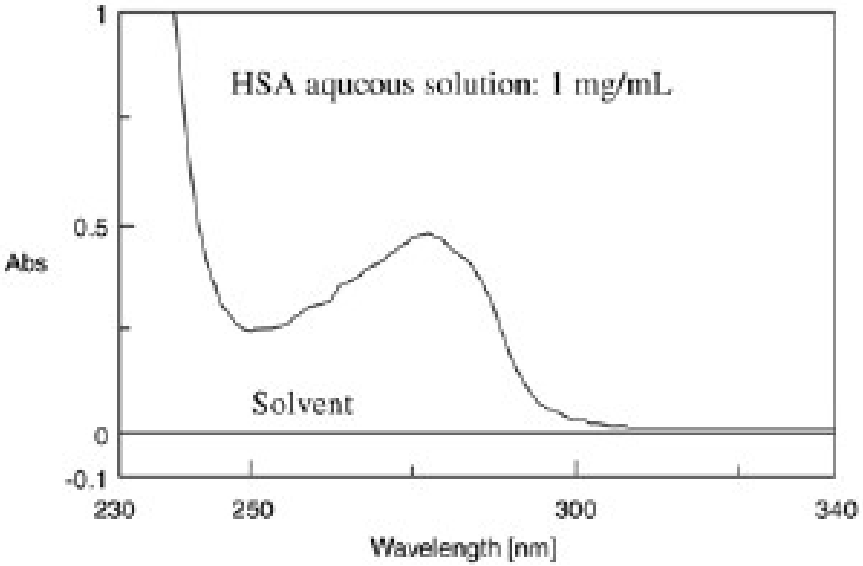
$$Abs = \varepsilon \cdot C \cdot \ell$$

Absorption is mainly due to the presence of Tryptophan, Tyrosine, Phenylalanine

Extinction coefficent can be calculated from the primary sequence:
http://www.expasy.org/

Buffer absorption is **always** subtracted.

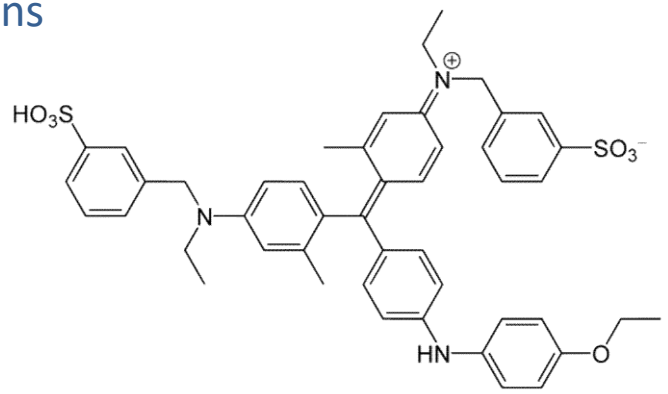
Instead of using a cuvette, for expensive samples absorbance can be measured on a drop of protein (1-2 μL)



Concentration: Bradford assay

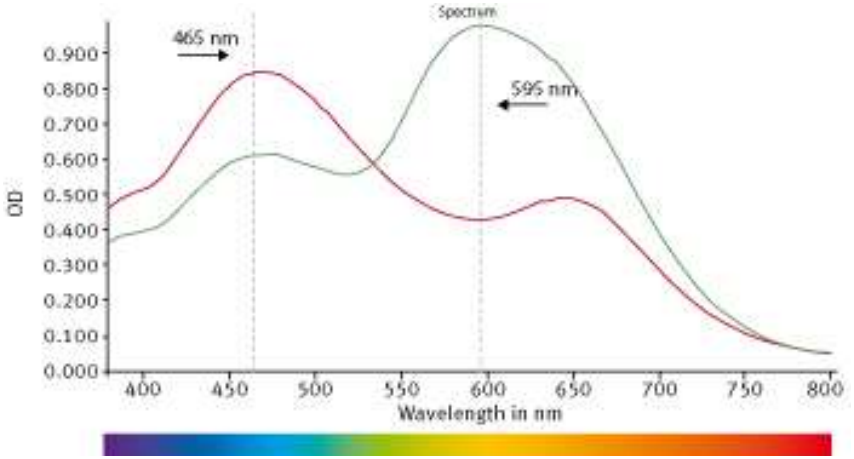
Based on binding of proteins to **Coomassie Brilliant Blue G-250**

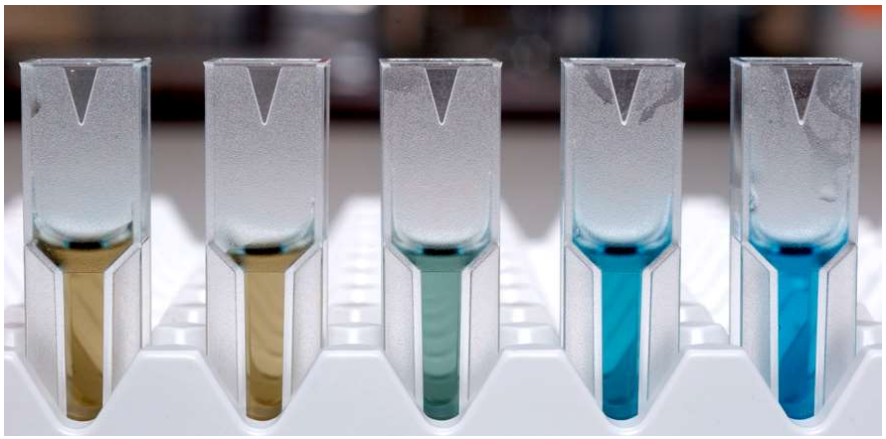
- Binding due to ionic and hydrophobic interactions
- Dependent on protein nature
- Approximate concentration can be evaluated using BSA or γ-globulin as standard



Protein concentration measured using the **color shift** of the dye

- Absorbance at **595 nm** is measured and compared with a standard solutions

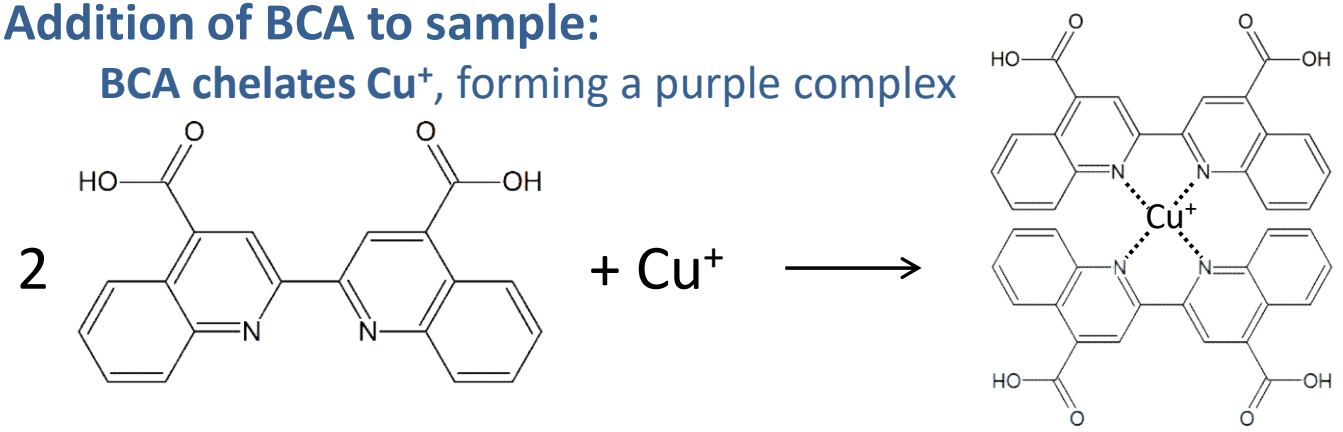




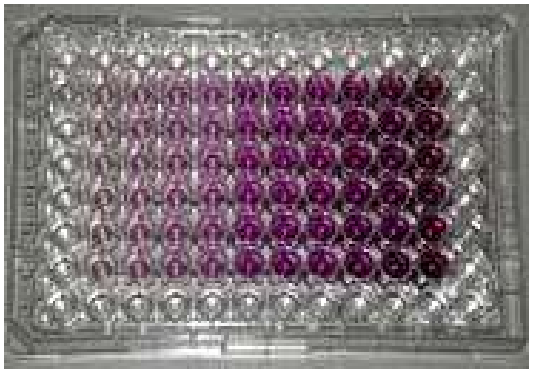
Concentration: Bicinchoninic Acid (BCA) Assay

1. **Addition of Cu^{2+} to protein sample:**
   - **Peptide bond** is oxidized and reduces **Cu^{2+} to Cu^{+}**: independent from protein!!
   - Temperature dependent reaction
   - $[Cu^{+}]$ proportional to number of peptide bonds, ~ mass of protein in solution

2. **Addition of BCA to sample:**
   - **BCA chelates Cu^{+}**, forming a purple complex

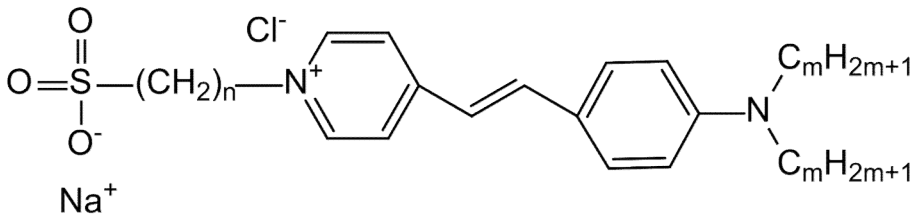


3. **Quantification obtained by absorbance measurement at 562 nm:**
   - Standardization using BSA is advisable

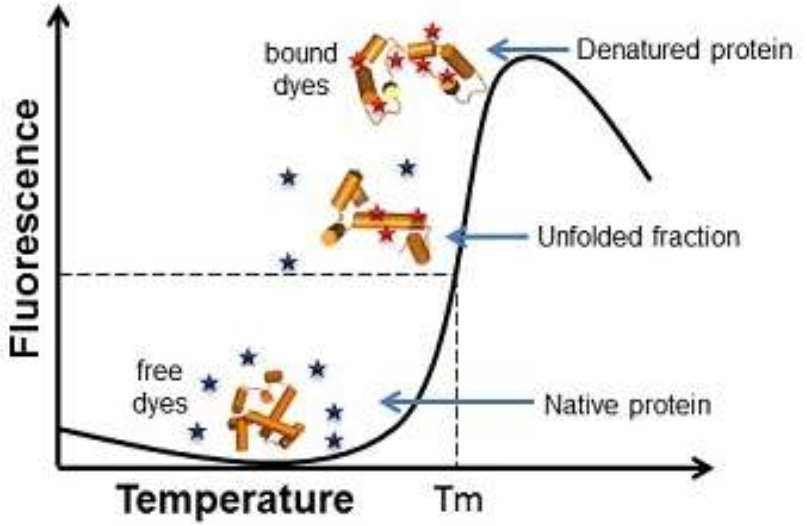


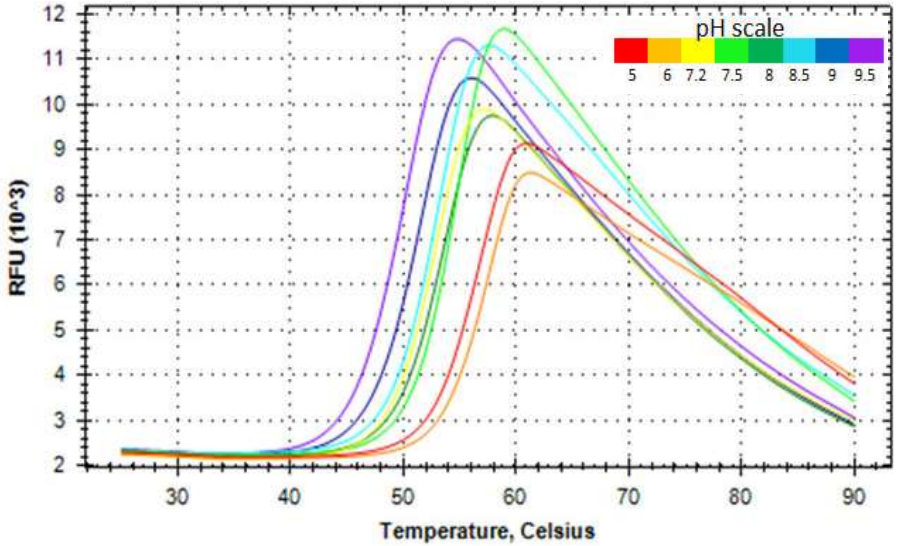
Stability: Fluorescent Thermal Shift Assay (TSA)

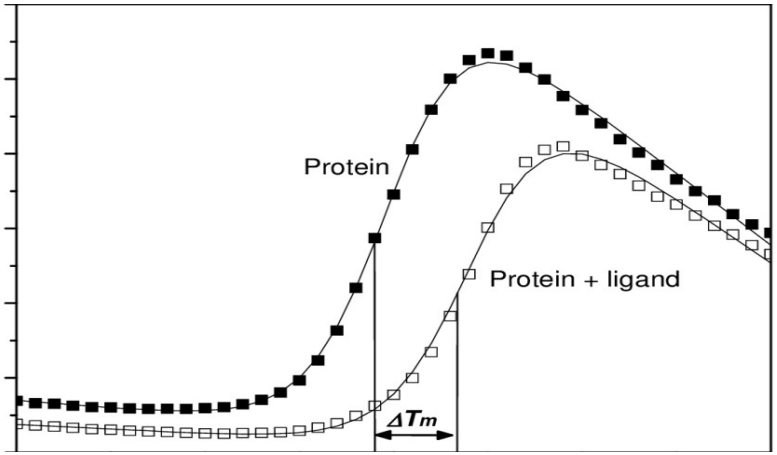
Sypro Orange dye: fluorescence only when the dye is bound to the hydrophobic surfaces of the protein. Fluorescence is quenched by water.



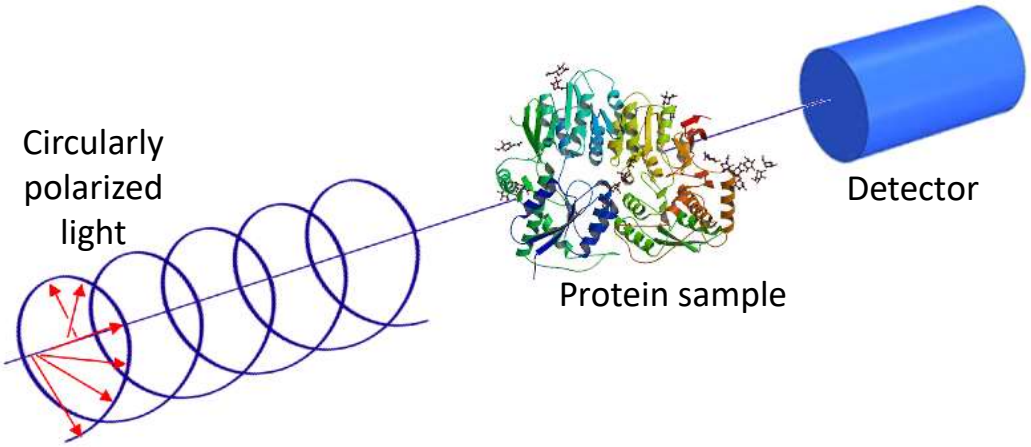
- Protein stability: **fluorescence vs temperature**
- Optimization of protein buffer/conditions: **fluorescence vs different buffers**
- Analysis of ligand binding: **comparison between melting curve of *apo* protein and protein/ligand solution**



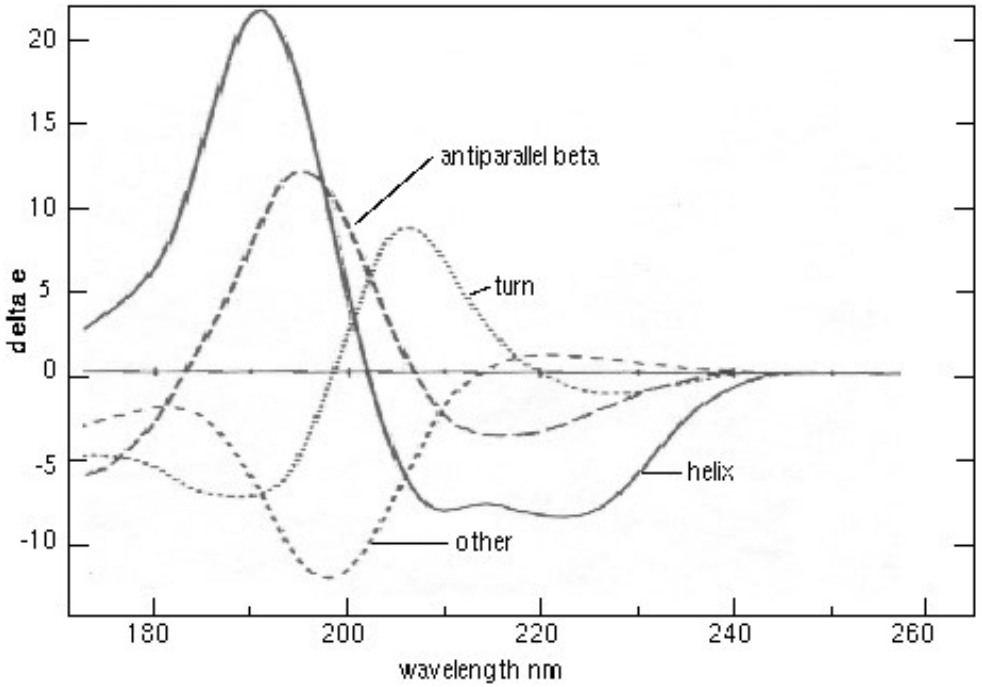


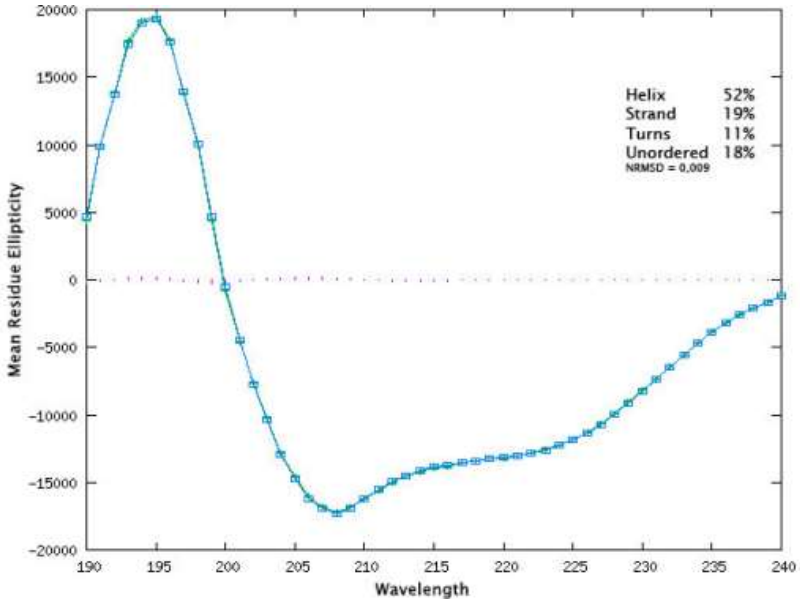


Stability & Protein folding: Circular Dichroism (CD)



Absorption of Left and Right Circularly Polarized (LCP and RCP) light is measured.
Ellipticity measured difference between absorption of LCP and RCP.
Ellipticity is plot against wavelength.

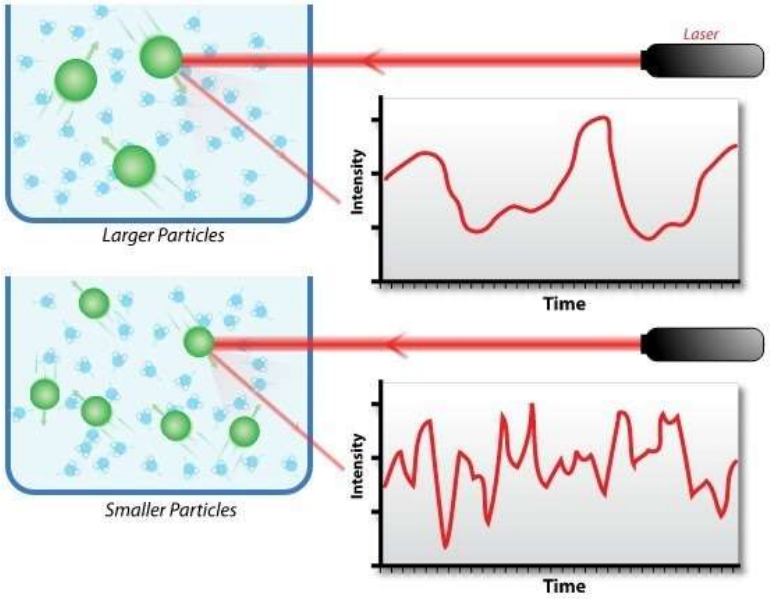




CD spectrum gives information on:

- Folding of protein
- **Stability & melting temperature**
- Conformational changes

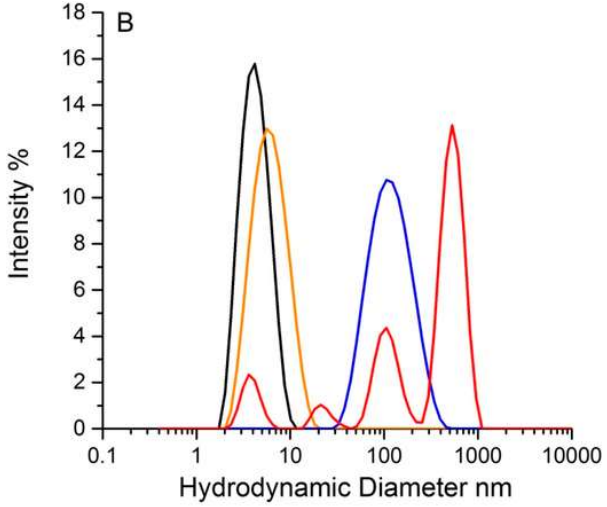
Oligomeric state:
Dynamic Light Scattering (DLS)



Measurement of fluctuations of scattered light in time.

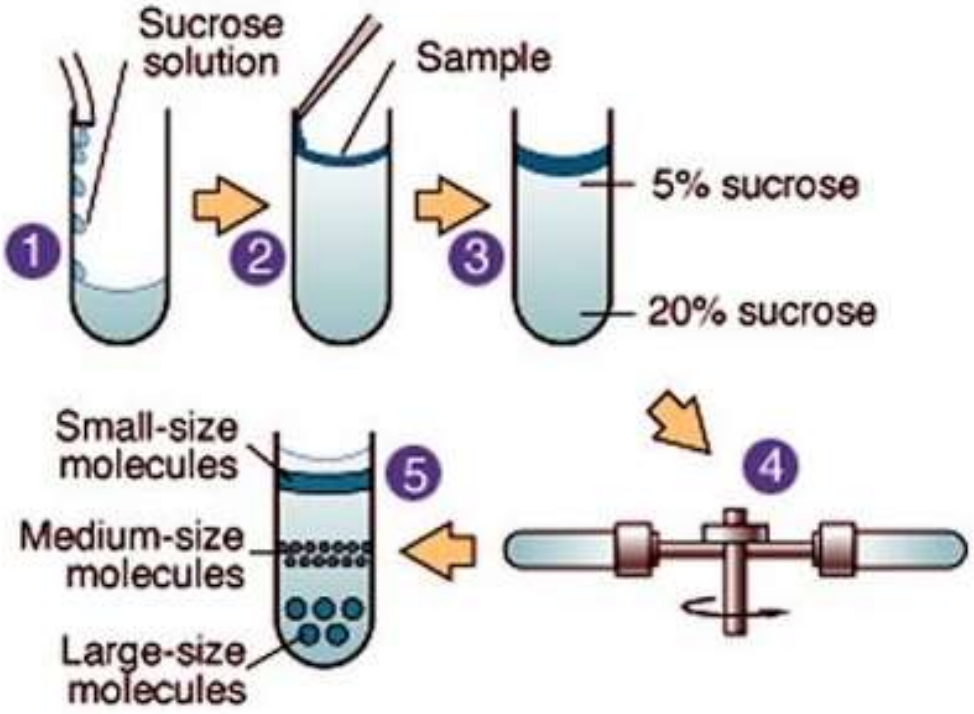
Autocorrelation function gives the diffusion coefficient of particles in solution.

If temperature and viscosity of the solution are given, Stokes-Einstein equation allows to calculate hydrodynamic radius of the particle.



Used to determine:
- Dimension of a protein (also combined with SEC)
- Aggregation state
- Oligomerization state
- Interactions between proteins

Oligomeric state:
Sucrose Gradient Ultracentrifugation

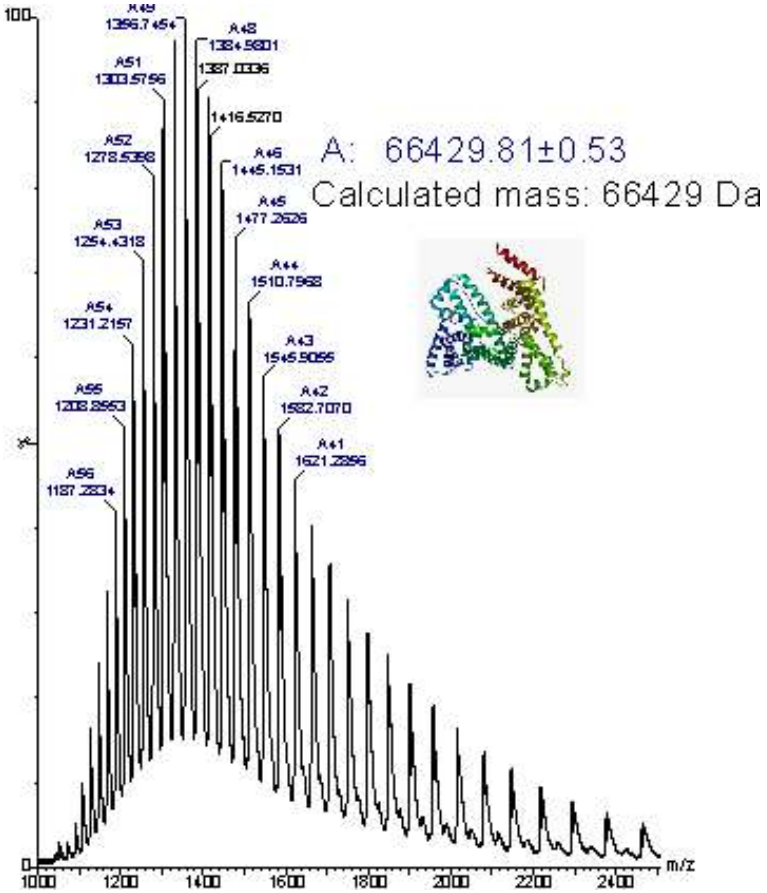


Also known as
Equilibrium Gradient Centrifugation

Separates particles by **density**: particles reach a level in the centrifuge tube in which their density matches the sucrose density

Requires high *g* (centrifugation force):
> 150.000 *g*

Used to determine:
- Molecular Weight (MW)
- Oligomerization state
- Protein-protein interactions
- Shape of the protein



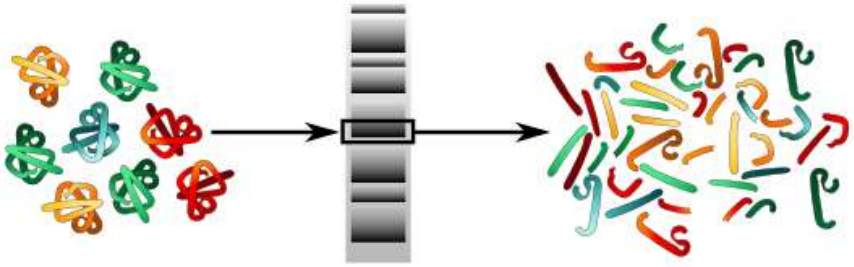
Chemical heterogeneity: Mass Spectrometry (MS)

Integrity of the protein (no proteolytic activity) and presence of post-translational modifications can be detected

Mild ionization methods (ESI or MALDI), charging of the protein without fragmentation

Deconvolution of the spectrum allows determination of exact mass of the protein

MassSpec after enzymatic digestion:



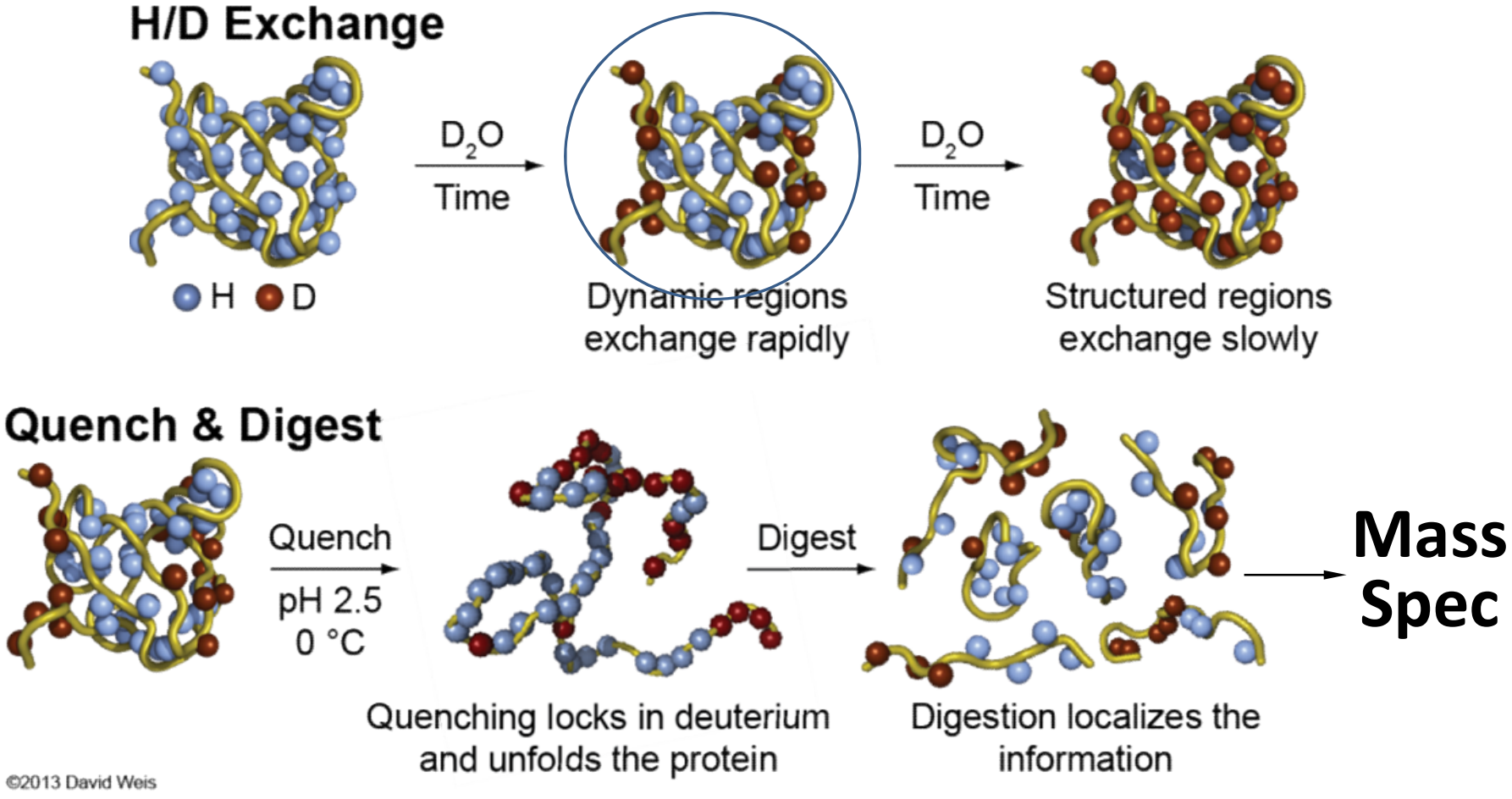
LC-MS →

Database of protein MS spectra or sequence-based expected digestion products

↓

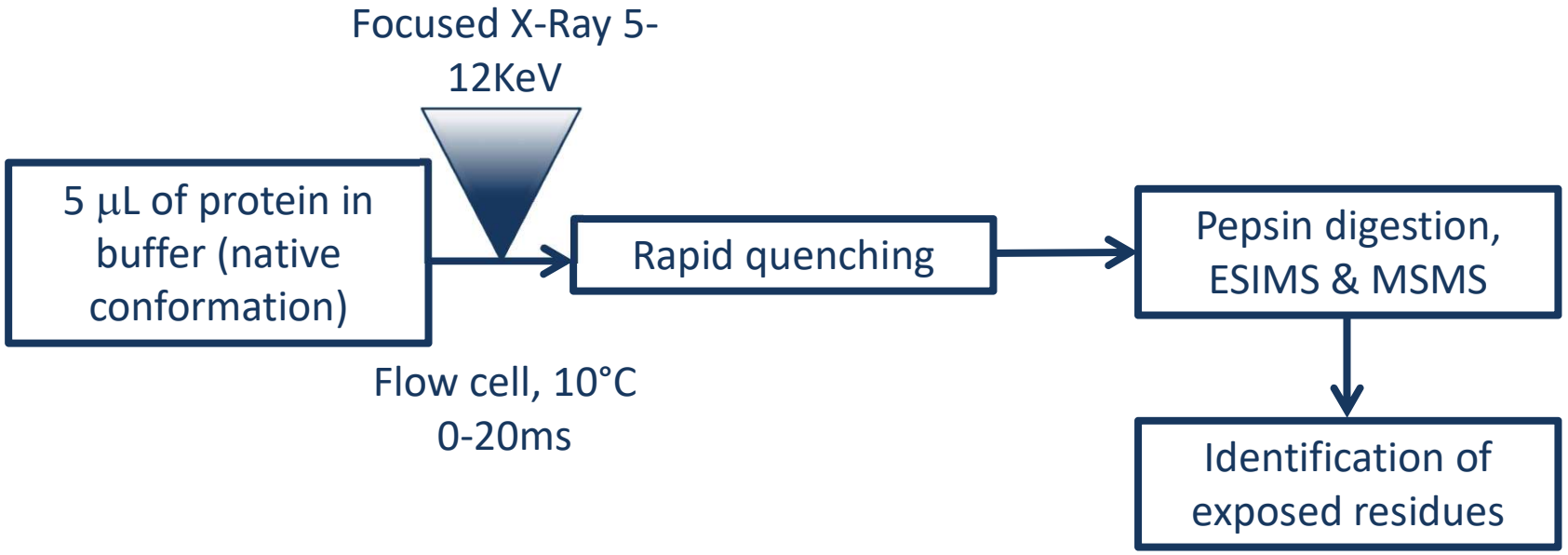
PROTEIN IDENTIFICATION & LOCALIZATION OF POST-TRANSLATIONAL MODIFICATIONS

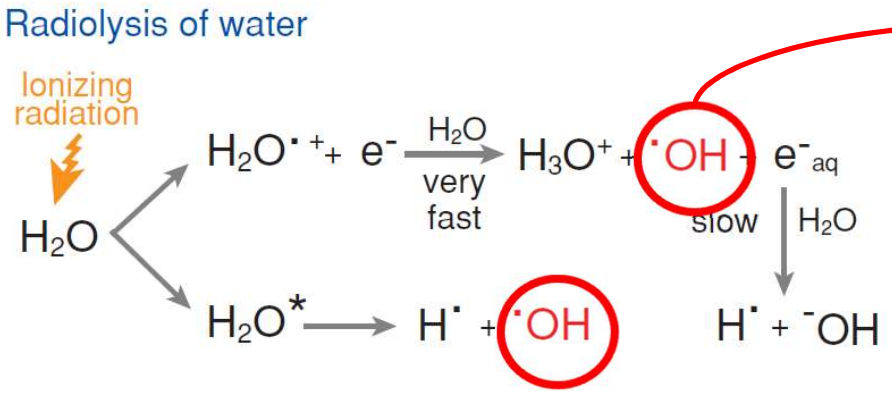
Surface mapping:
Deuterium Exchange Mass Spectrometry (DXMS)



Surface regions and flexible linkers are more affected by deuterium exchange.

Surface mapping:
Mass Spectrometry Radiolytic Footprint

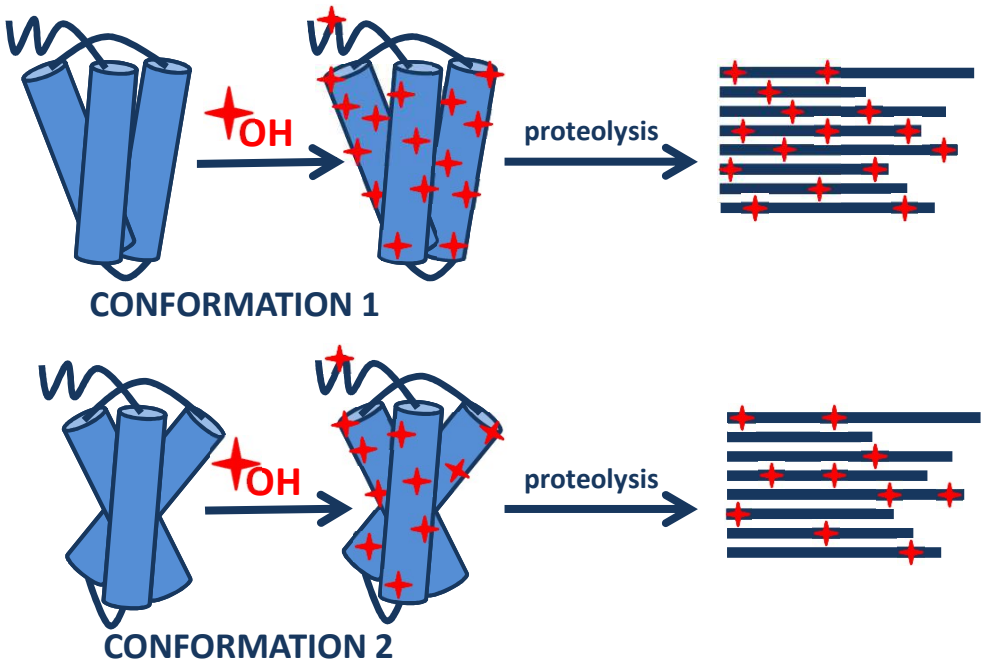




Reactivity of side chains:

Cys > Met > Trp > Tyr > Phe > Cystine > His > Leu ~ Ile > Arg ~ Lys ~ Val > Ser ~ Thr ~ Pro > Gln ~ Glu > Asp ~ Asn > Ala > Gly

Conformational change:
Mass Spectrometry Radiolytic Footprint



Different rates of radiolysis for the two conformations are correlated with different exposition to the solvent of residues involved.

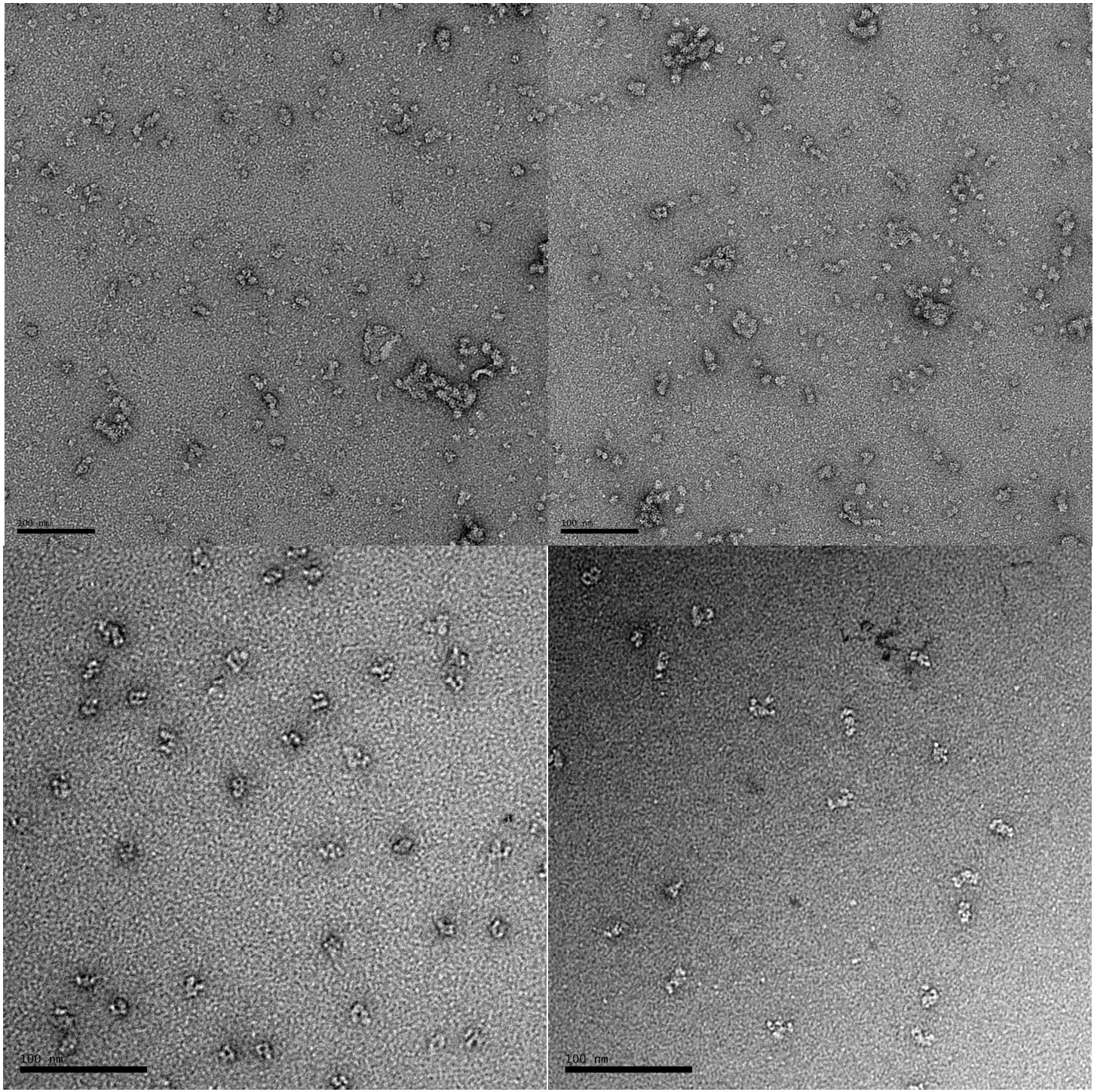
Mapping of residues involved in change of conformation.

Conformational heterogeneity: Negative Staining Transmission Electron Microscopy NS-TEM

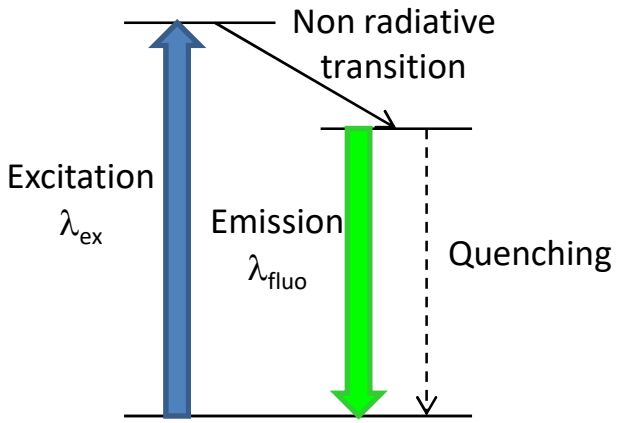
Aggregation state

Homogeneity of the sample

Conformational homogeneity


100 nm
100 nm
100 nm
100 nm

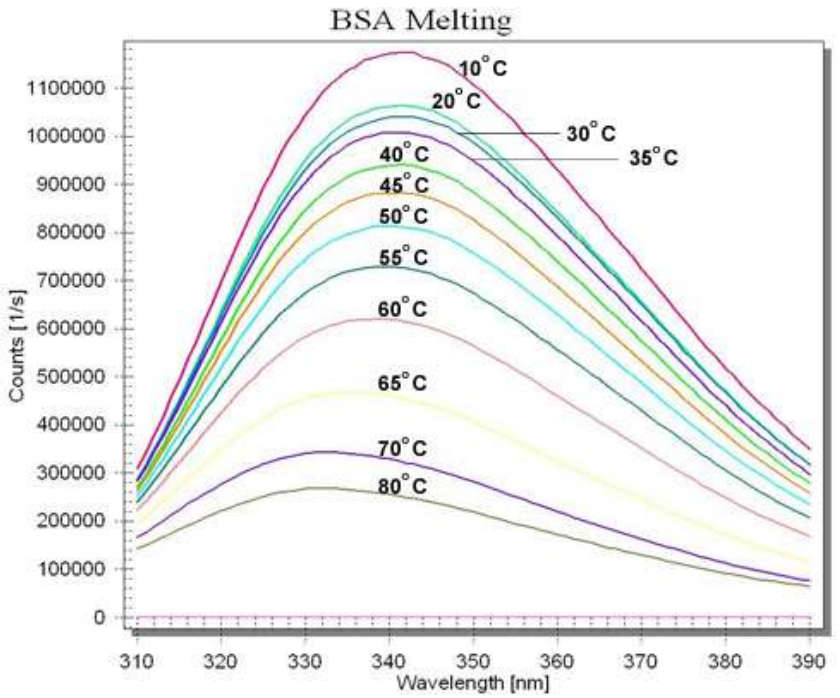
Ligand binding: Spectrofluorimetry



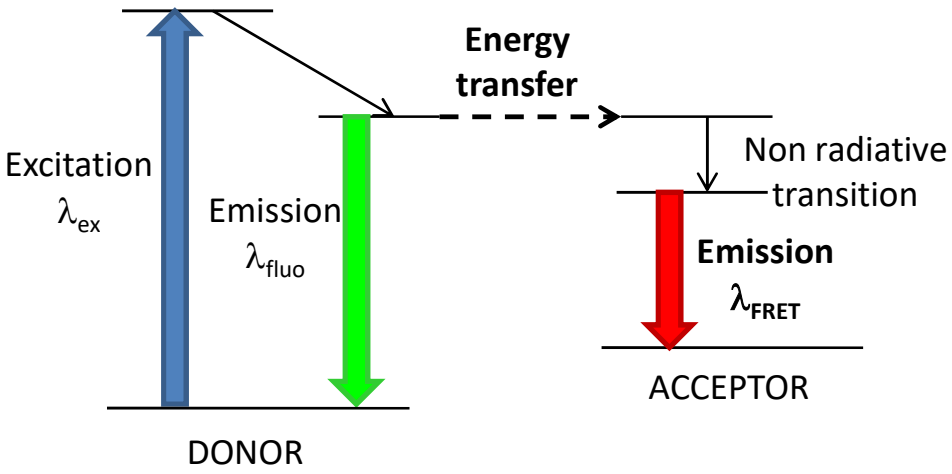
Excitation at a certain wavelength λ_{ex}
Emission at λ_{fluo} ($> \lambda_{ex}$)

Aromatic residues display fluorescence
Use of fluorescent probes that increase/quench their fluorescence due to interaction with the protein

Used for:
- Study of pH dependence
- Conformational changes
- Accessibility of a specific site of the protein (with fluorophore)
- Thermal stability
- FSEC (Fluorescence Size Exclusion Chromatography)
- …

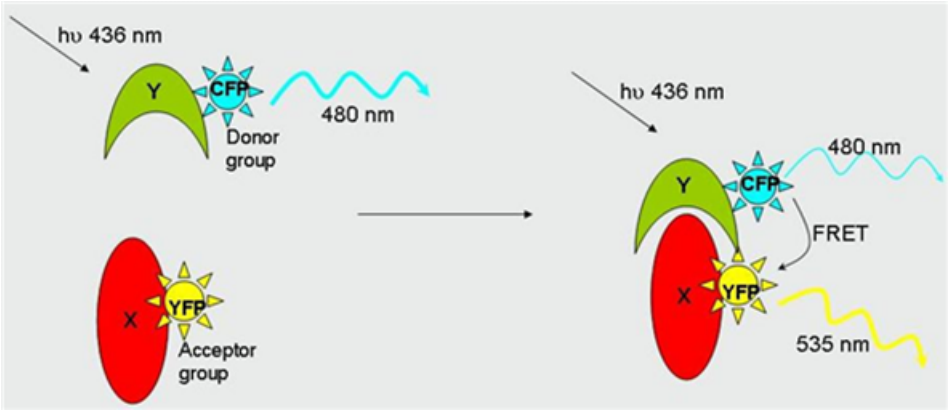


Fluorescence Resonance Energy Transfer (FRET)



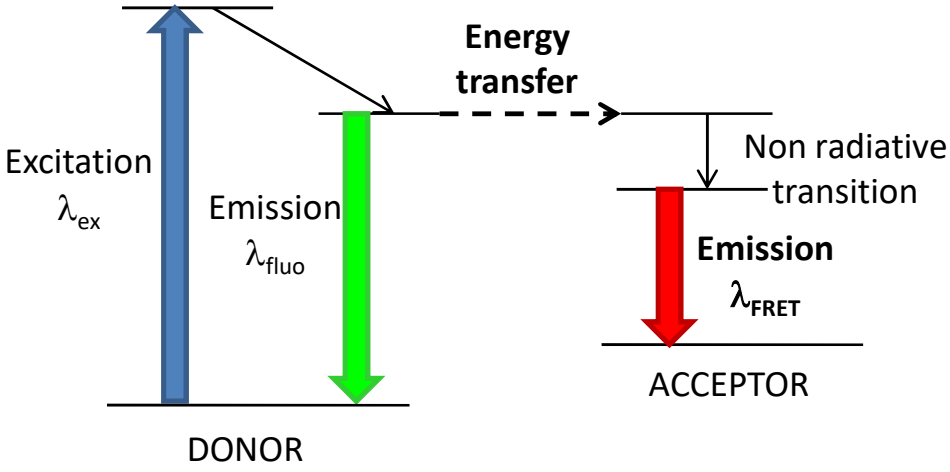
Energy transfer between fluorophores based on radiative dipole-dipole coupling

Depends on:
- **Distance** between donor and acceptor
- **Energy overlap** of donor and acceptor
- Relative **dipole orientation** of donor and acceptor



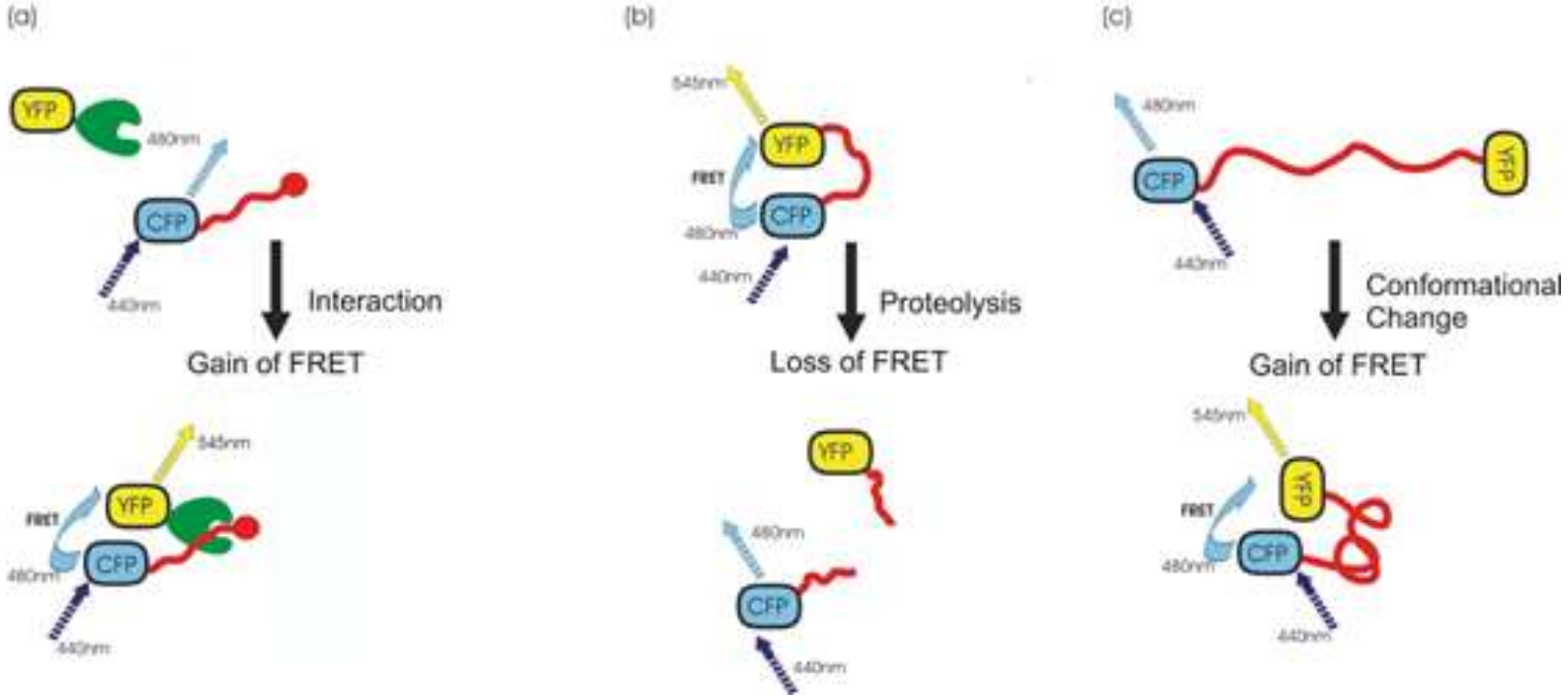
Used to prove:
- Domain-domain interaction
- Protein-protein interaction
- Spatial arrangement of domains

Fluorescence Resonance Energy Transfer (FRET)



Energy transfer between fluorophores based on radiative dipole-dipole coupling

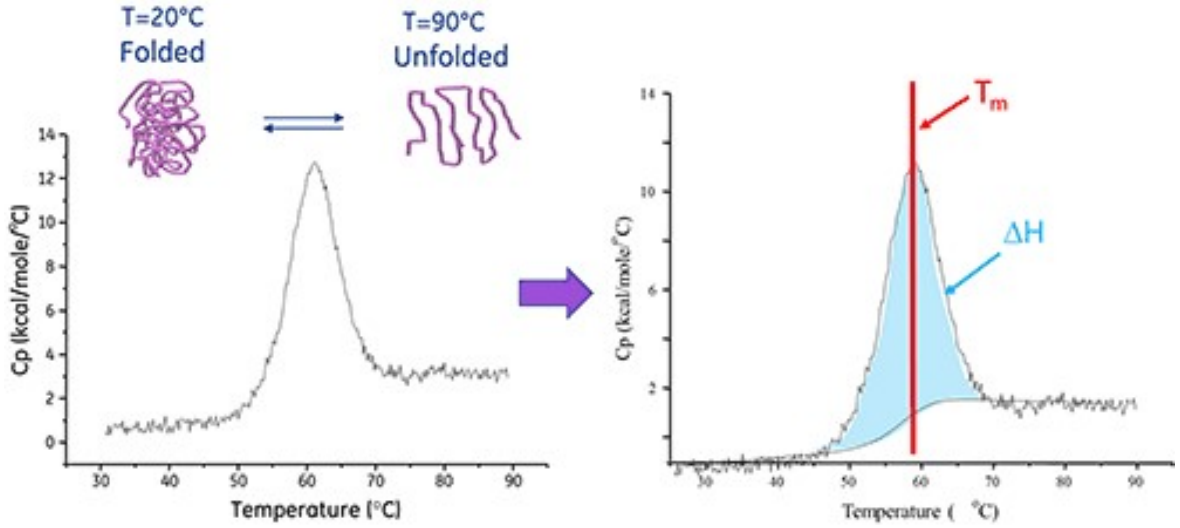
Depends on:
- **Distance** between donor and acceptor
- **Energy overlap** of donor and acceptor
- Relative **dipole orientation** of donor and acceptor



Stability, folding and ligand binding: Differential Scanning Calorimetry (DSC)

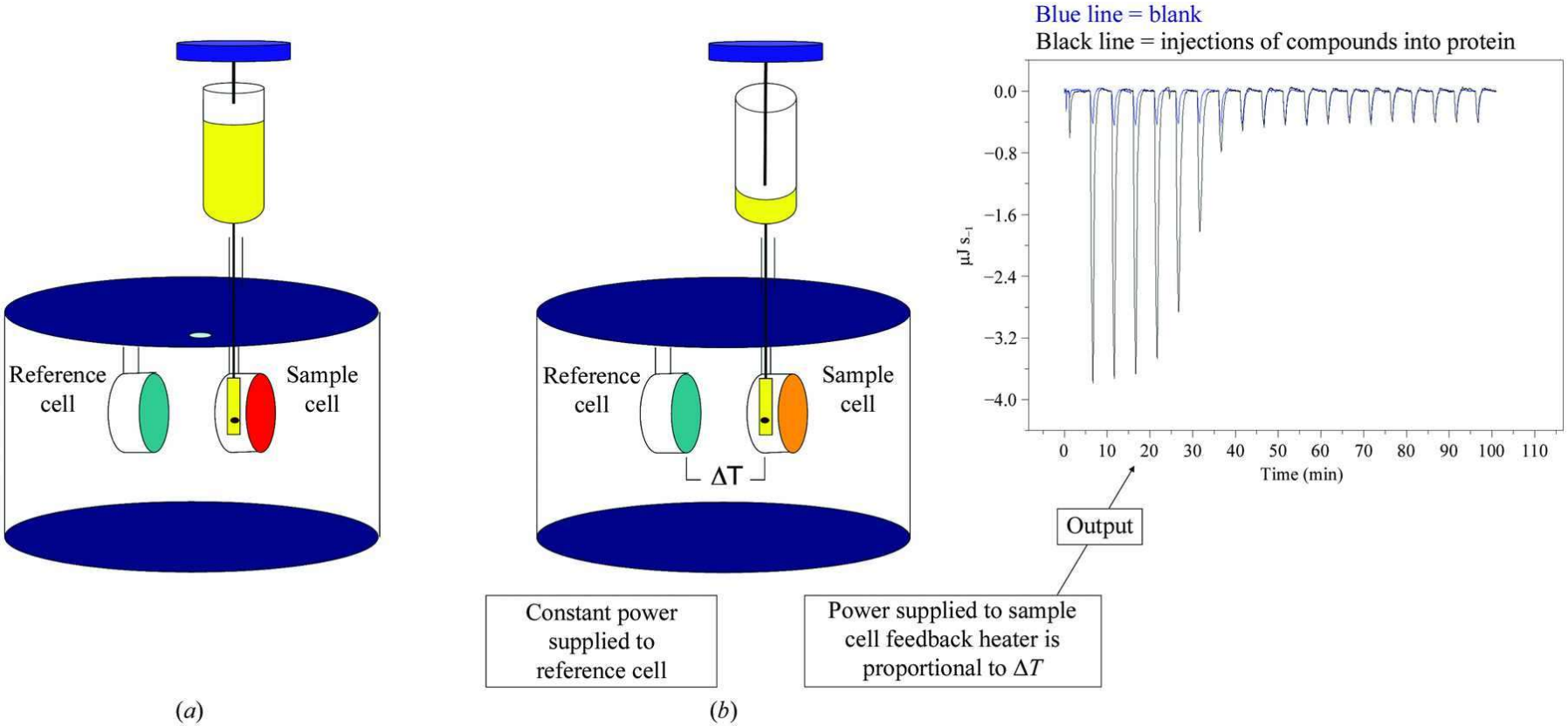
Difference of heat absorption between sample and reference is measured as a function of temperature

Detection of phase transitions, exothermic and endothermic processes



Used to prove:
- Conformational changes (i.e. pH dependence)
- Unfolding of protein in chaotropic agent
- Thermal stability of protein (i.e. aggregation)
- Binding to ligands
- Lipid-protein interactions

Ligand binding:
Isothermal Titration Calorimetry (ITC)



Energy supplied to the reference cell used as probe to follow binding of ligand/protein to the target protein.

References

- Post-expression strategies: Columbus L., “Post-expression strategies for structural investigations of membrane proteins.”, **Curr Opin Struct Biol. 2015**, *32*:131-8.
- Purification strategies: Kim Y. *et al.*, “High-throughput protein purification and quality assessment for crystallization.”, **Methods 2011**, *55(1)*:12-28; Konczal J. & Gray C.H., “Streamlining workflow and automation to accelerate laboratory scale protein production.”, **Protein Expr Purif. 2017**, *133*:160-169.
- **GE Healthcare Handbook**, “Purifying Challenging Proteins”, http://www.gelifesciences.com/file_source/GELS/Service%20and%20Support/Documents%20and%20Downloads/Handbooks/pdfs/Purifying%20Challenging%20Proteins.pdf
- **GE Healthcare Handbook**, “Recombinant Protein Purification Handbook”, https://www.gelifesciences.com/gehcls_images/GELS/Related%20Content/Files/1336168762999/litdoc18114275_20120505001440.pdf
- **Amersham Handbook**, “Protein Purification”, download from Gaudet Lab website, Harvard University, http://labs.mcb.harvard.edu/Gaudet/Resources_Files/GEHealthcare_chromatography/Don%27t%20move/18113229AB.pdf
- Protein stability: Deller M.C., Kong L., and Rupp B., “Protein stability: a crystallographer’s perspective”, **Acta Crystallogr F Struct Biol Commun. 2016**, *72(Pt 2)*: 72–95.