- **Class Today**

1. finish up proteomics

2. Lecture 7 and Lecture 8

- **Announcement**

1. The CW1 will be uploaded onto LMO tomorrow (April 9th) and due on April 23rd (Tuesday) at 5 pm.

Poll Question

1. What is a 'proteotypic' peptide?

A) A post-translationally modified peptide

B) A stable isotope-containing peptide

C) A peptide which is unique to a specific protein

D) A peptide which is typical of all other peptides

Poll Question

2. Selected reaction monitoring (SRM) is useful for which of the following?

A) Comparing the levels of hundreds or thousands of proteins

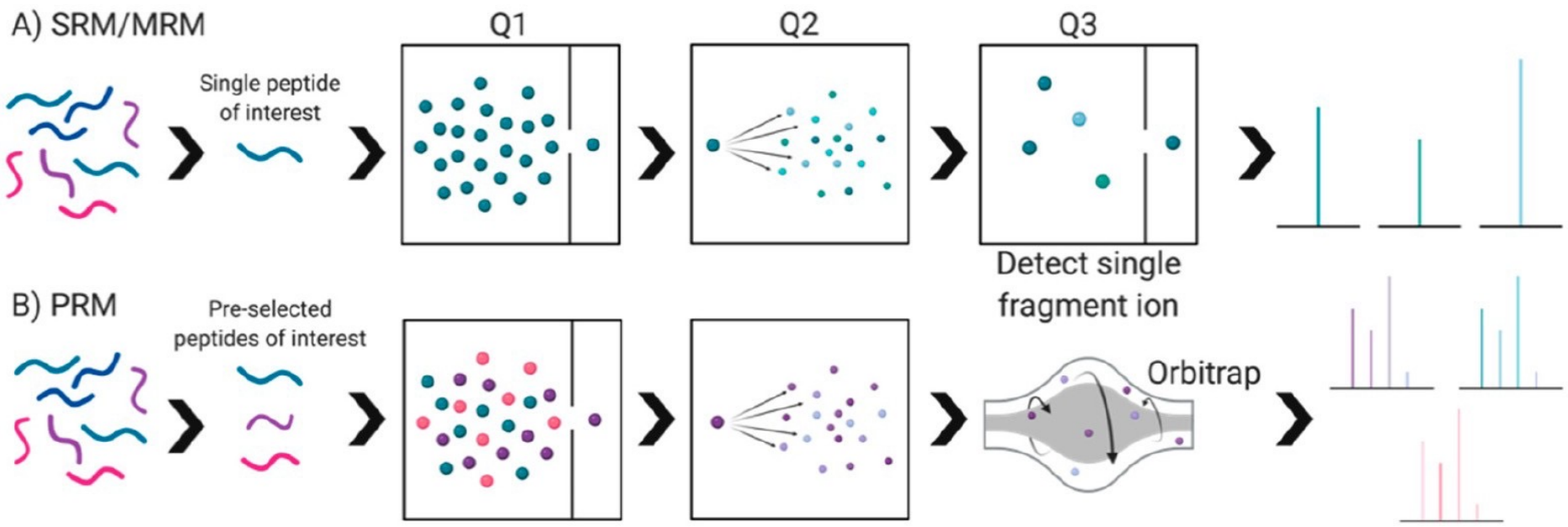
B) Comparing the levels of a specific protein

C) Identifying a protein in a gel band/spot

D) Identifying the post-translational modification position of an unknown protein

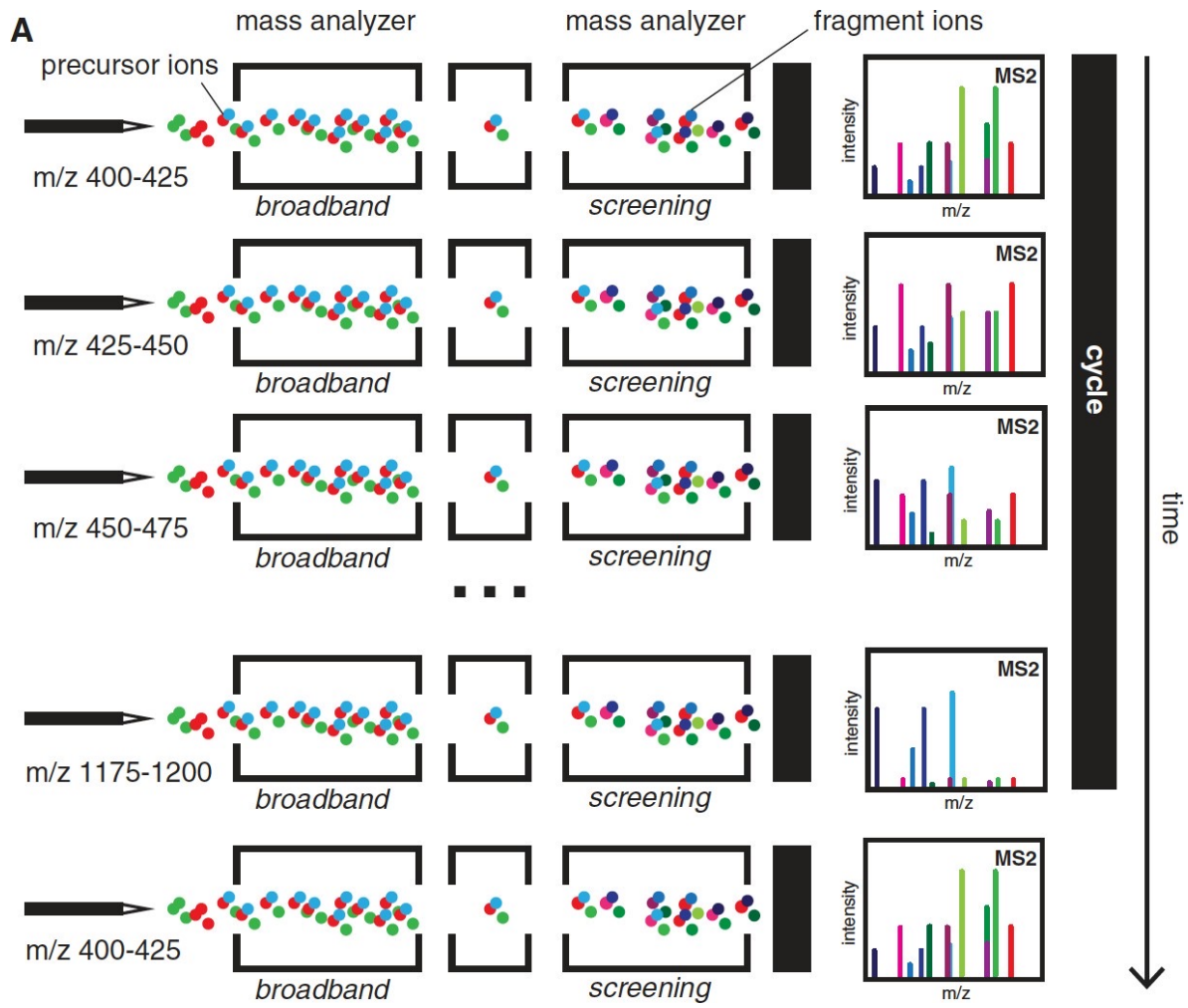
Parallel Reaction Monitoring (PRM): Review

- Quadrupole Orbitrap or Quadrupole TOF
- Produce **full** MS2 spectra for each precursor



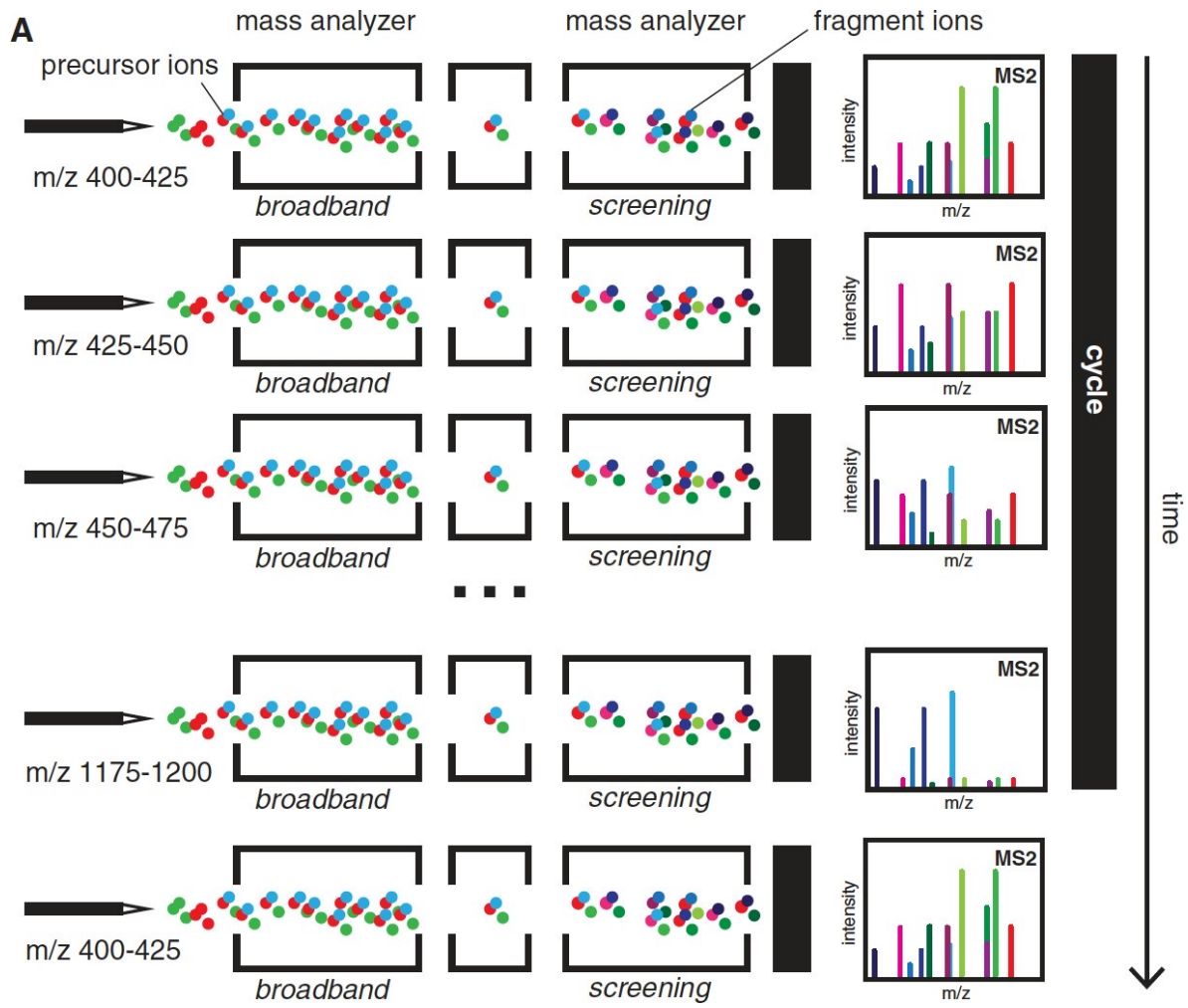
Data-Independent Acquisition (DIA)/SWATH

- **MS2 data-independent acquisition methods** are usually implemented in quadrupole-orbital ion traps, and quadrupole TOF hybrid mass analyzers.
- These methods rely on the use of one or several broadband isolation windows that isolate all peptide ions in a sample within each cycle.
  - 32 consecutive 25 Da precursor isolation windows



Data-Independent Acquisition (DIA)/SWATH

- All peptide ions within a window are simultaneously fragmented, and the resulting fragment ion maps can be subjected to targeted data analysis to identify the coeluting fragment ion groups from the peptides of interest.
- Generate data for all peptides and therefore, the targets of interest can be selected post-acquisition.
- Can only be considered targeted as long as there is a hypothesis to be tested.
  - different protein isoforms, splicing variants, and specific modified peptides



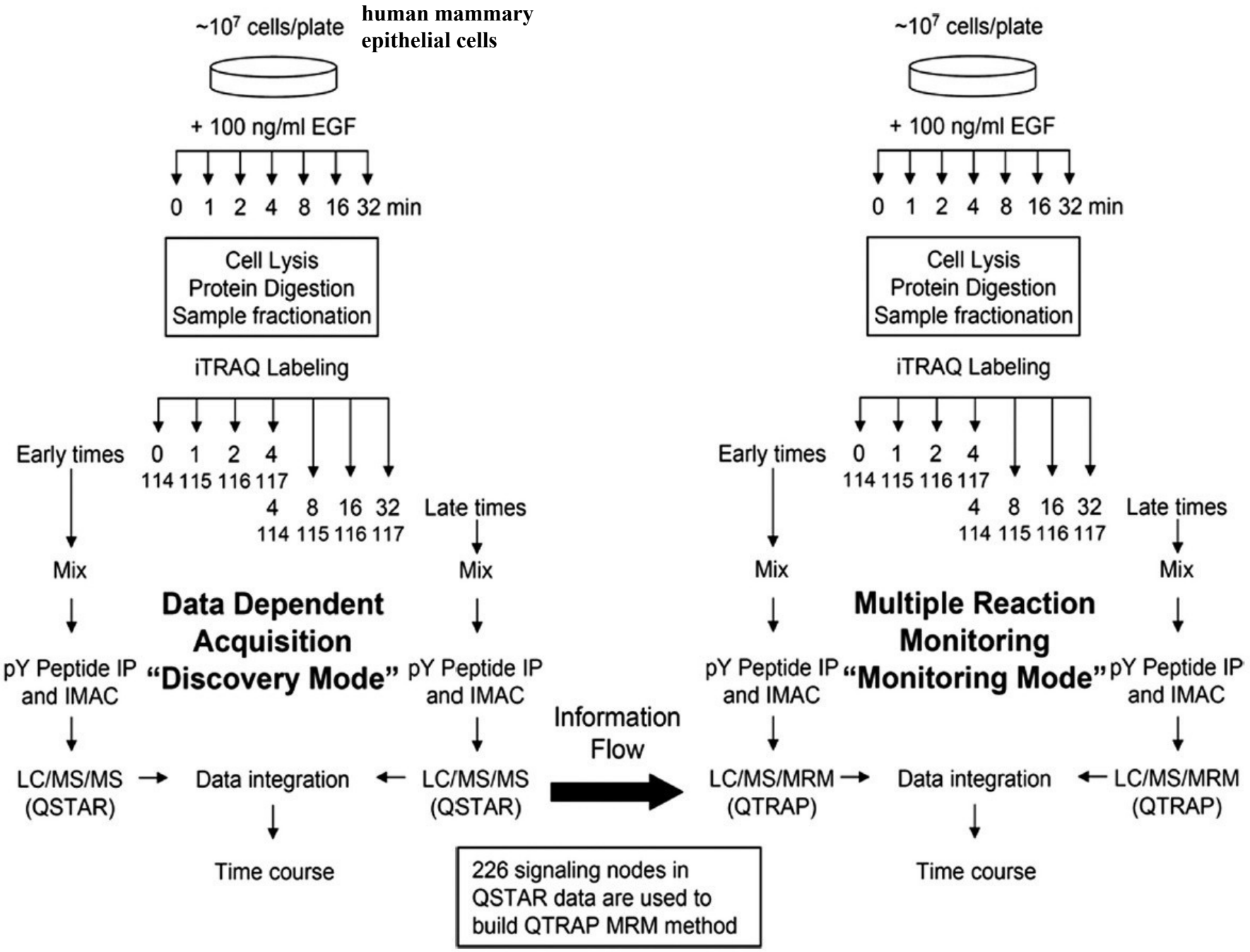
Case study: identify and quantify hundreds of phosphorylation sites by using DDA/MRM method

-temporal dynamics of protein tyrosine phosphorylation within the epidermal growth factor receptor (EGFR) signaling network

Reference paper: Wolf-Yadlin A, Hautaniemi S, Lauffenburger DA, White FM. Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. Proc Natl Acad Sci U S A. 2007 Apr 3;104(14):5860-5. doi: 10.1073/pnas.0608638104. Epub 2007 Mar 26. PMID: 17389395; PMCID: PMC1851582.

Combination of discovery DDA and targeted MRM method

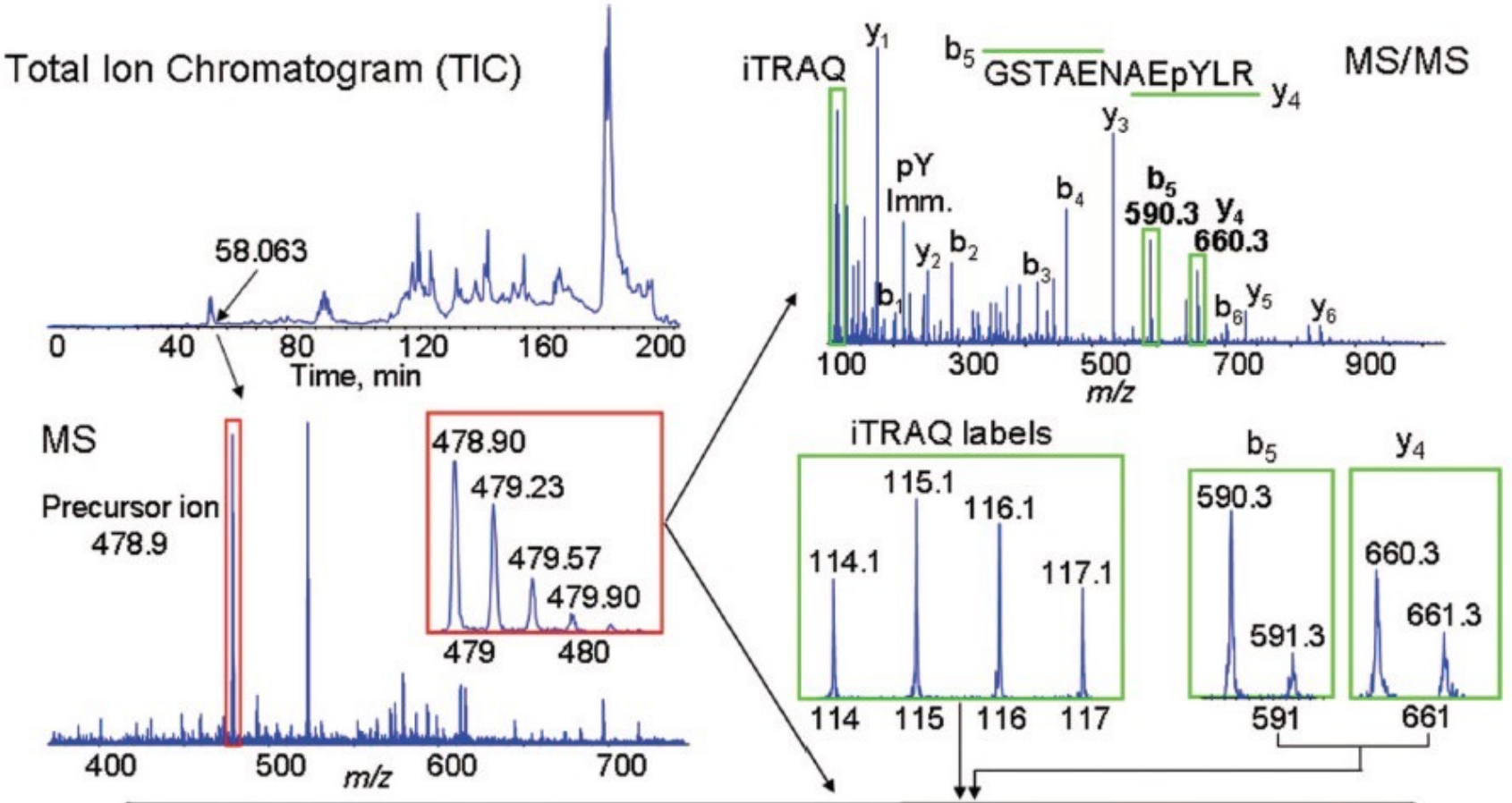
Quantify 226 phosphorylated peptides covering 208 phosphorylation sites in 143 proteins across seven time points of EGF stimulation.



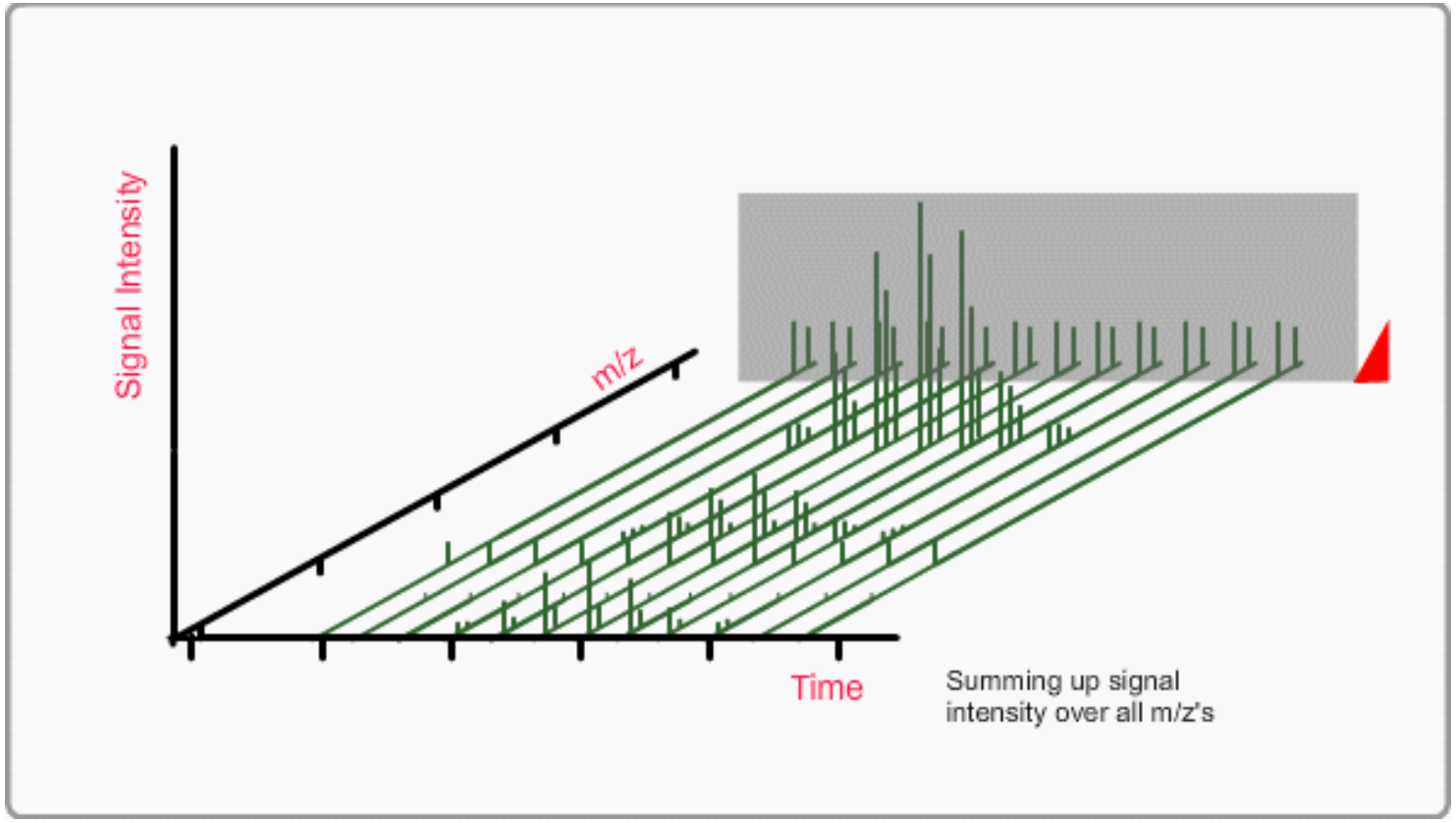
Construction of the MRM method from DDA data

<u>Parameters obtained from DDA:</u>

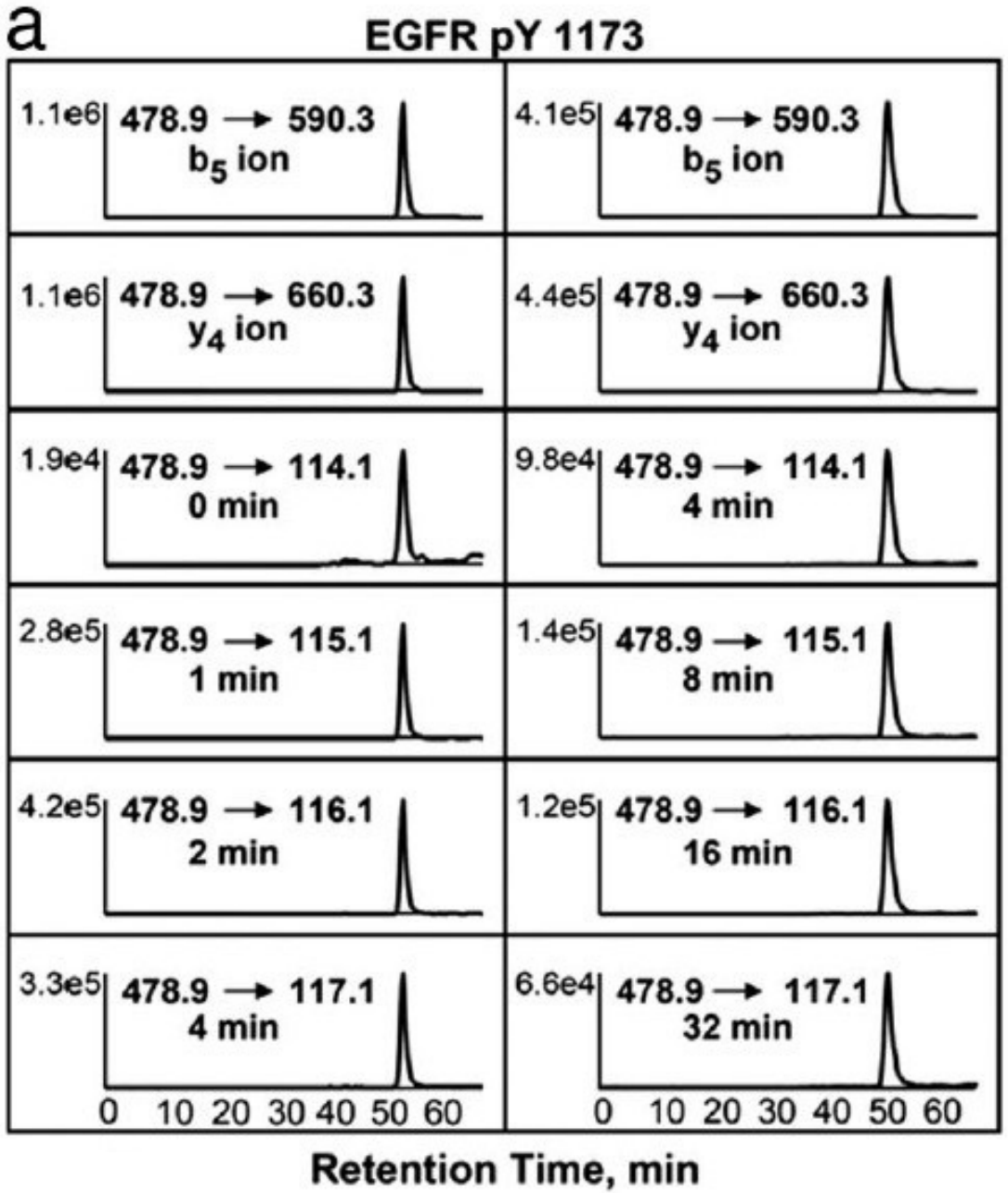
1. Retention time
2. m/z and z of the precursors
3. m/z and z of characteristic b- and y-type fragment ions

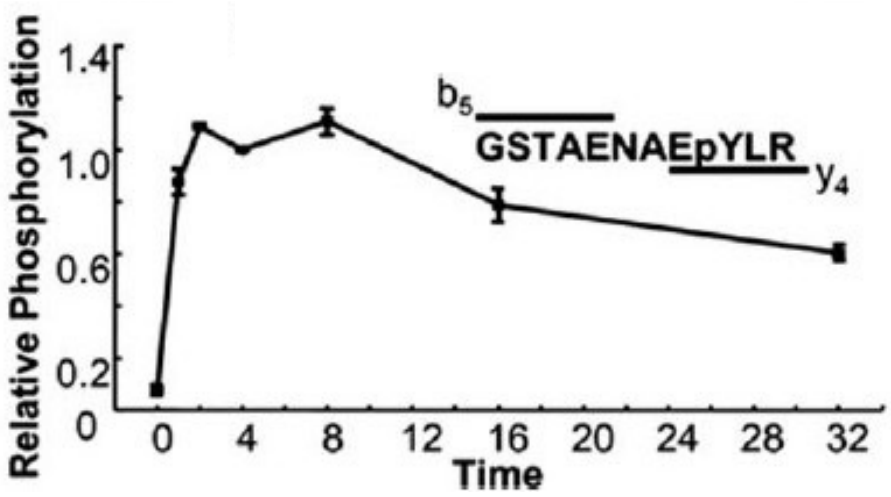


Parameters to build MRM Acquisition Method								
Protein	Peptide	pY	Elution Order	Prec. Ion	Frag. Ion	CE (V)	Time (ms)	Ion
EGFR	GSTAENAEpYLR	1173	15	478.9	590.3	27	30	b5
EGFR	GSTAENAEpYLR	1173	15	478.9	660.3	27	30	y4
EGFR	GSTAENAEpYLR	1173	15	478.9	114.1	41	30	iTRAQ
EGFR	GSTAENAEpYLR	1173	15	478.9	115.1	41	30	iTRAQ
EGFR	GSTAENAEpYLR	1173	15	478.9	116.1	41	30	iTRAQ
EGFR	GSTAENAEpYLR	1173	15	478.9	117.1	41	30	iTRAQ



Total ion chromatogram (TIC): **a chromatogram created by summing up intensities of all mass spectral peaks belonging to the same scan.**


a
EGFR pY 1173
1.1e6
478.9 → 590.3
b5 ion
4.1e5
478.9 → 590.3
b5 ion
1.1e6
478.9 → 660.3
y4 ion
4.4e5
478.9 → 660.3
y4 ion
1.9e4
478.9 → 114.1
0 min
9.8e4
478.9 → 114.1
4 min
2.8e5
478.9 → 115.1
1 min
1.4e5
478.9 → 115.1
8 min
4.2e5
478.9 → 116.1
2 min
1.2e5
478.9 → 116.1
16 min
3.3e5
478.9 → 117.1
4 min
6.6e4
478.9 → 117.1
32 min
0 10 20 30 40 50 60
0 10 20 30 40 50 60
Retention Time, min

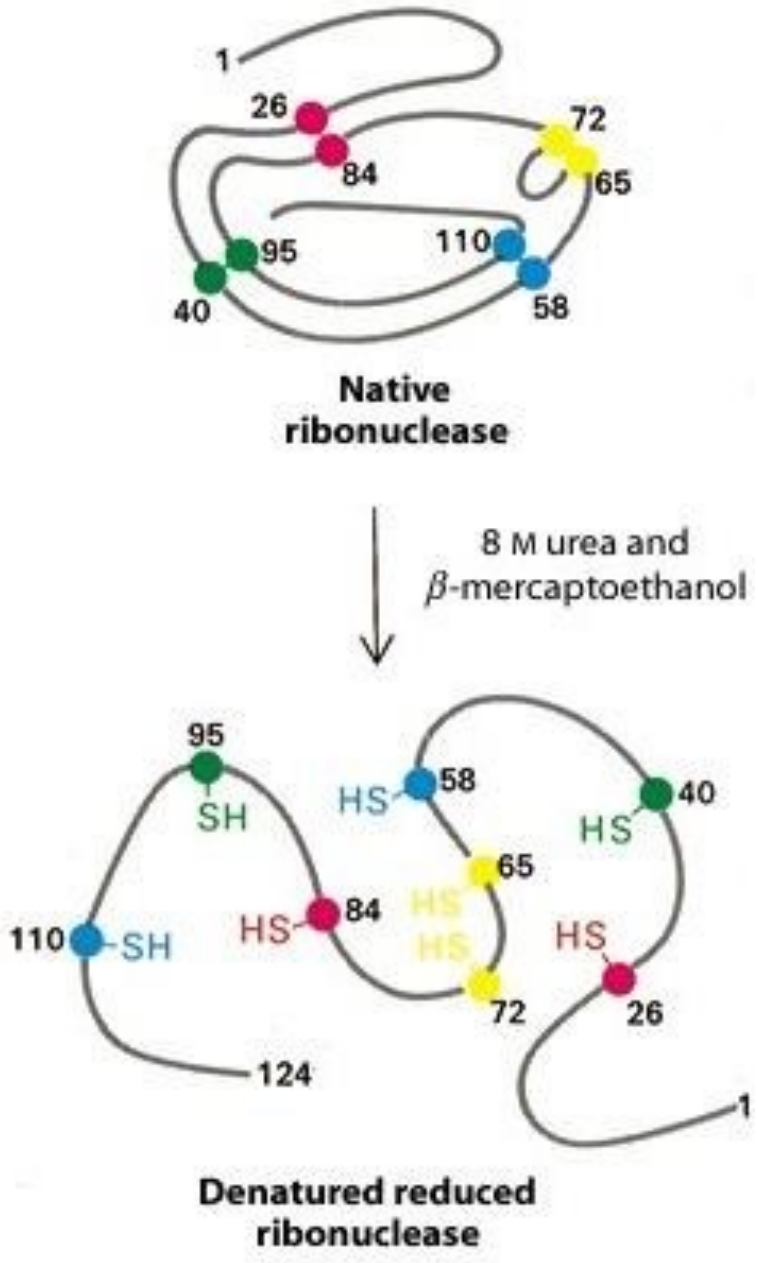

Relative Phosphorylation
1.4
1.0
0.6
0.2
0
b5
GSTAENAEpYLR
y4
0 4 8 12 16 20 24 28 32
Time

LECTURE 7: PROTEIN FOLDING AND UNFOLDING

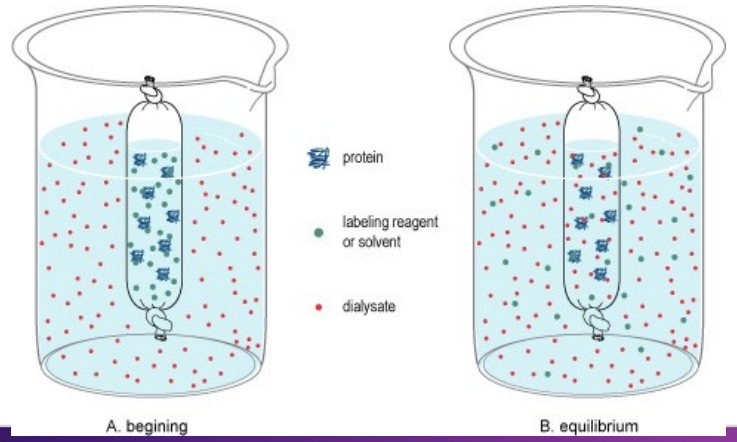
Bio312

Instructor: Dr. Lanlan Han
E-mail: Lanlan.Han@xjtlu.edu.cn

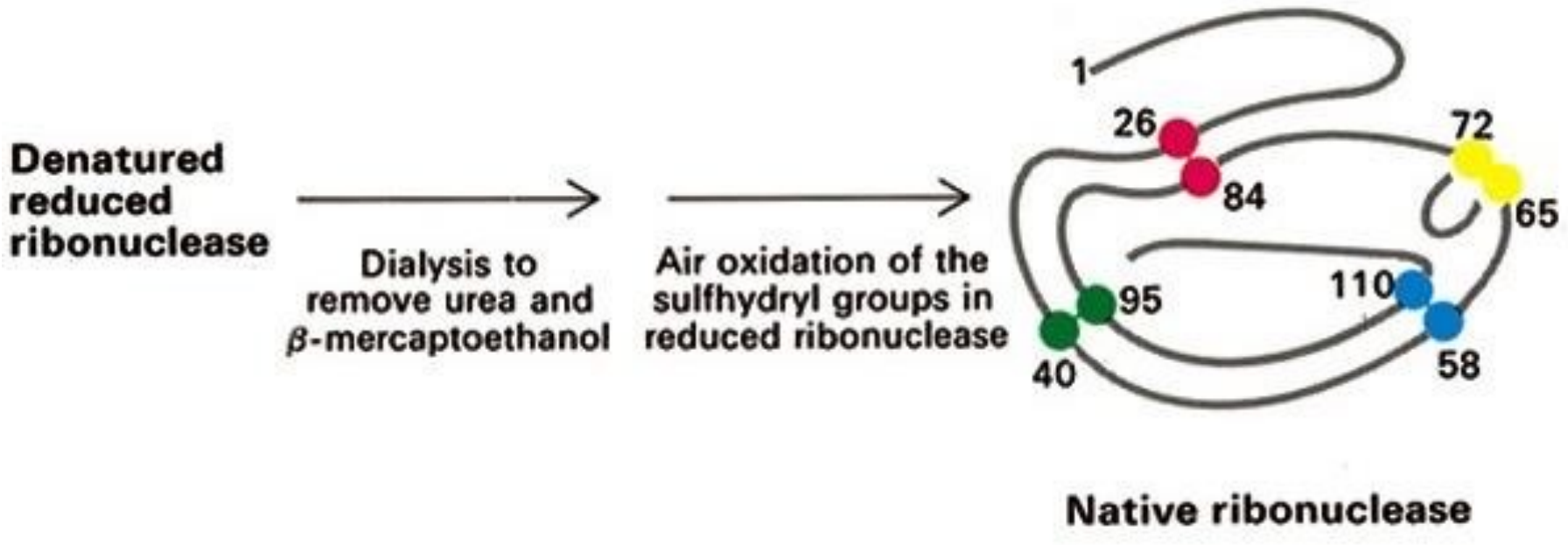
Christian Anfinsen Experiment$_1$ (Nobel Prize, 1972)



- **Denaturation** of ribonuclease A (4 disulfide bonds) with 8 M urea containing β-mercaptoethanol to random coil, no activity
  - Urea disrupts noncovalent interactions (ionic, H bonding)
  - β-mercaptoethanol reduces the disulfide bond

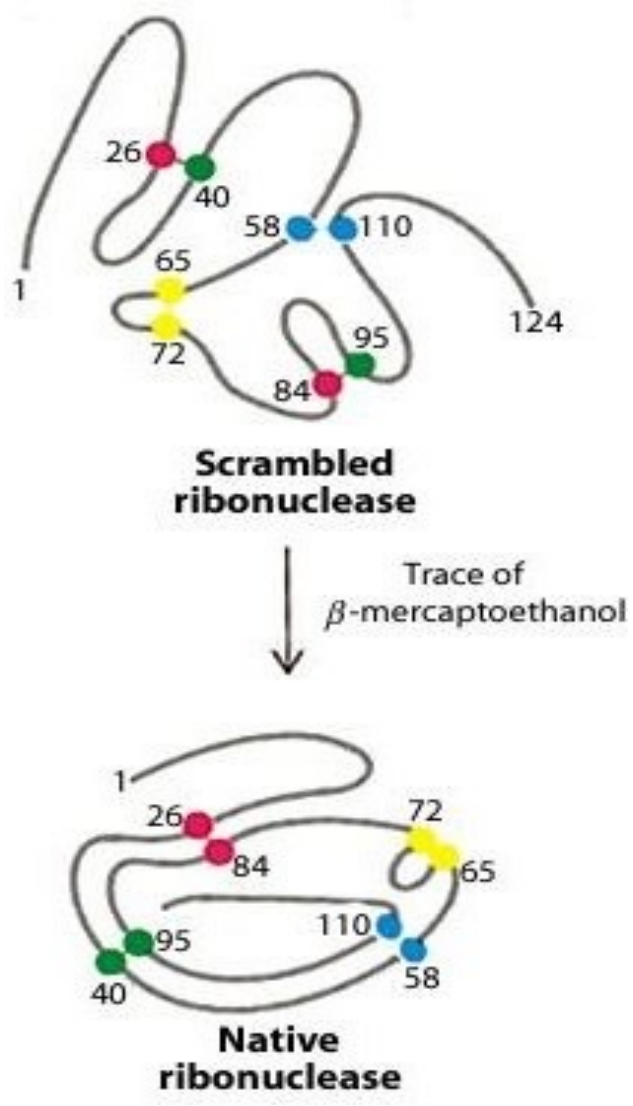


Anfinsen Experiment$_2$



- After renaturation, the refolded protein has native activity
- ***Conclusion***: All the information necessary for folding the peptide chain into its native structure is contained in the primary amino acid sequence of the peptide.

Anfinsen Experiment$_3$



- Remove β-mercaptoethanol only, oxidation of the sulfhydryl group, then remove urea → scrambled protein, no activity
- Further addition of trace amounts of β-mercaptoethanol converts the scrambled form into native form.
- ***Conclusion***: The native form of a protein is the thermodynamically most stable structure.

The Levinthal Paradox

- How long does it take for a protein folds up into the native state?

  - Consider just for the peptide backbone, for example, there are 3 conformations per amino acid in the unfolded state. For a 100 a.a. protein, we have $\mathbf{3^{100}}$ conformations.
  - If the chain can sample 10^{12} conformations/sec, it takes 5×10^{35} sec (2×10^{28} year)
  - Actually, protein folding is fast (millisecond to sec)
  - **Conclusion**: Protein folding is not random, must have specific pathways.

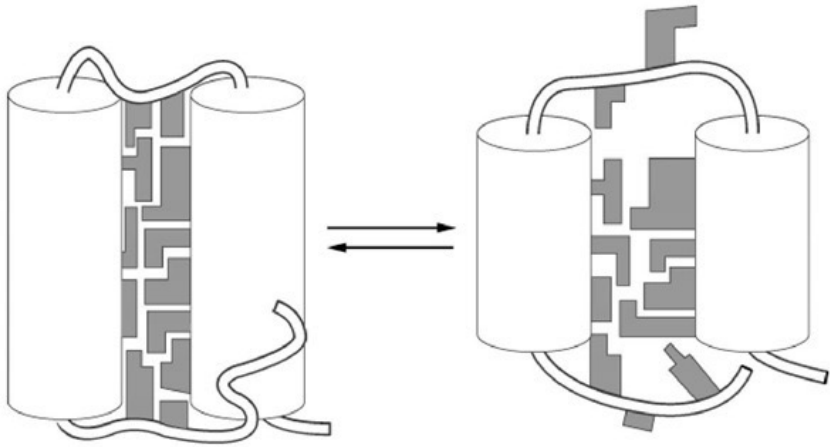
Protein Folding

- The primary structure determines its tertiary structure.
  - **Anfinsen Experiment**

- Forces behind protein folding (native state)
  - Hydrophobic effect
    - The clustering of hydrophobic AAs from different parts of the polypeptide chain causes the polypeptide chain become compact.
  - Noncovalent interactions maintaining folding
    - H bonding
    - Van der Waals forces
    - Salt bridges (electrostatics)

Protein Folding Intermediate

- Protein folding can occur quite rapidly, but one or more partially folded intermediate states often exist.
  - **Molten Globule state**: a loosely collapsed state with fluctuating tertiary interactions and very weak secondary structure

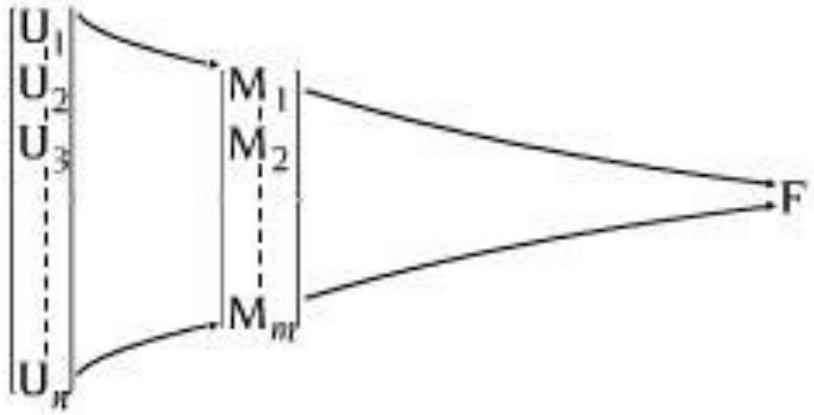


Native Molten Globule

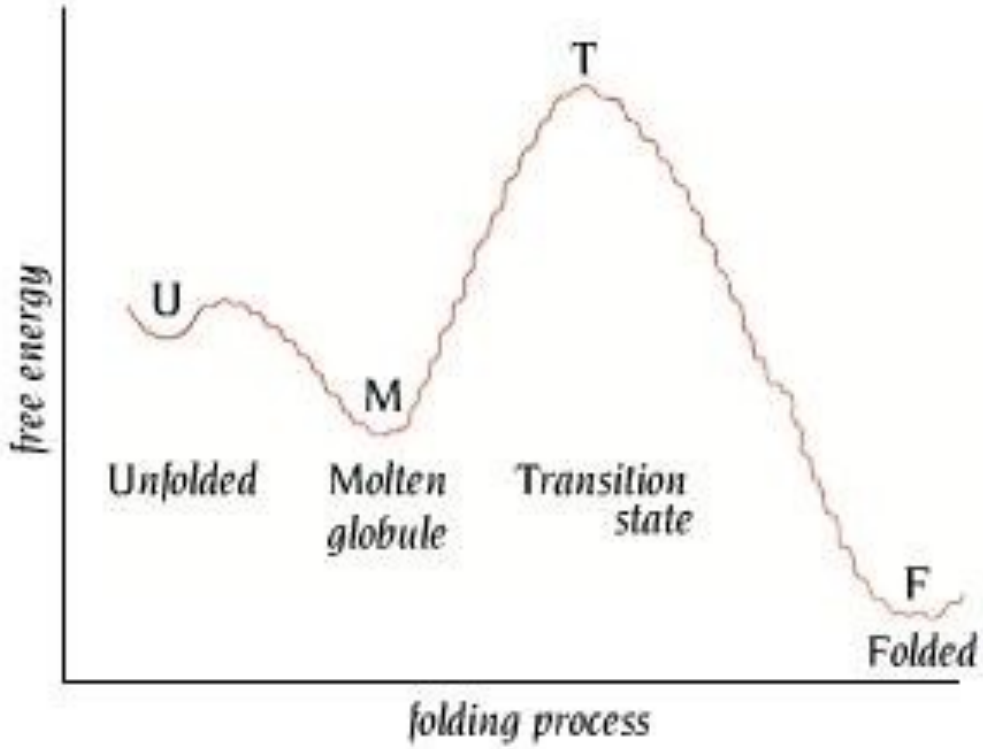
- It is an intermediate of the folding transition: U→MG→F
- It is a compact globule, yet expanded over a native radius
- Native-like secondary structure, can be measured by CD and NMR proton exchange rate
  - no detectable near UV CD signal
  - quenched signal with broadened NMR chemical peaks

http://what-when-how.com/wp-content/uploads/2011/05/tmp482_thumb.jpg

Unfolded State

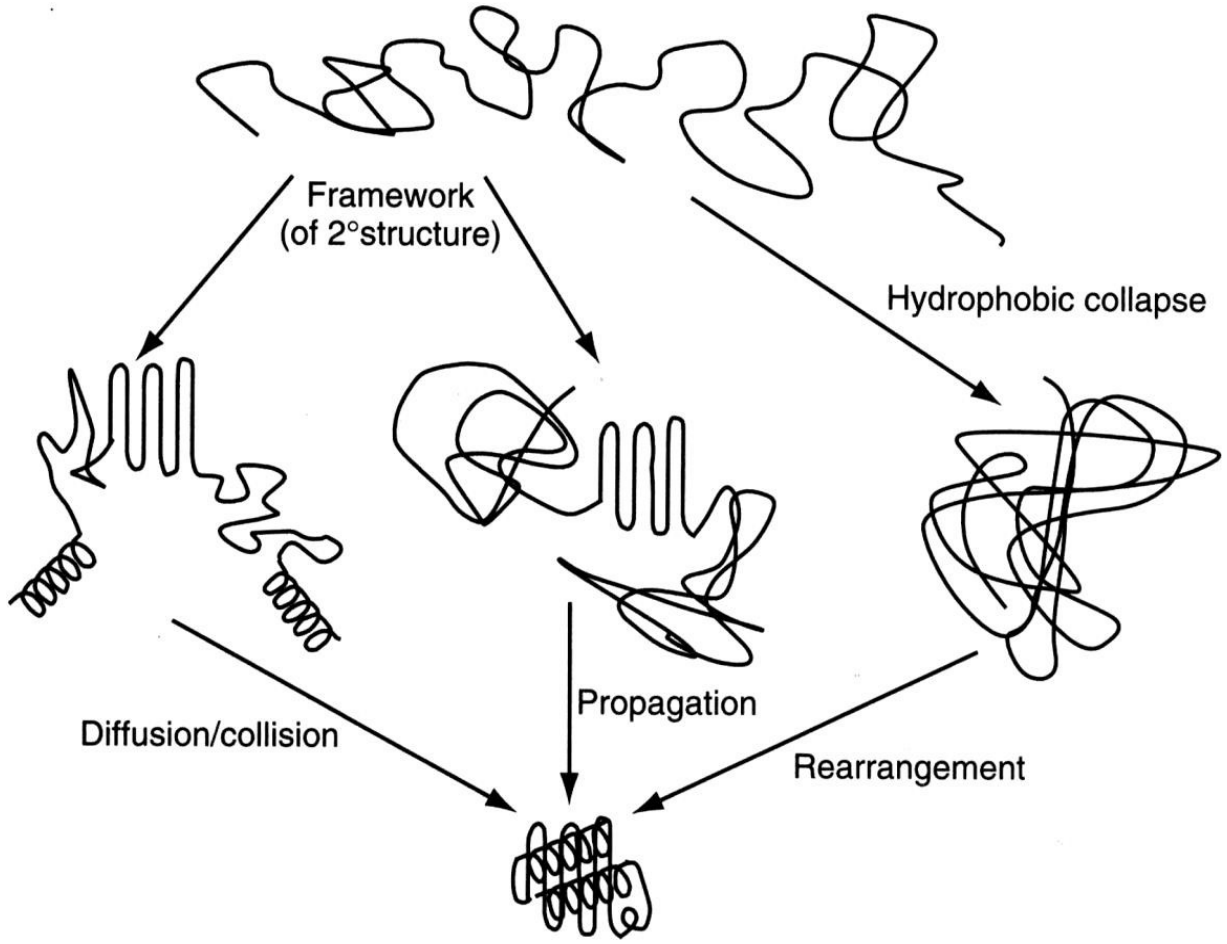


- The unfolded state is an ensemble of a large number of molecules with different conformations.



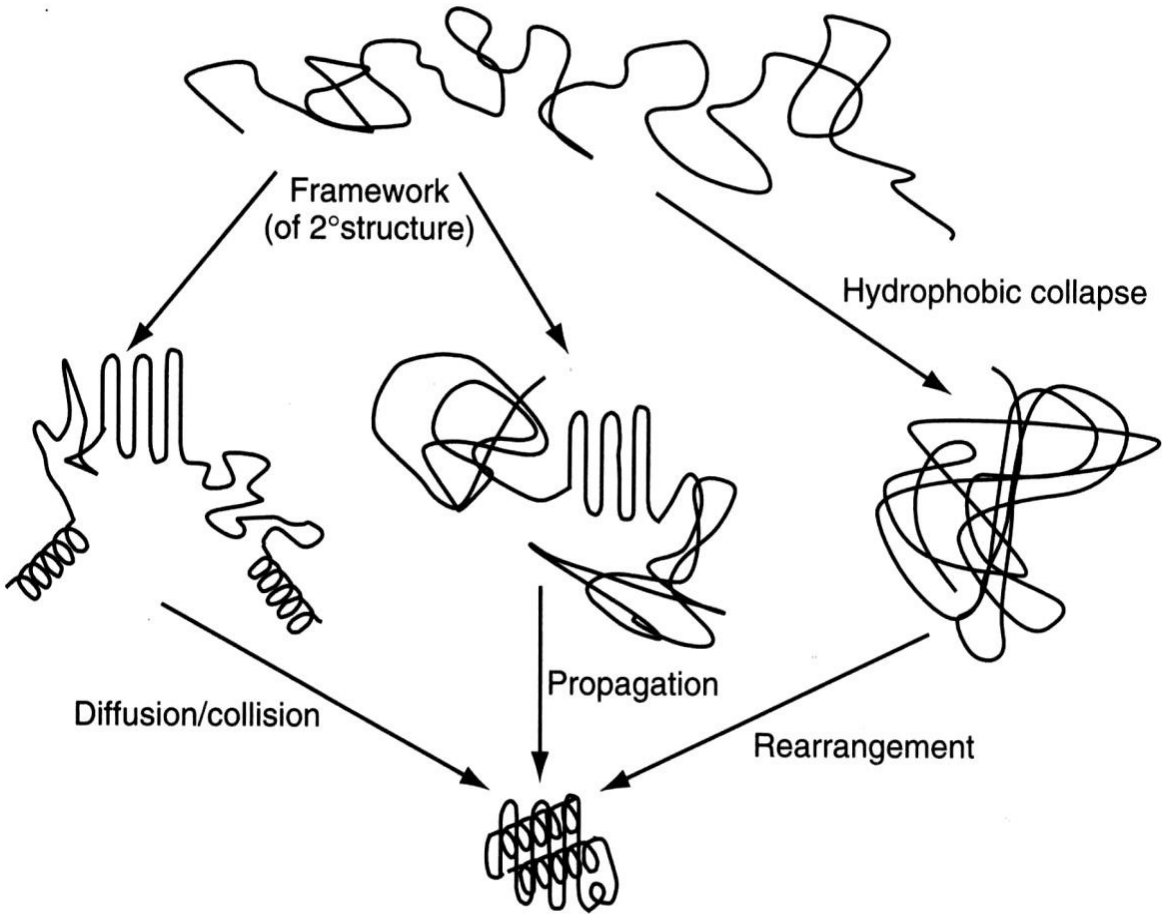
Three Classic Models of Protein Folding: Framework model

- The Framework model proposed that local elements of native local secondary structure could form independently of tertiary structure (Kim and Baldwin). These elements would diffuse until they collided, successfully adhering and coalescing to give the tertiary structure (diffusion-collision model)(Karplus & Weaver).



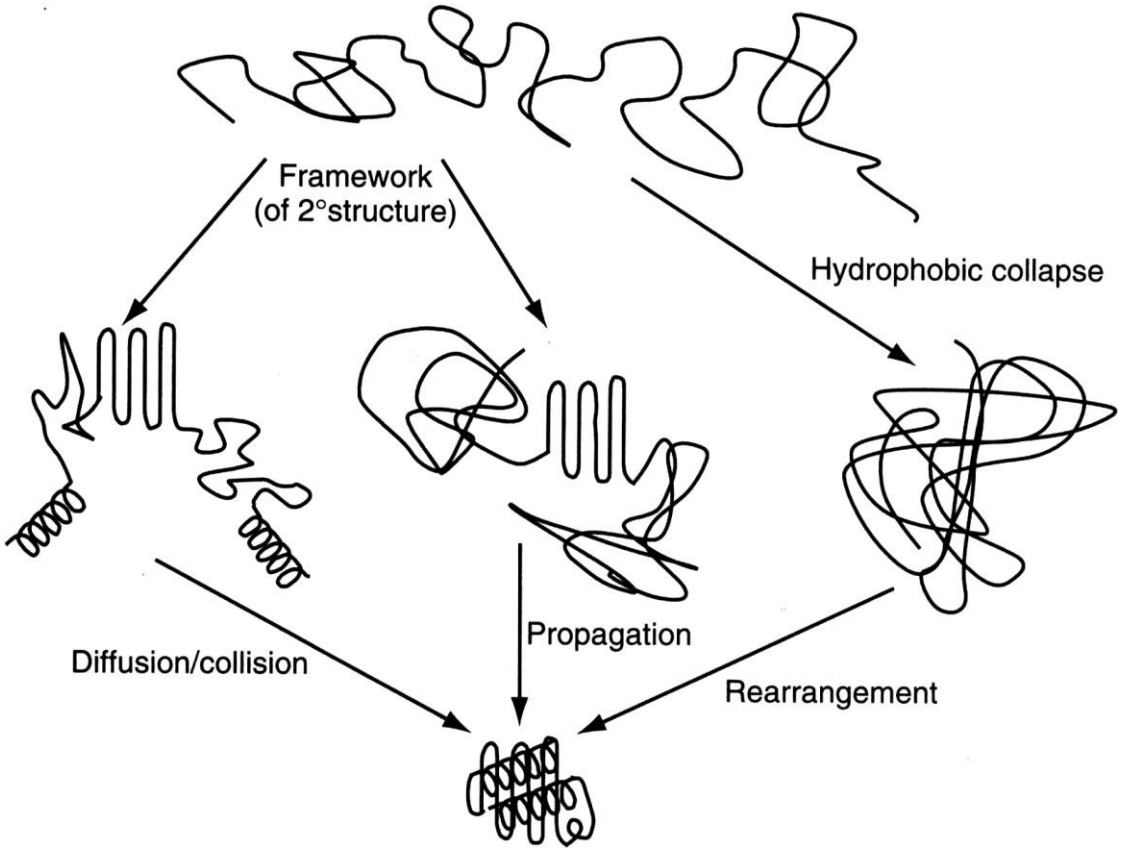
Three Classic Models of Protein Folding: The classic Nucleation Model

- The classic nucleation model postulated that some neighboring residues in the sequence would form native secondary structure that would act as a nucleus from which the native structure would propagate in a stepwise manner. Thus, the tertiary structure would form as a necessary consequence of the secondary structure.



Three Classic Models of Protein Folding: The hydrophobic-collapse Model

- The hydrophobic-collapse model hypothesized that a protein would collapse rapidly around its hydrophobic sidechains and then rearrange from restricted conformational space occupied by the intermediate. Here the secondary structure would be directed by native-like tertiary structure (Ptitsyn & Kuwajima).



The Folded Protein Is A Thermodynamic Compromise

- Stability is defined as a net loss of free energy.

$$\Delta G = \Delta H - T\Delta S$$

- Enthalpy (H): heat released when interactions form
- Entropy (S): a measurement of randomness or disorder.
  - Protein folding tends to increase the total entropy of a system and its surroundings.
    - Protein entropy ↓, but solvent (water) entropy ↑, ⇒ Net ↑

➢The difference in free energy between the folded and unfolded states is actually small.

The Folding Funnel

- The funnel is a collections of geometrically similar structure to native proteins.

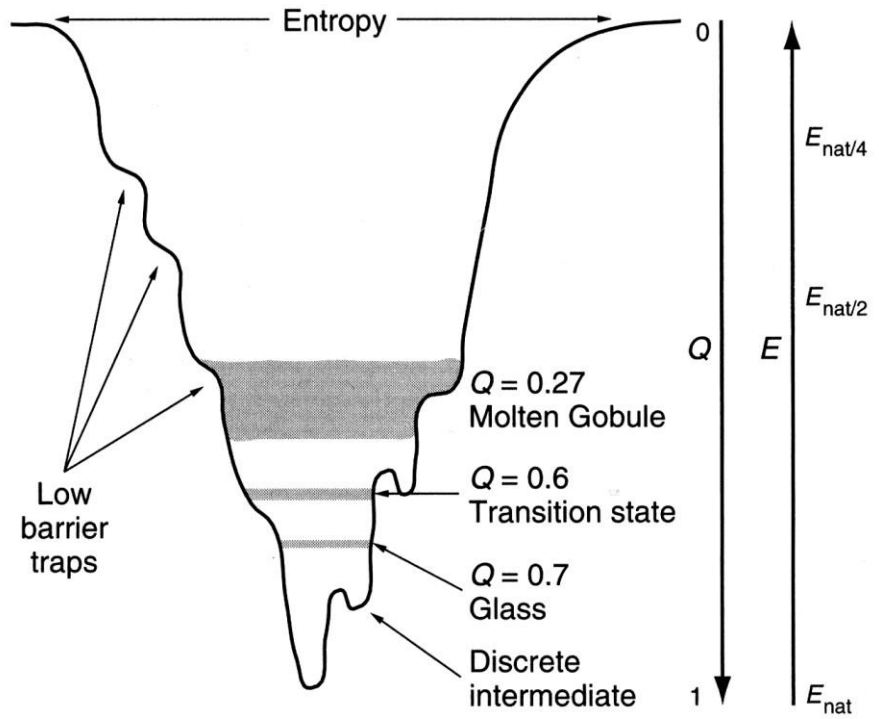
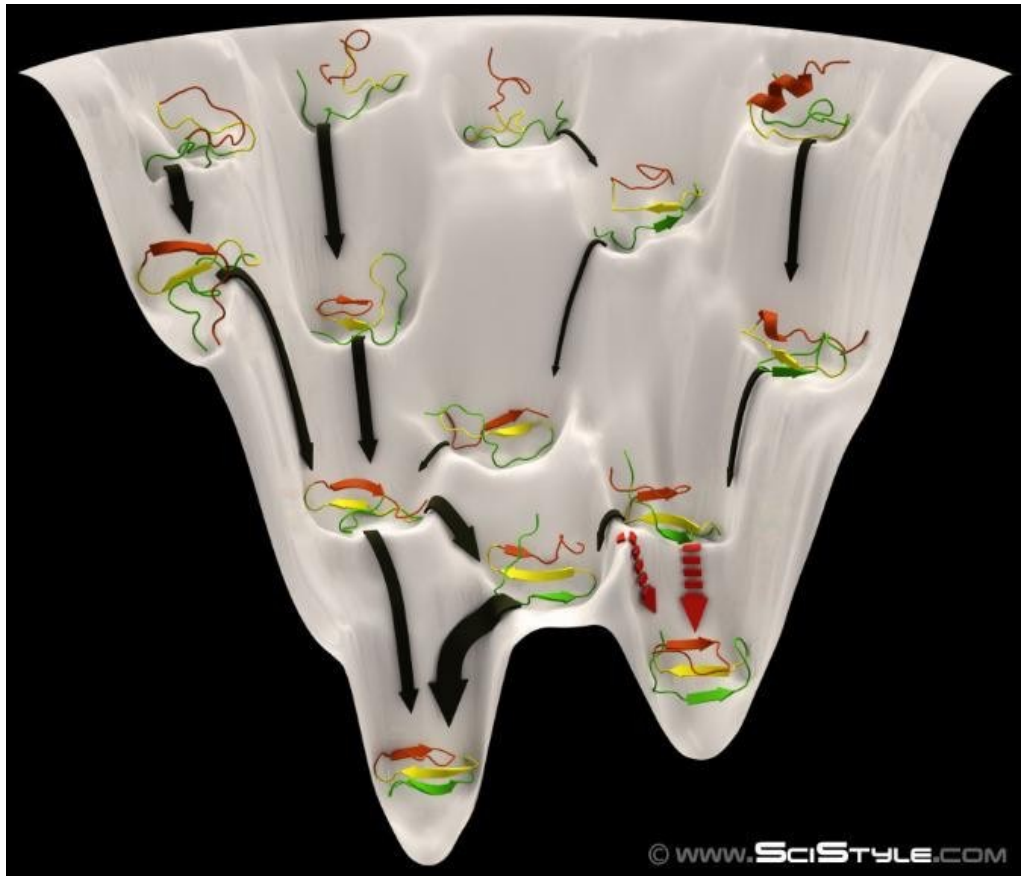
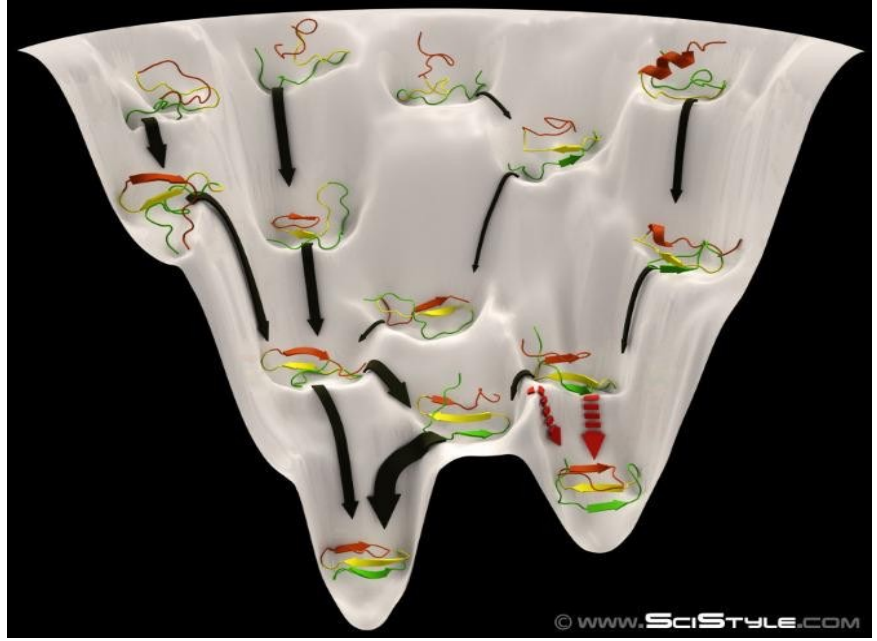


Figure 19.16 Cross section through a folding funnel. *E* corresponds to free energy.
[Courtesy of P. G. Wolynes]



The Folding Funnel



- A new view of protein folding suggested that there is no single route, but a large ensemble of structures follow a many dimensional funnel to its native structure.

- Progress from the top to the bottom of the funnel is accompanied by an increase in the native-like structure as folding proceeds.
- Energy landscape theory of protein folding
  - Assumes the native state is a deep free energy minimum, corresponding to a single well-defined tertiary structure usually in a cell.

Protein Folding in vivo

- Co-translational protein folding
  - Translation of mRNA produces a linear polymer of amino acids that usually folds spontaneously into a more compact, stable structure.
  - *Assumption*: as soon as the nascent chain is extruded, it will start to fold co-translationally (i.e., acquire secondary structures, super-secondary structures, domains) until the complete polypeptide is produced and extruded

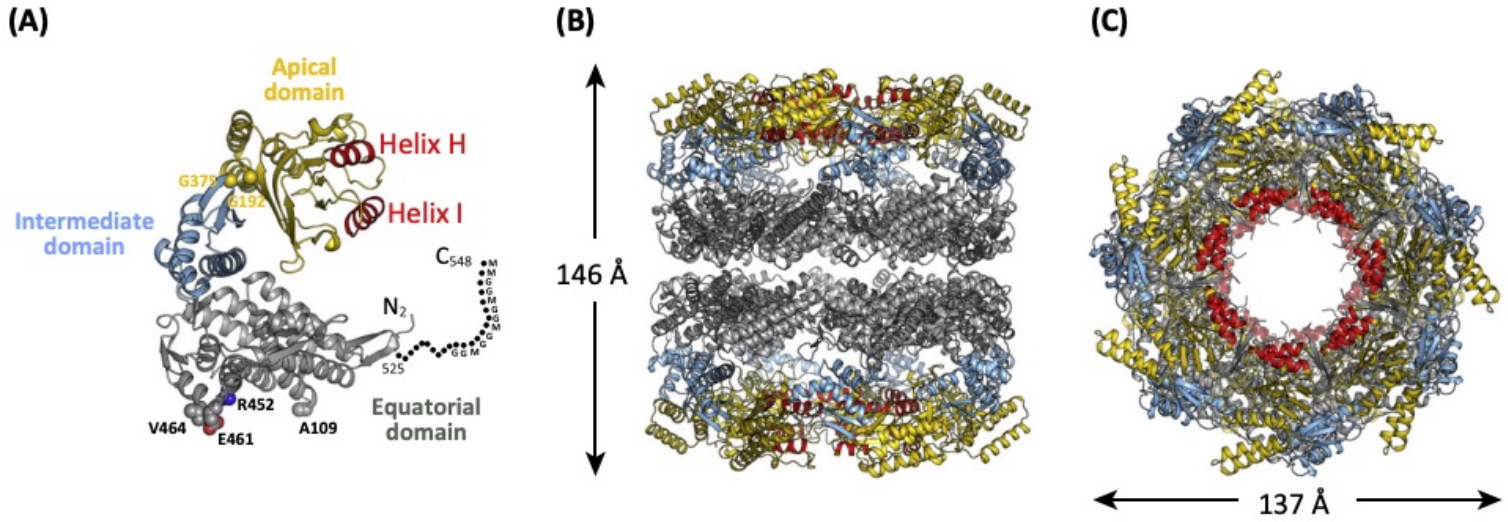
- Some proteins need help to fold
  - Specific environment (S-S $\rightleftharpoons$ -SH HS-)
  - Chaperones
  - Cofactors

Chaperones

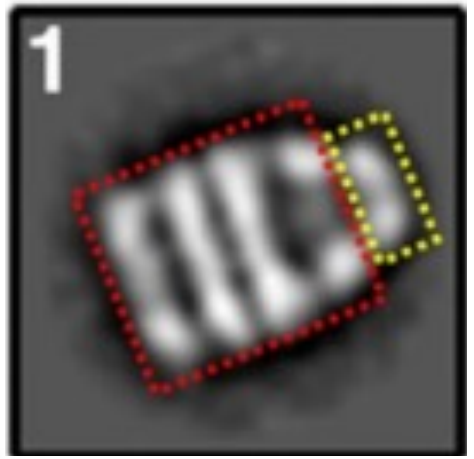
- Assist protein folding primarily by preventing inappropriate interactions between non-native polypeptides that would otherwise lead to aggregation
  - GroEL-GroES complex forms large 800–1000 kDa double-ring structure with an internal chamber in each ring.
  - GroEL-GroES complex recognizes hydrophobic amino acids exposed by non-native proteins and mediates folding through ATP-regulated cycles of binding and release

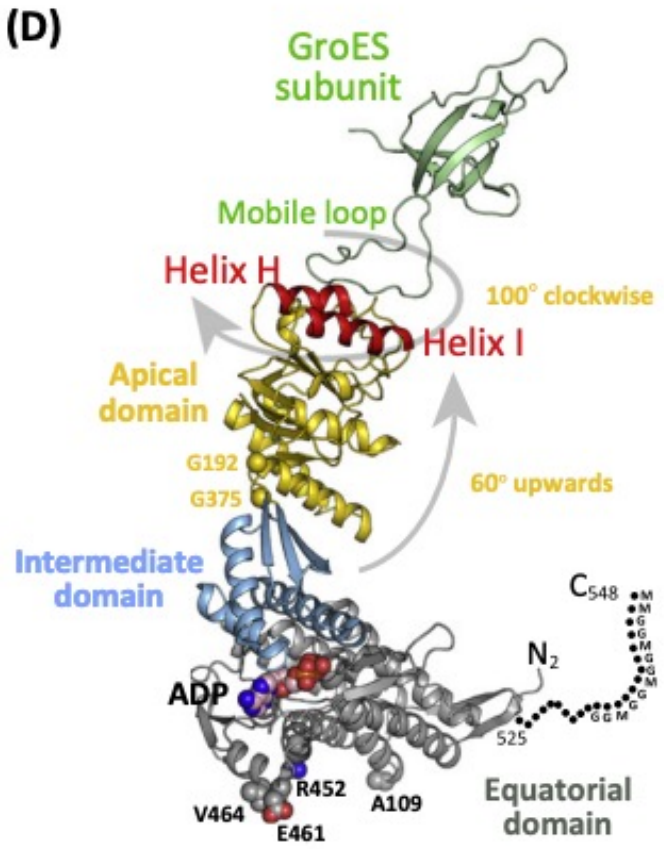
GroEL (Hsp 60, Cpn60)

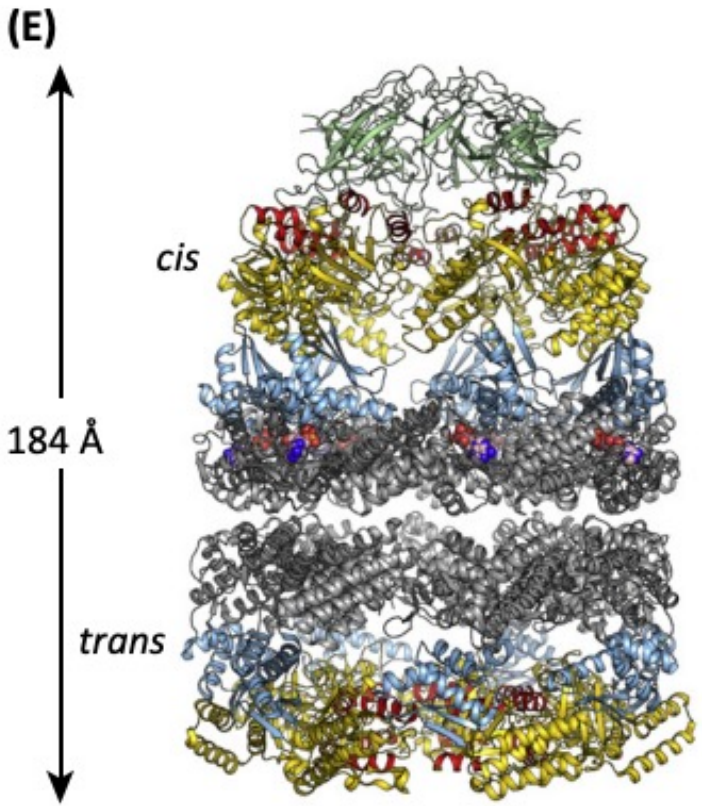
- GroEL of *Escherichia coli* is a cylindrical complex of two heptameric rings of 57 kDa subunits.
- Each subunit is composed of:
  - an equatorial ATP-binding domain,
  - an intermediate hinge domain,
  - an apical domain that binds non-native Substrate protein (SP) and GroES.

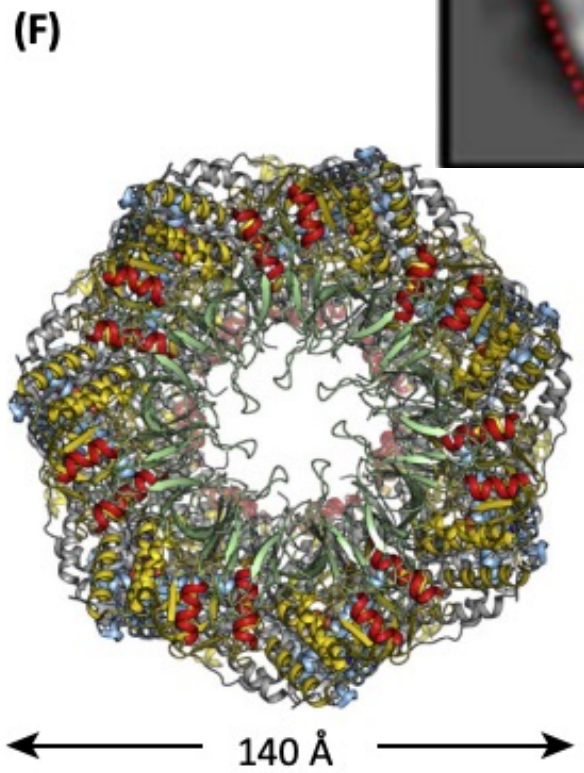


GroEL-GroES Complex



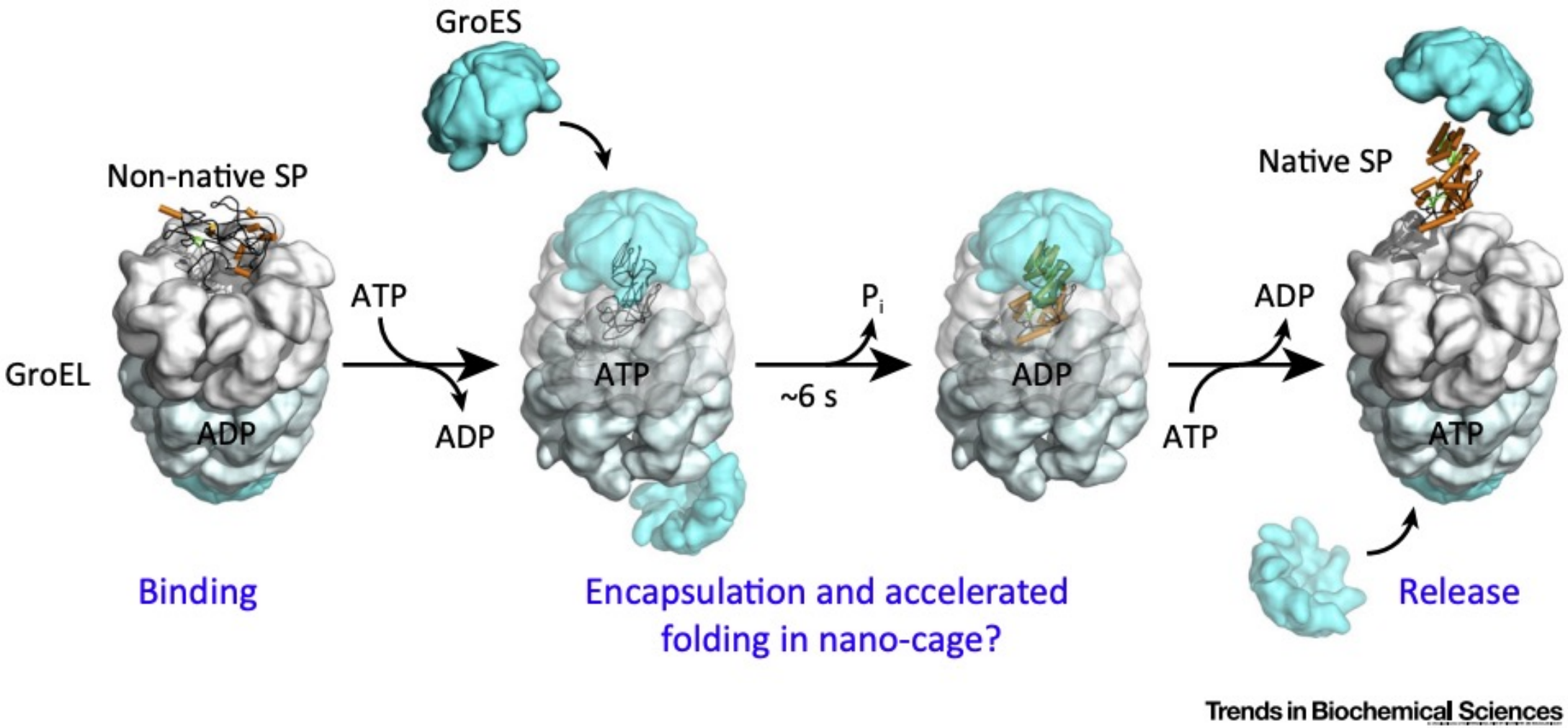






Trends in Biochemical Sciences

Protein Folding in the GroEL/ES Chaperonin Cage



Protein Denaturation

- Changing their physical or chemical environment
  - Heating
  - Adding a chemical denaturant (e.g., urea, guanidinium chloride)
  - Changing pH
  - Applying high pressure

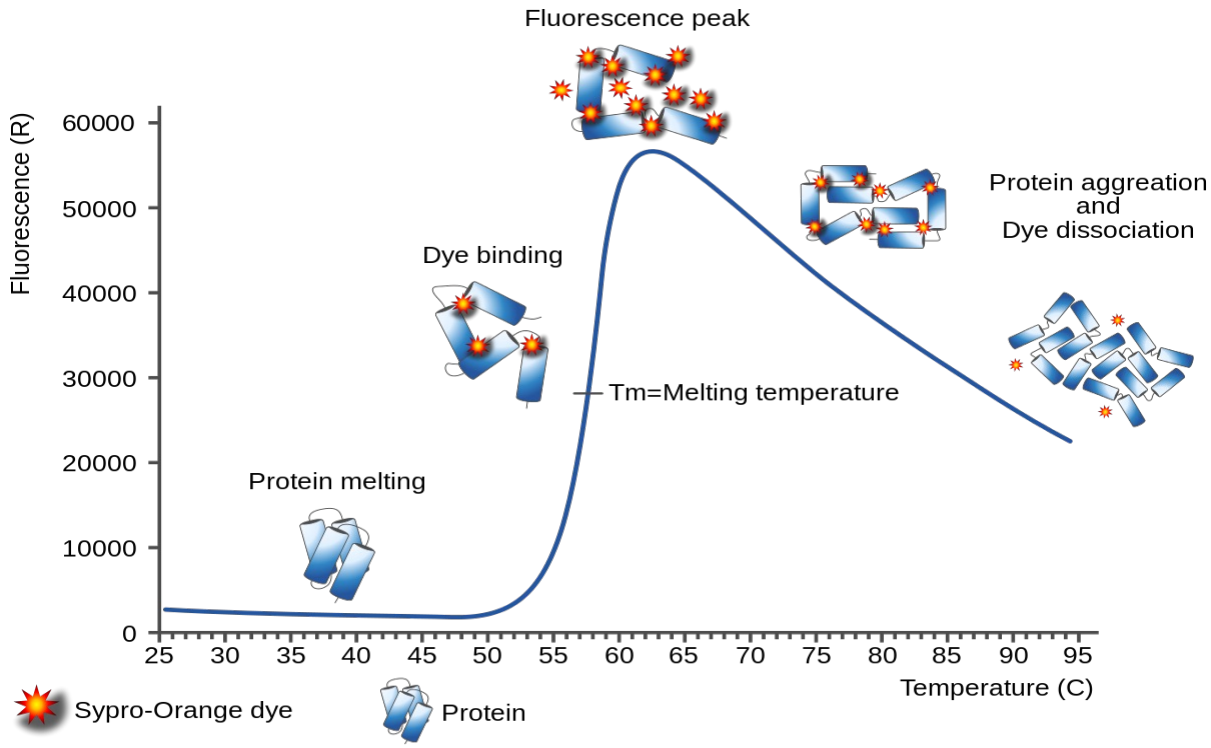
- A denatured protein makes many interactions with solvent water.
  - Considerable conformational freedom
  - High configurational entropy

Protein Refolding

- Misfolding of recombinant protein results in the formation of “inclusion bodies”.
- Many Proteins, especially small ones, denature *reversibly*
  - Regain native structures spontaneously when returning to conditions that favor folding
  - Renature methods: gradient dialysis/super dilution/affinity chromatography, etc.
- Many proteins, especially larger proteins denature irreversibly
  - Usually aggregate or precipitate after removing denaturant

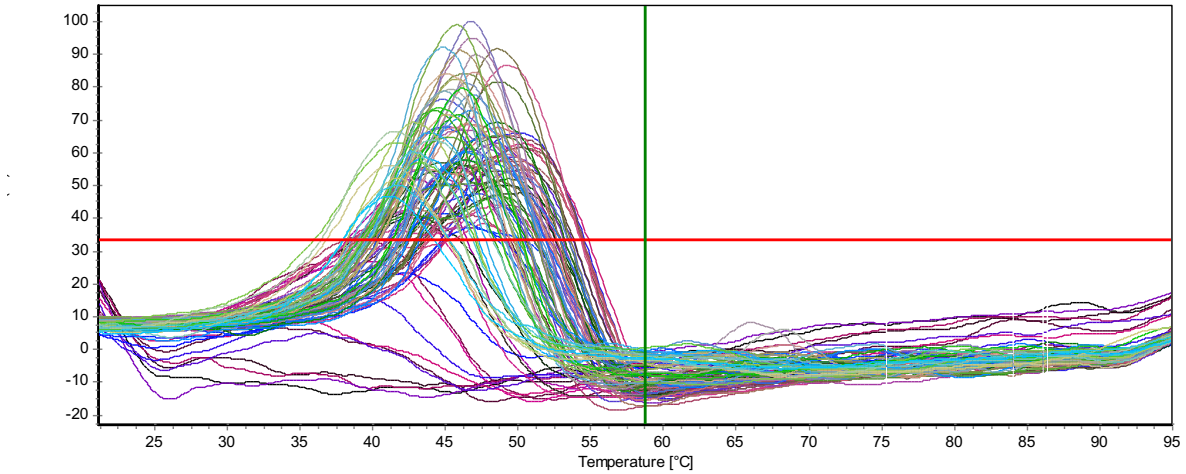
Thermal Shift Assay

- *Principle*: Protein hydrophobic core is buried. When the protein solution is heated up, the protein is denatured, and hydrophobic region is exposed to SYPRO™ orange fluorescent dye. The fluorescent signal can be determined in real time instrument.



Applications

1. Optimal storage buffer/conditions
2. Conditions for crystallography
3. Effect of mutations on protein stability
4. Protein-ligand interaction
5. Small molecule drug screening



Fluorescence Spectroscopy in Protein Folding/Unfolding Study

- Certain molecules remit the absorbed energy in the form of light--fluorescence.
- Of the proteinogenic amino acids, only Trp (Tyr) has significant fluorescence.
- Trp (or Tyr) fluorescence is sensitive to their environment which changes when protein folds/unfolds.
  - Increases in nonpolar environment (folded state)
  - Decreases in polar solvent (unfolded state)

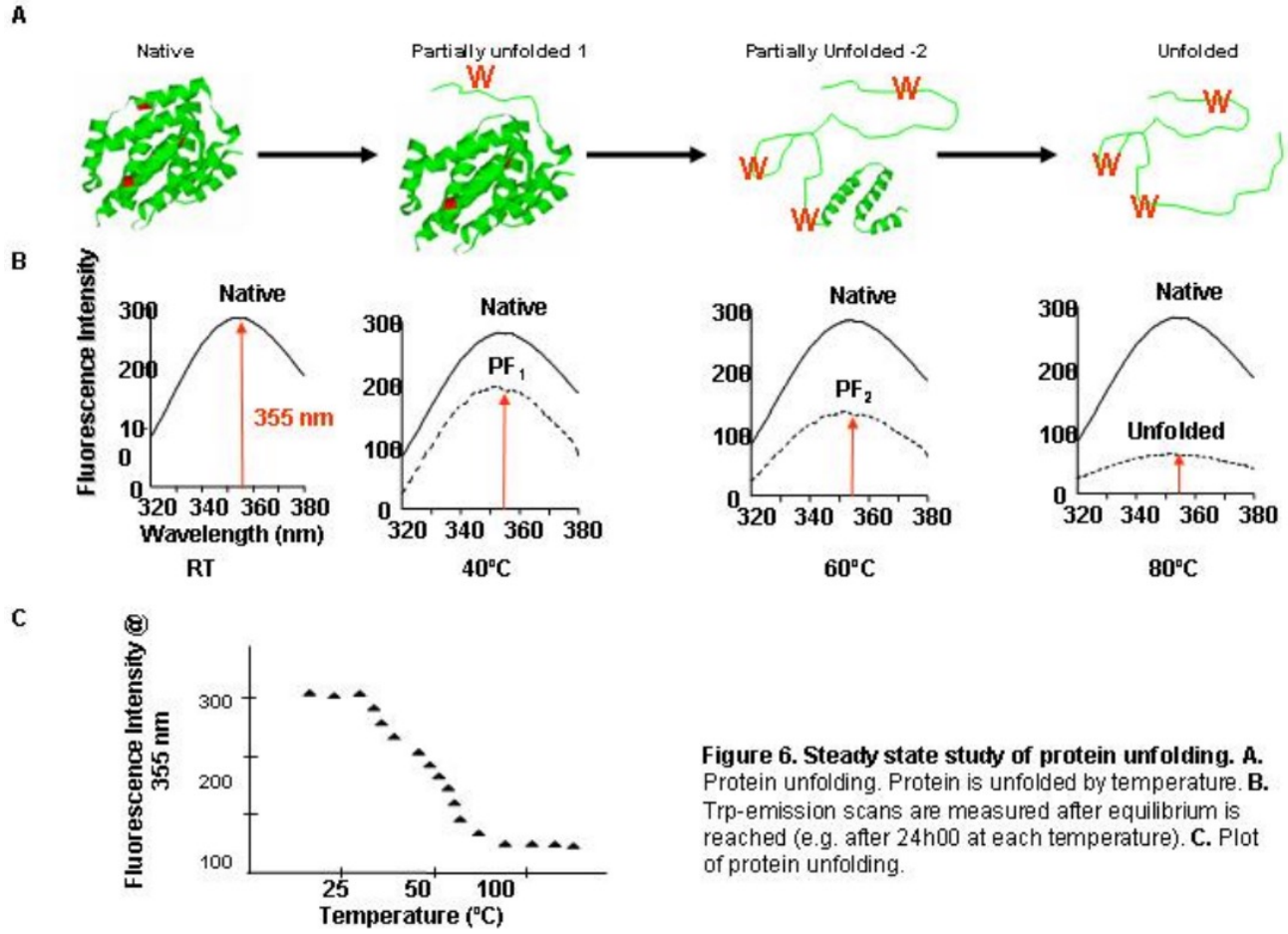


Figure 6. Steady state study of protein unfolding. A. Protein unfolding. Protein is unfolded by temperature. **B.** Trp-emission scans are measured after equilibrium is reached (e.g. after 24h00 at each temperature). **C.** Plot of protein unfolding.

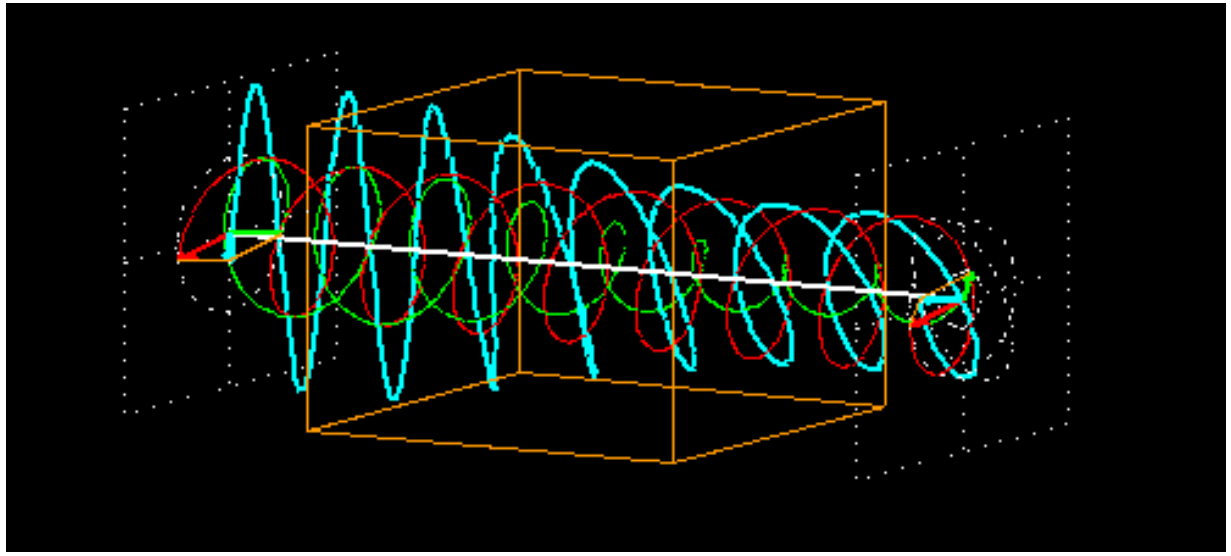
Protein Structure and Analytical Methods

Structure	Information Gained	Analytical Techniques
Primary	Amino acid sequence	Chemical method (Edman Sequencing); Mass Spectrometry
Secondary	Folding and composition of α helix, β sheet, random coils	Circular Dichroism (CD); Infrared Spectroscopy (IR)
Tertiary	Overall 3D structure/shape of a single polypeptide chain	X-ray crystallography; Nuclear Magnetic Resonance (NMR); Cryo-EM
Quaternary	Overall 3D structure/shape of two or more polypeptide chains	Size Exclusion Chromatography (SEC); Small Angle X-ray Scattering (SAXS) and Small-Angle Neutron Scattering (SANS)

LECTURE 8: CIRCULAR DICHROISM AND INFRARED SPECTROSCOPY

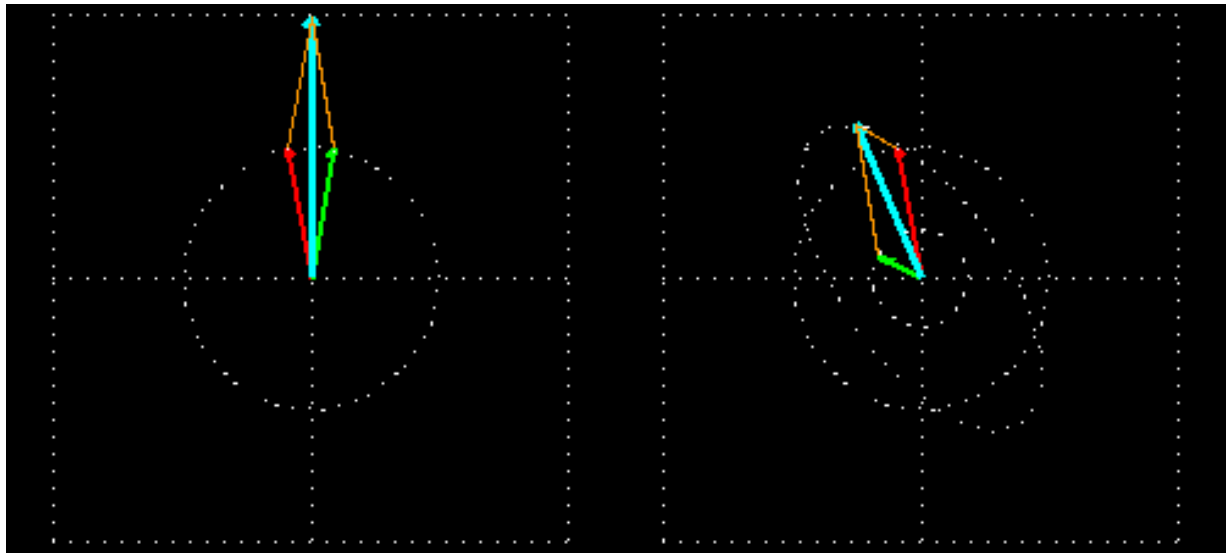
Circular Dichroism (CD)

- In circular polarized light, the field vector rotate about the axis of travel; the rotation can be clockwise or counterclockwise.
- Chiral molecules will absorb right and left circular polarized light **differently** and have different indices of refraction for the two waves.
- The resulting field vector does not oscillate along a straight line, but it *rotates and traces out an ellipsoid path.*



Before

After the sample



Circular Dichroism (CD)

- CD data are reported as the differential extinction coefficient:

$$\Delta\varepsilon = \varepsilon_L - \varepsilon_R = \Delta A/(bc)$$

- or the ellipticity:

$$\theta = 2.303 * \Delta A * 180/(4\pi)$$

- or mean residue ellipticity (MRE) in degrees cm^2 $dmol^{-1}$ per residue

$$MRE = {}^{\theta_{obs}}/_{(10nlc)}$$

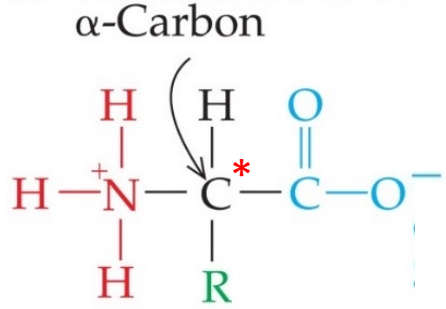
➢The CD data is measured as the function of wavelength.

➢For most biological samples, the observed CD signal or $\Delta\varepsilon$ is very small.

Circular Dichroism (CD) in Protein Study

➢Proteins are made up of chiral amino acids, and due to this fact, the structure of proteins are chiral.

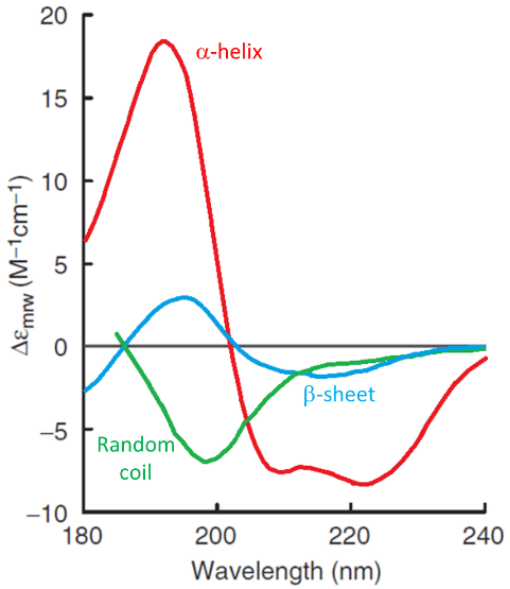
- α-Helix can be left-handed or right-handed



- In proteins, the chromophores of interest include the
  - peptide bond (absorption below 240 nm)
  - aromatic amino acid side chains (absorption in the range 260 to 320 nm)
  - disulfide bonds (weak broad absorption bands centered around 260 nm)
  - Cofactors (PLP: ~330 nm, FAD: 300-500 nm etc.)

Circular Dichroism (CD) for Protein Structure$_1$

- Different wavelength probe different levels of structure:
  1. **Far UV** (smaller wavelengths, 190-250 nm)
     - ➢ probing individual aa --> proportion of 2° structure in a protein

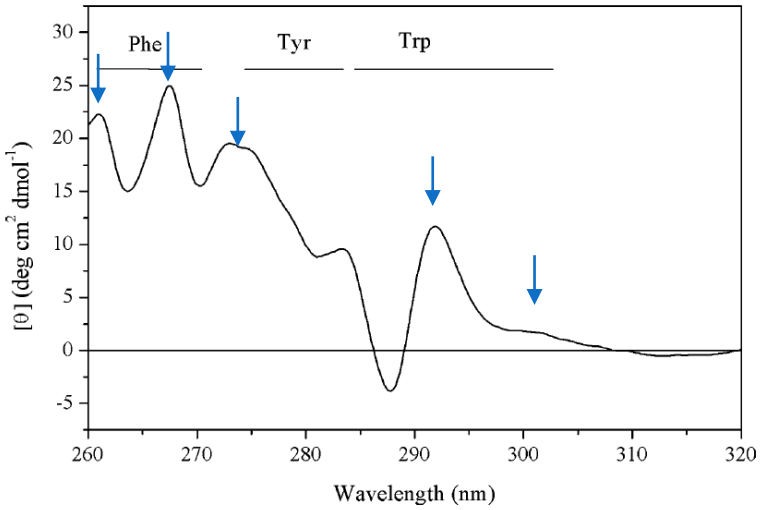


- When a sample is scanned over a range of wavelengths of polarized light, the various secondary structures (α helix, β sheet, and coils) all have specific (characteristic) CD signals.

- α helix : negative bands at 208 and 222 nm and positive band at 193 nm
- β sheet : negative band at 218 nm and positive band at 195 nm
- Random coil: negative band at 195 nm and low ellipticity above 210 nm

Circular Dichroism (CD) for Protein Structure$_2$

- Different wavelength probe different levels of structure:
  2. **Near UV** (longer wavelengths, 260-320 nm)
     - ➢ probing larger chiral structure (e.g., helices, sheets)--> 3^o structure information



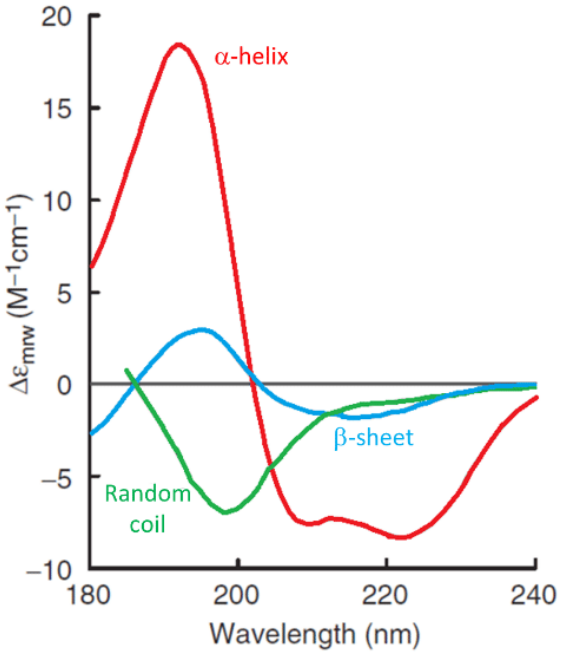
Note: CD **cannot** give any information about the residue specific information about the 3^o structure

– just **compare changes** in response to heating or addition of a ligand, for example.

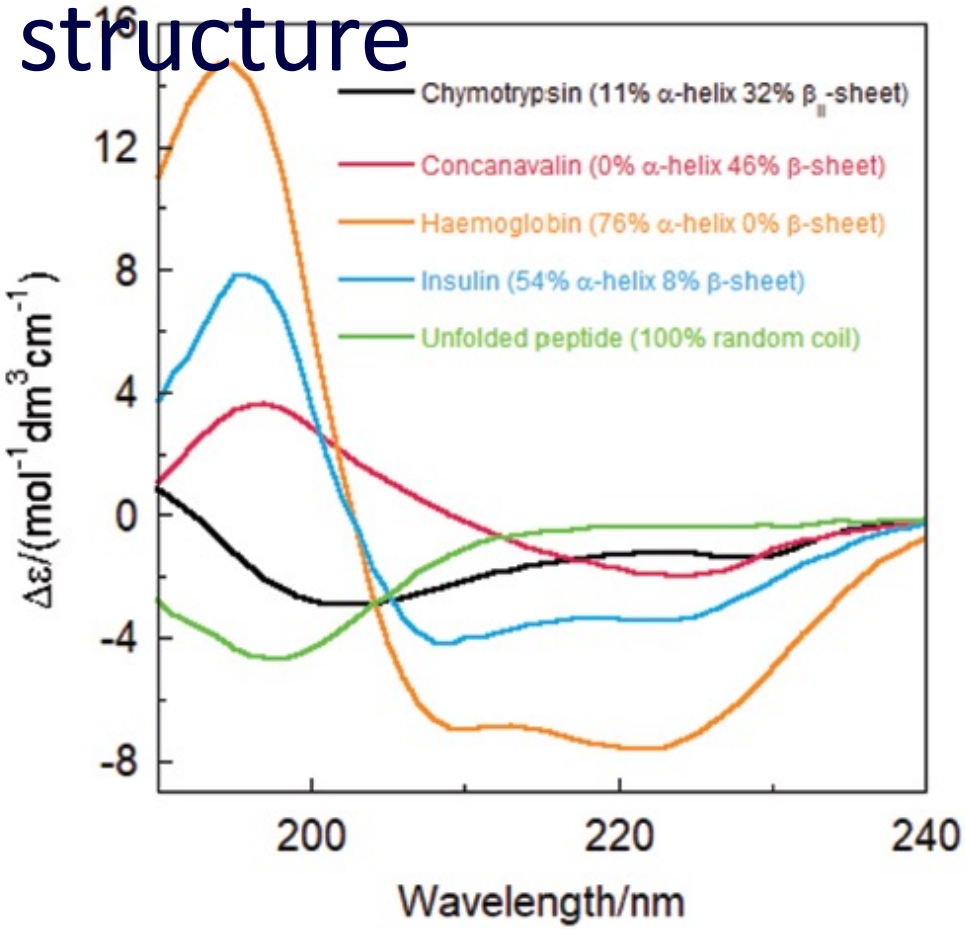
- Trp shows a peak close to 290 nm with fine structure between 290 and 305 nm
- Tyr shows a peak between 275 and 282 nm
- Phe shows weaker but sharper bands with fine structure between 255 and 270 nm

Applications of CD spectroscopy:

1. Relative amounts of secondary structure



- α helix : negative bands at 208 and 222 nm and positive band at 193 nm
- β sheet : negative band at 218 nm and positive band at 195 nm
- Random coil: negative band at 195 nm and low ellipticity above 210 nm



Far UV CD spectra of proteins with different 2º structure content

Alison Rodger, Doug Marshall; Beginners guide to circular dichroism. *Biochem (Lond)* 12 April 2021; 43 (2): 58–64. doi: https://doi.org/10.1042/bio_2020_105

2. Determine 2º Structure Composition

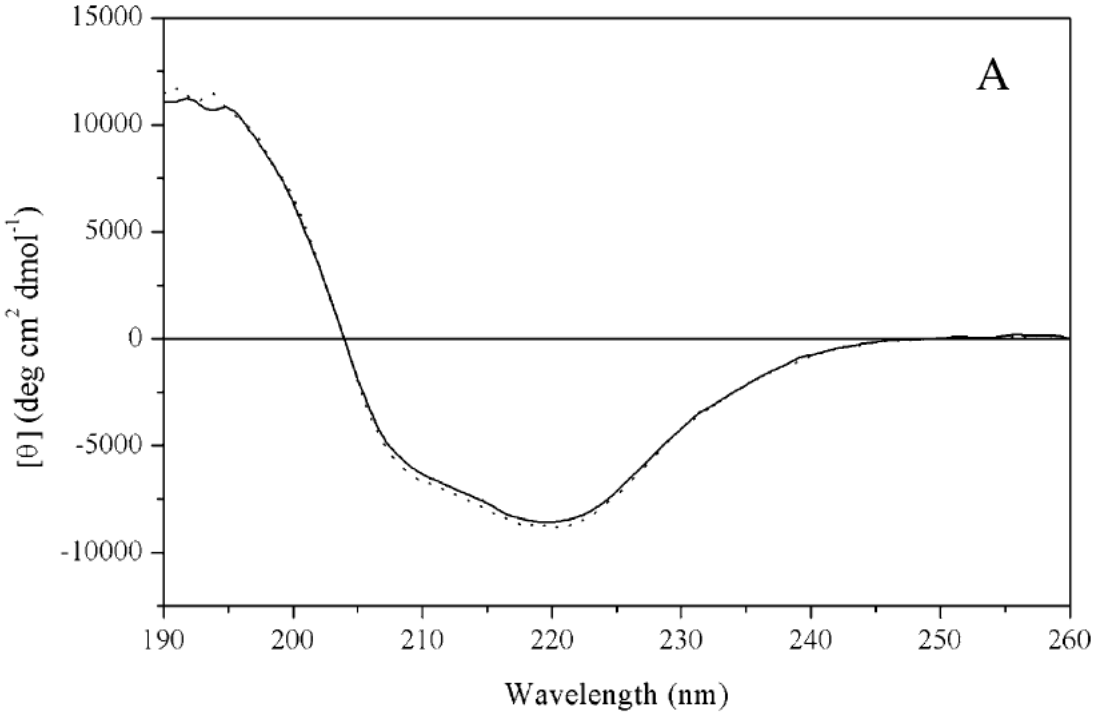
- The CD spectrum of a protein is a linear combination of the spectra of its secondary structure elements, plus a noise term (aromatic chromophores and prosthetic groups).

$$\theta_\lambda = \sum \varepsilon_i S_{\lambda i} + noise$$

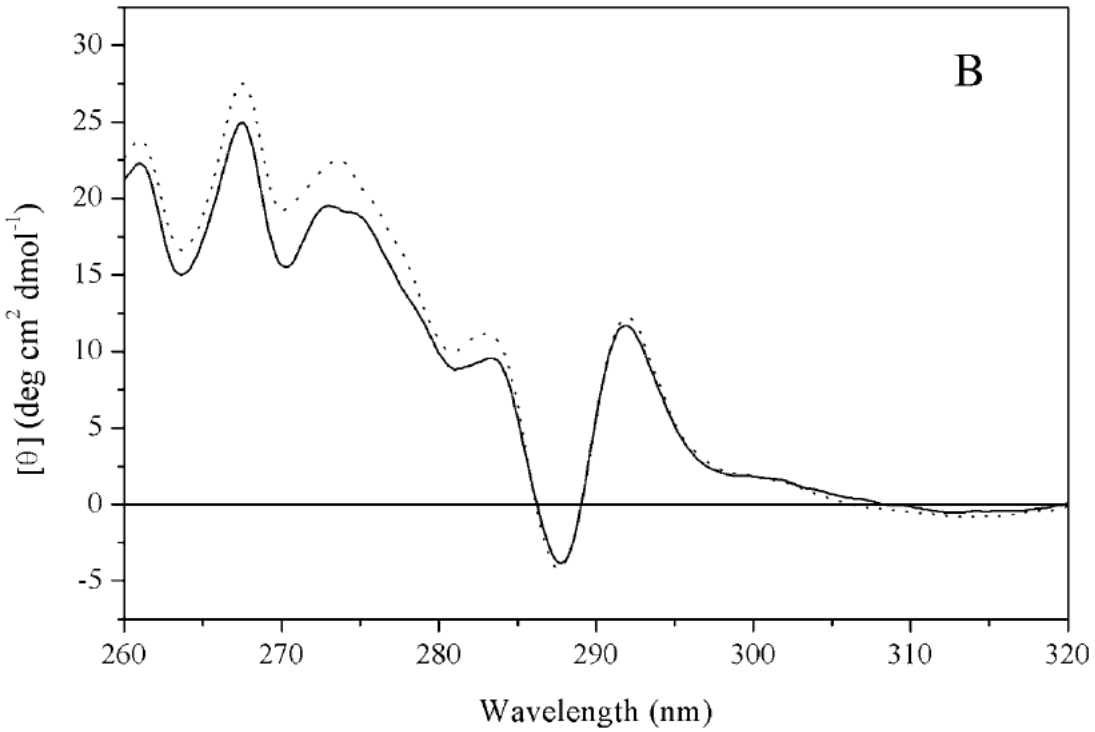
  - Where θ_λ is the CD of the protein as a function of wavelength; ε_i is the fraction of each secondary structure, i; $S_{\lambda i}$ is the ellipticity at each wavelength of each i^{th} secondary structure.
- A good reference set of CD spectra from proteins of known structure (e.g., by X-ray crystallization) is needed.
  - Protein Circular Dichroism Data Bank and the Dichroweb site.
- Note: reliable analysis of secondary structure, it is necessary to ensure that the concentration of the protein solution is ***accurately*** known.

3. Comparison of WT vs Mutant Protein (Structural Integrity)

wild-type (solid line)
mutant R23Q (dotted line): no activity



Far UV scan



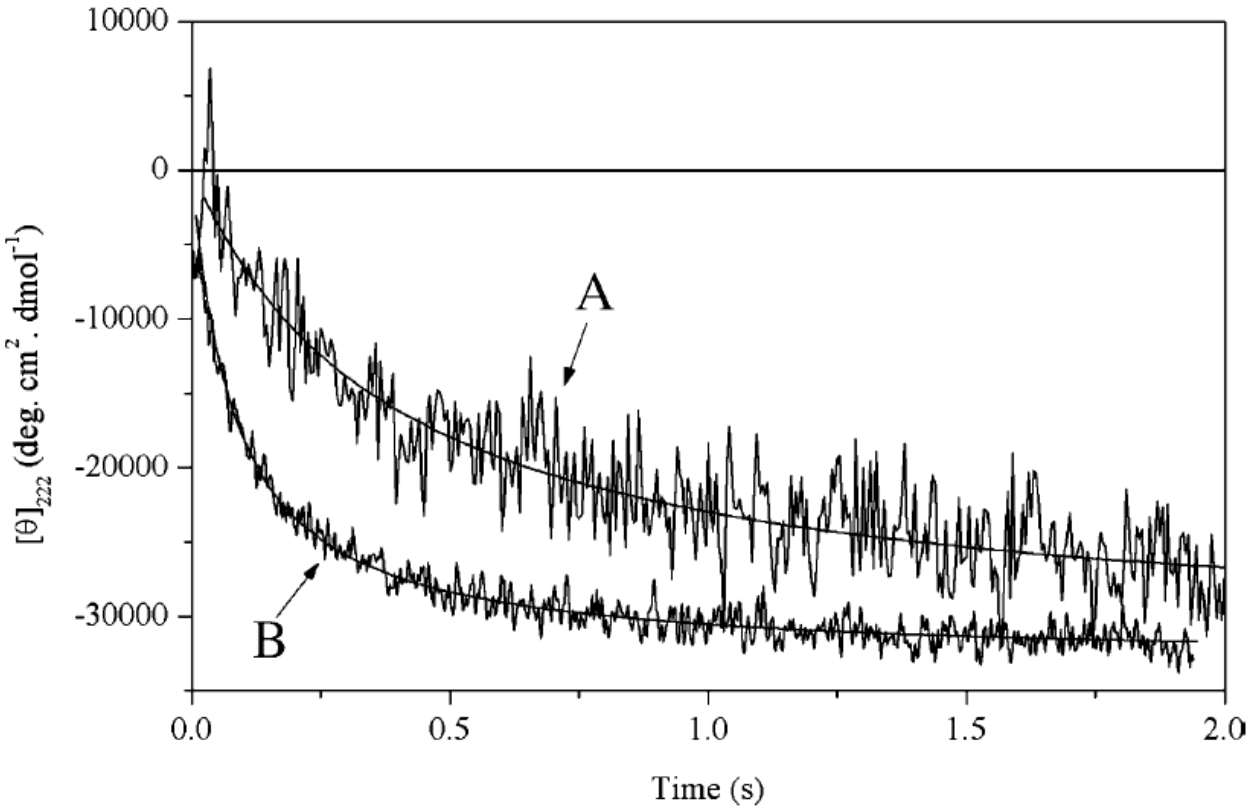
Near UV scan

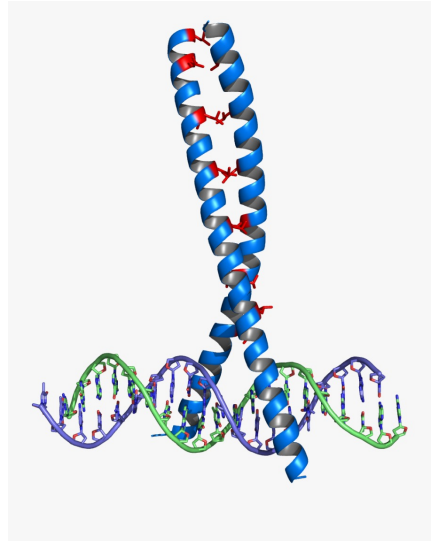
S.M. Kelly et al. / Biochimica et Biophysica Acta 1751 (2005) 119– 139

T. Krell, M.J. Horsburgh, A. Cooper, S.M. Kelly, J.R. Coggins. J. Biol. Chem. 271 (1996) 24492–24497

4. Structural Stability (Folding or Unfolding)

The refolding of a leucine zipper peptide monitored by stopped flow CD at 222 nm.





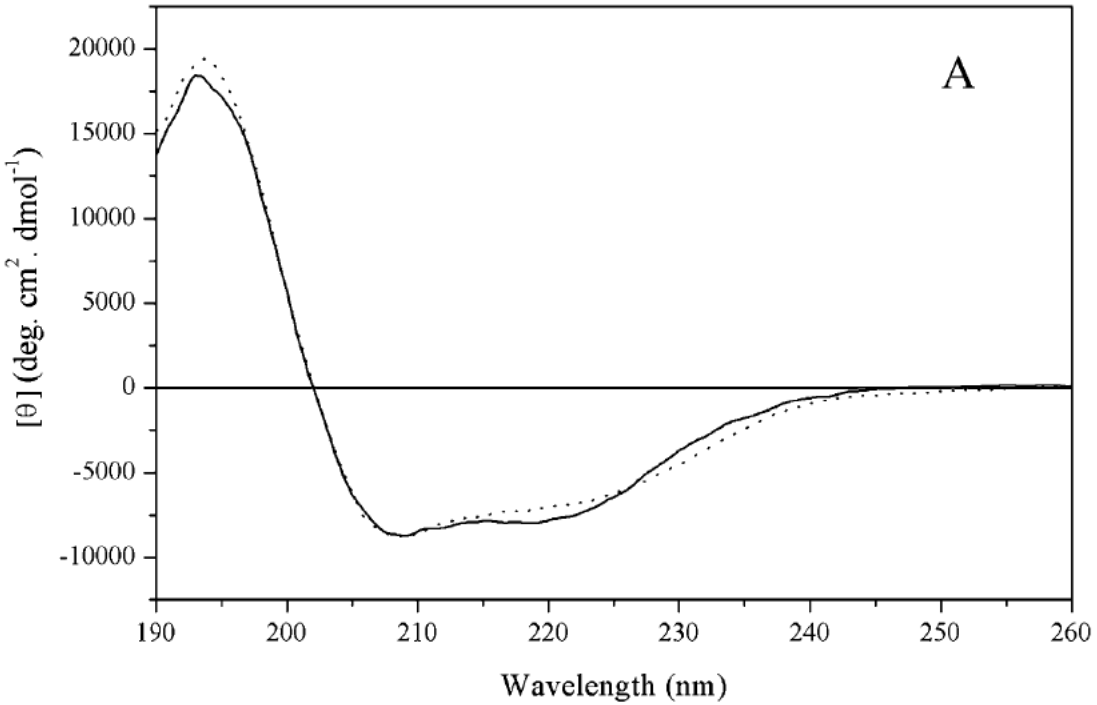
33-residue peptide at concentrations of 6 µM (A) and 26 µM (B) was refolded after denaturation in GdmCl.

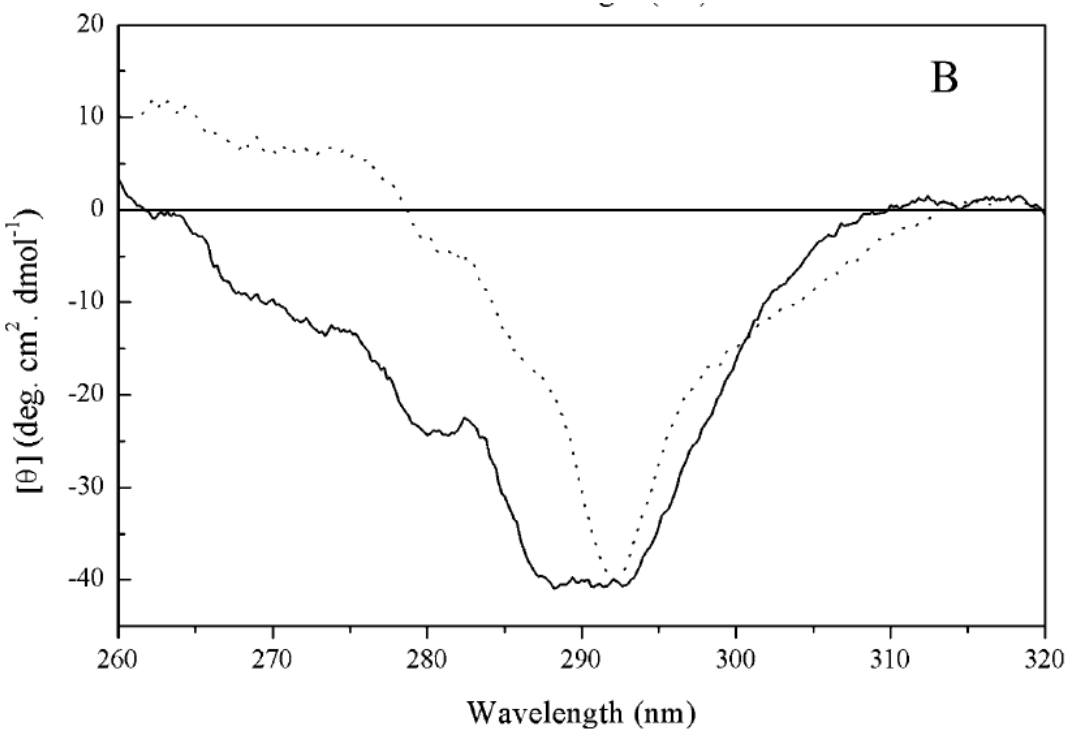
Zitzewitz JA, Bilsel O, Luo J, Jones BE, Matthews CR. Biochemistry. 1995 Oct 3;34(39):12812-9. doi: 10.1021/bi00039a042. PMID: 7548036.

5. Conformational Changes in Protein-Ligand Binding$_1$

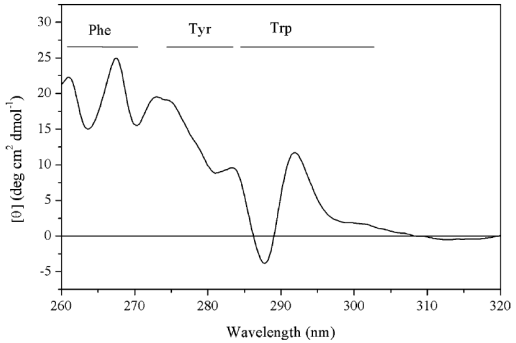
molybdate-sensing protein ModE (solid line)

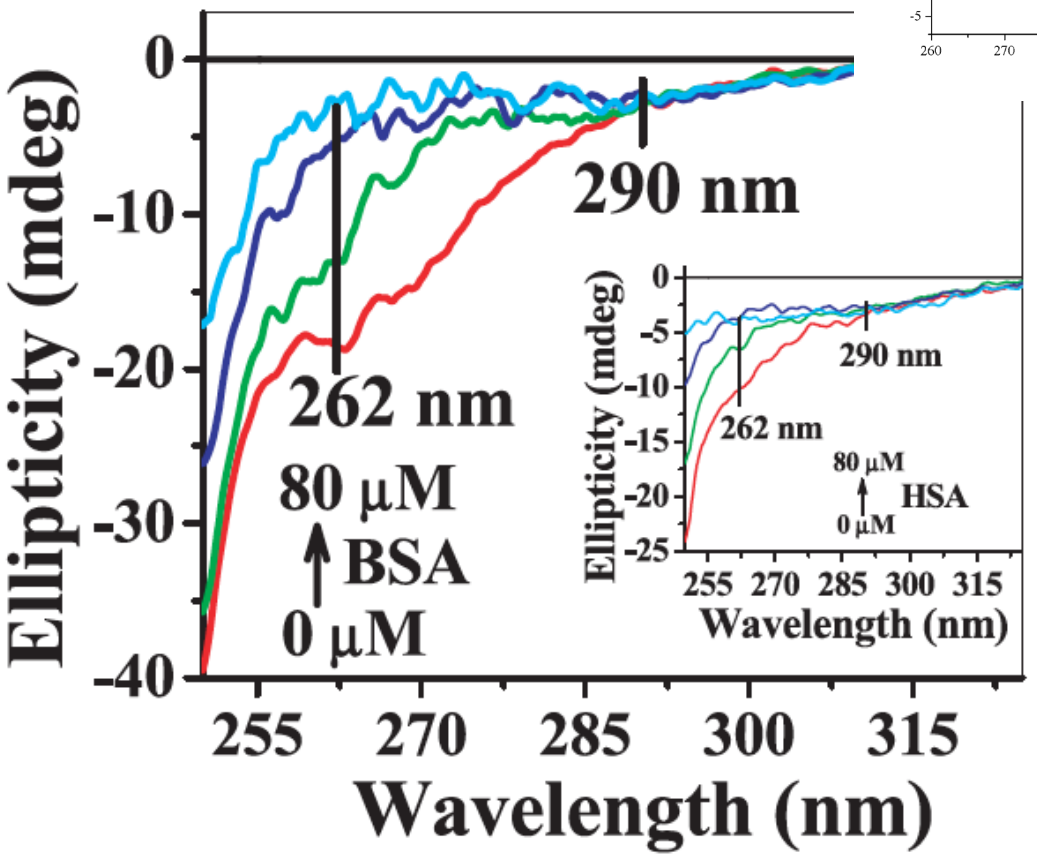
molybdate-sensing protein ModE + 1 mM molybdate (dotted line)

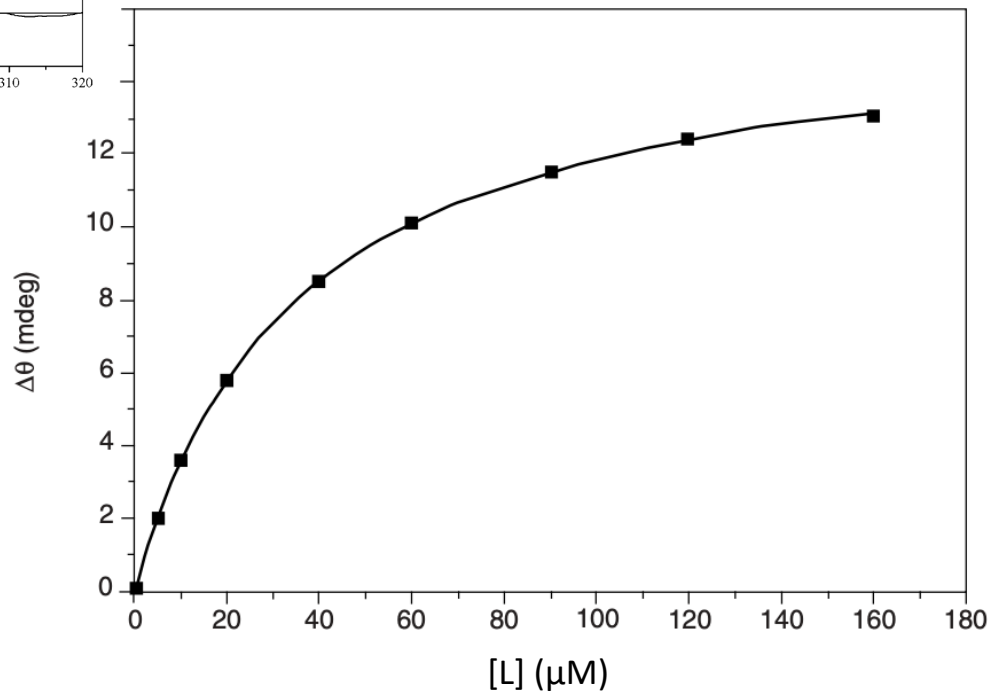




5. Conformational Changes in Protein-Ligand Binding$_2$





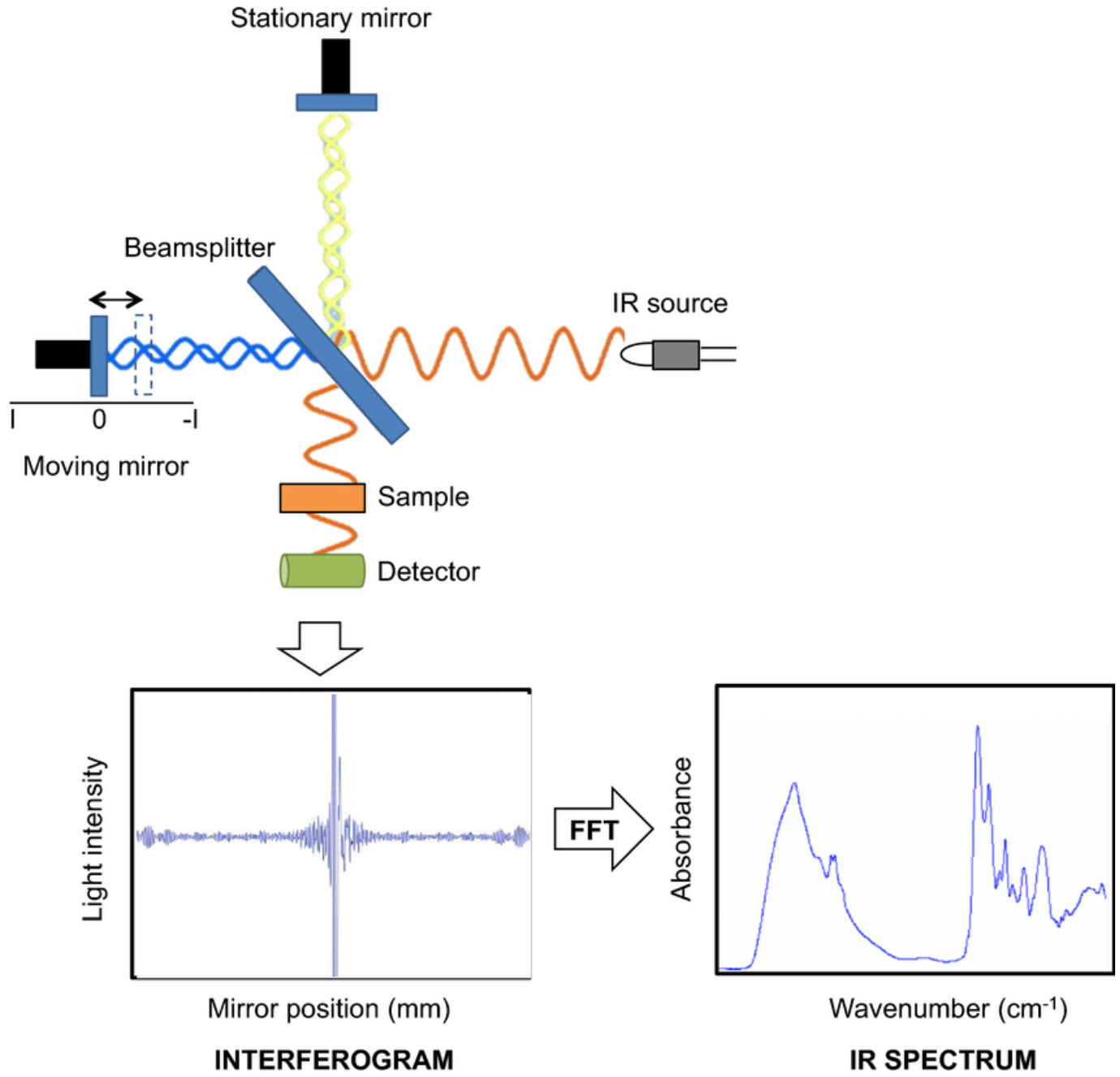


$$\Delta\theta = \Delta\theta_{max}\left([L_t]/(K_d + [L_t])\right)$$

Infrared (IR) Spectroscopy

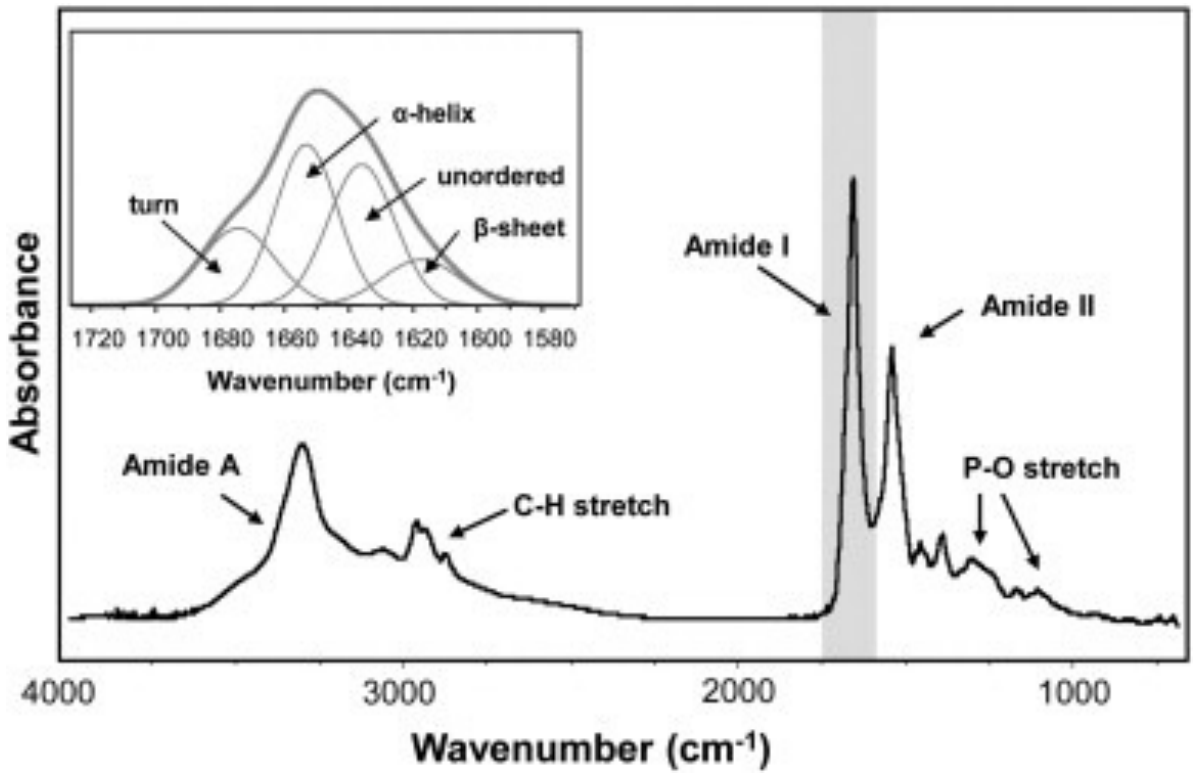
- IR spectroscopy studies the interaction between matter and infrared radiation.
- IR radiation is absorbed at characteristic wavelengths (wave number, $1/\lambda$) by different groups.
  - C=O groups absorb in the IR at 1550-1750 cm^{-1}
- IR is not a great technique for studying proteins, because water absorbs IR radiation strongly in regions that overlap the protein absorbance bands.

Fourier Transform IR (FTIR)



Campanella, Beatrice & Palleschi, Vincenzo & Legnaioli, S.. (2021). Introduction to vibrational spectroscopies. ChemTexts. 7. 10.1007/s40828-020-00129-4.

FTIR Spectra



- FTIR can be used to determine the amounts of each kind of 2° structure and the conformation analysis.
- Amide I and amide II band shape is very characteristic for a protein's 2° structure.

Lisa M. Miller, Megan W. Bourassa, Randy J. Smith. Biochimica et Biophysica Acta (BBA) - Biomembranes, Volume 1828, Issue 10, 2013, Pages 2339-2346, ISSN 0005-2736