Methods for determining protein structure

- Sequence:
  - Edman degradation
  - Mass spectrometry
- Secondary structure:
  - Circular Dichroism
  - FTIR
- Tertiary, quaternary structure:
  - NMR
  - X-ray crystallography

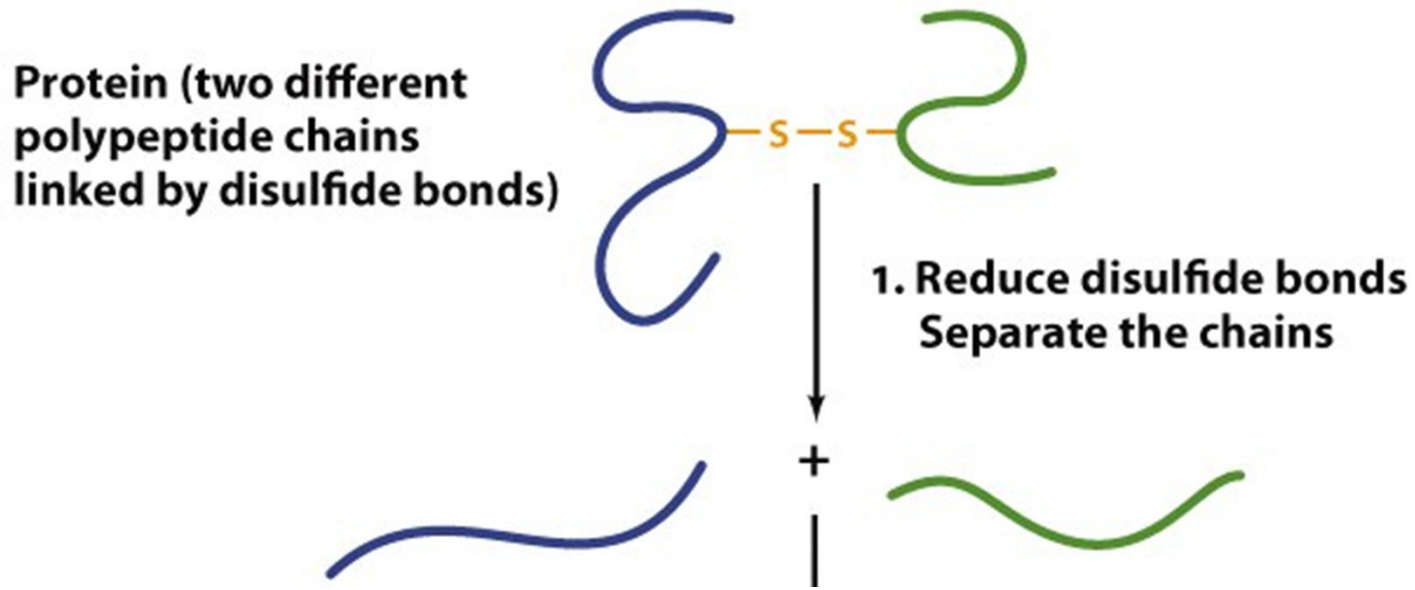
Protein sequencing approaches depend on what is known and what is the goal

- Protein is unknown, from organism with no DNA sequence information – starting from scratch
  - Purify protein & separate chains (if multimer)
  - Fragment and sequence each chain
  - Fragment differently and sequence
  - Reassemble sequence based on overlapping fragments

- Protein is unknown or known, and comes from an organism with known DNA sequence
  - Purify protein (& separate chains)
  - Fragment chain(s) and sequence or measure mass
  - Use sequence database to reassemble sequence

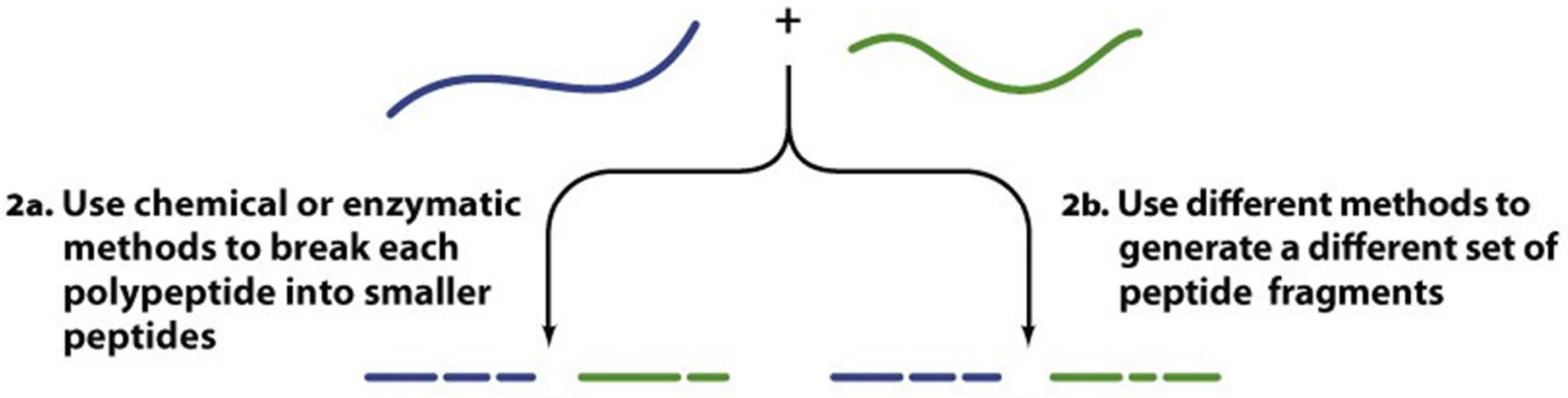
Protein sequencing from scratch

- Step 0: Purify the protein
- Step 1: Separate the chains (if multimeric)
  - If needed, reduce disulfides and block free thiols



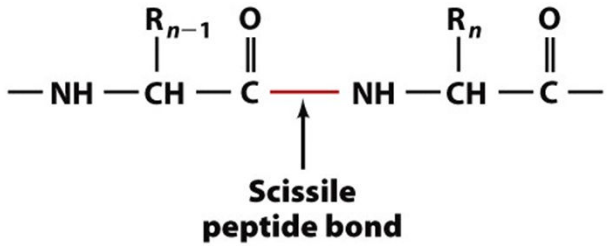
Protein sequencing from scratch

- Step 0: Purify the protein
- Step 1: Separate the chains (if multimeric)
- Step 2: Fragment each polypeptide
  - Enzymatically, with endopeptidase, chemically (e.g. with cyanogen bromide), or physically (e.g. through collision in MS)



Step 2: Fragment each polypeptide

Table 5-3 Specificities of Various Endopeptidases



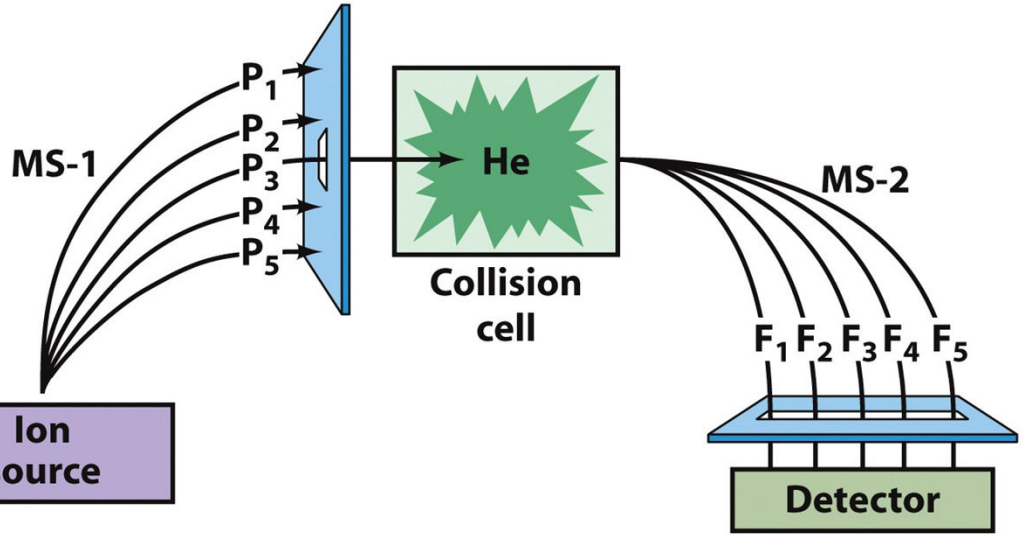
Enzyme	Source	Specificity	Comments
Trypsin	Bovine pancreas	R_{n-1} = positively charged residues: Arg, Lys; $R_n \neq$ Pro	Highly specific
Chymotrypsin	Bovine pancreas	R_{n-1} = bulky hydrophobic residues: Phe, Trp, Tyr; $R_n \neq$ Pro	Cleaves more slowly for R_{n-1} = Asn, His, Met, Leu
Elastase	Bovine pancreas	R_{n-1} = small neutral residues: Ala, Gly, Ser, Val; $R_n \neq$ Pro	
Thermolysin	*Bacillus thermoproteolyticus*	R_n = Ile, Met, Phe, Trp, Tyr, Val; $R_{n-1} \neq$ Pro	Occasionally cleaves at R_n = Ala, Asp, His, Thr; heat stable
Pepsin	Bovine gastric mucosa	R_n = Leu, Phe, Trp, Tyr; $R_{n-1} \neq$ Pro	Also others; quite nonspecific; pH optimum = 2
Endopeptidase V8	*Staphylococcus aureus*	R_{n-1} = Glu	

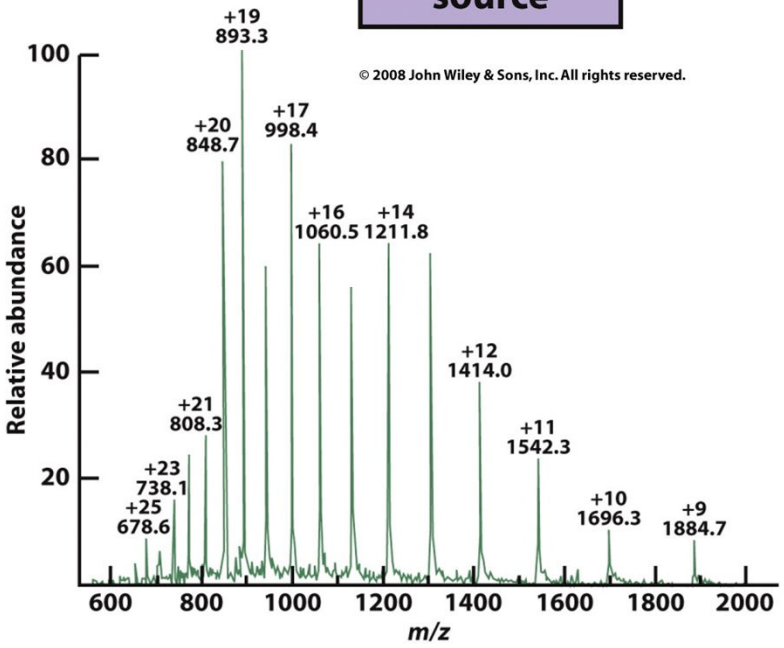
Cyanogen bromide (CNBr): R_{n-1} = Met

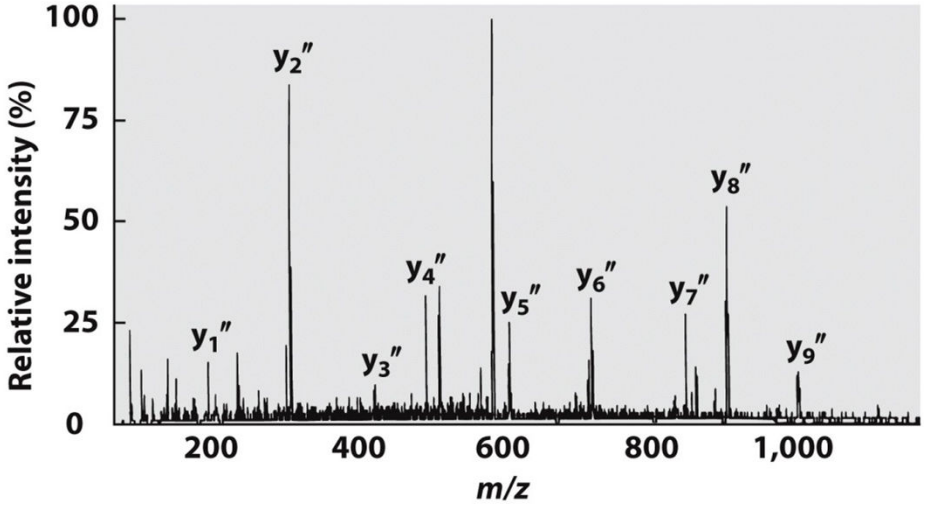
Protein sequencing from scratch

- Step 0: Purify the protein
- Step 1: Separate the chains (if multimeric)
- Step 2: Fragment each polypeptide
- Step 3: Sequence the fragments
  - Via, e.g., Edman degradation or Mass spectrometry

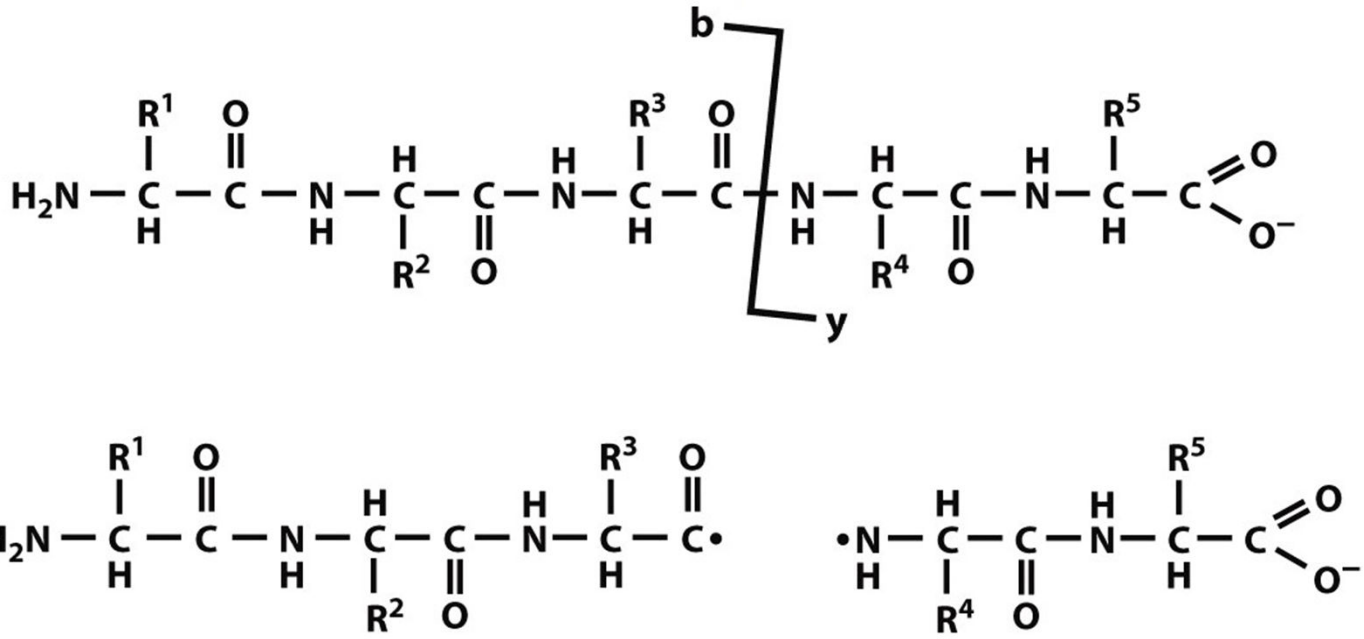
Sequence peptides with mass spectrometry (MS/MS)





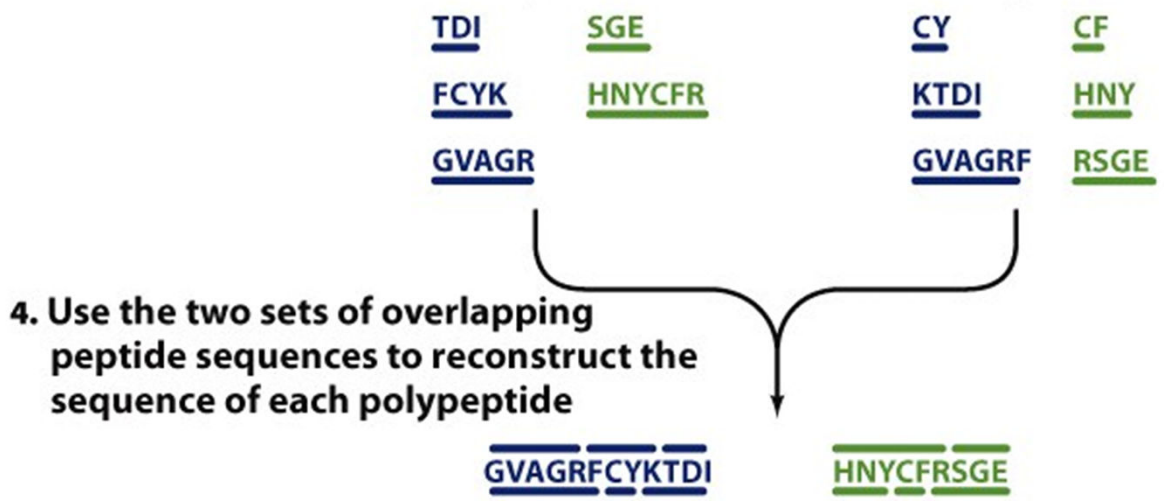


MS cleavage occurs mainly at peptide bonds, and charge is retained in one product



Protein sequencing from scratch

- Step 0: Purify the protein
- Step 1: Separate the chains (if multimeric)
- Step 2: Fragment each polypeptide
- Step 3: Sequence the fragments
- Step 4: Reconstruct the sequence

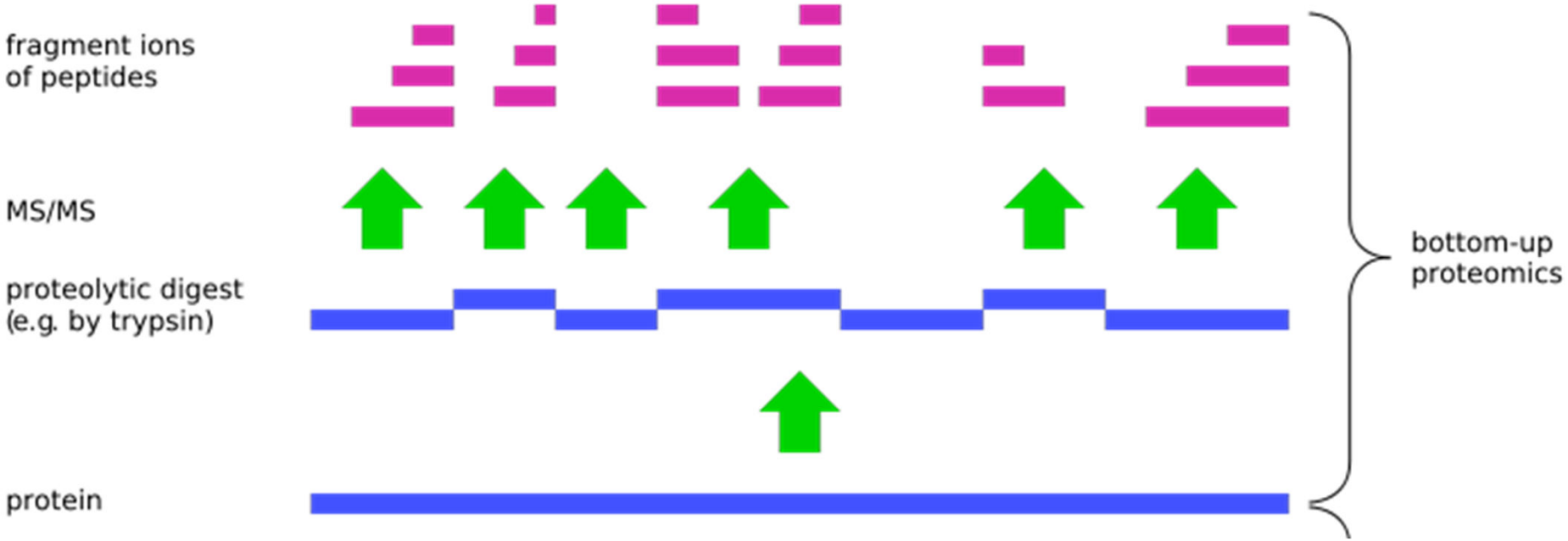


Protein sequencing approaches depend on what is known and what is the goal

- Protein is unknown, from organism with no DNA sequence information – starting from scratch
  - Purify protein & separate chains (if multimer)
  - Fragment and sequence each chain
  - Fragment differently and sequence
  - Reassemble sequence based on overlapping fragments

- Protein is unknown or known, and comes from an organism with known DNA sequence
  - Purify protein (& separate chains)
  - Fragment chain(s) and sequence or measure mass
  - Use sequence database to reassemble sequence

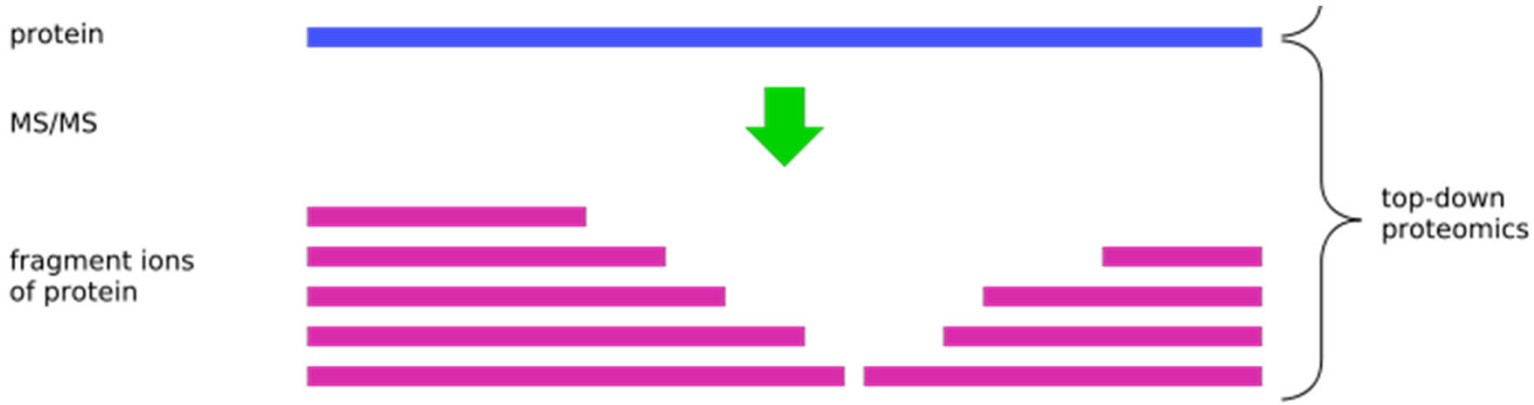
There are different approaches for using mass spectrometry to sequence a protein



Bottom-Up Proteomics

- Fragment protein (e.g. enzymatically) and separate fragments
- Ionize fragments, trap in the spectrometer, and measure m/z
- Select one m/z peak and fragment (e.g. by collision)
- Measure m/z of the smaller fragments and use a database to match the peaks to known sequences

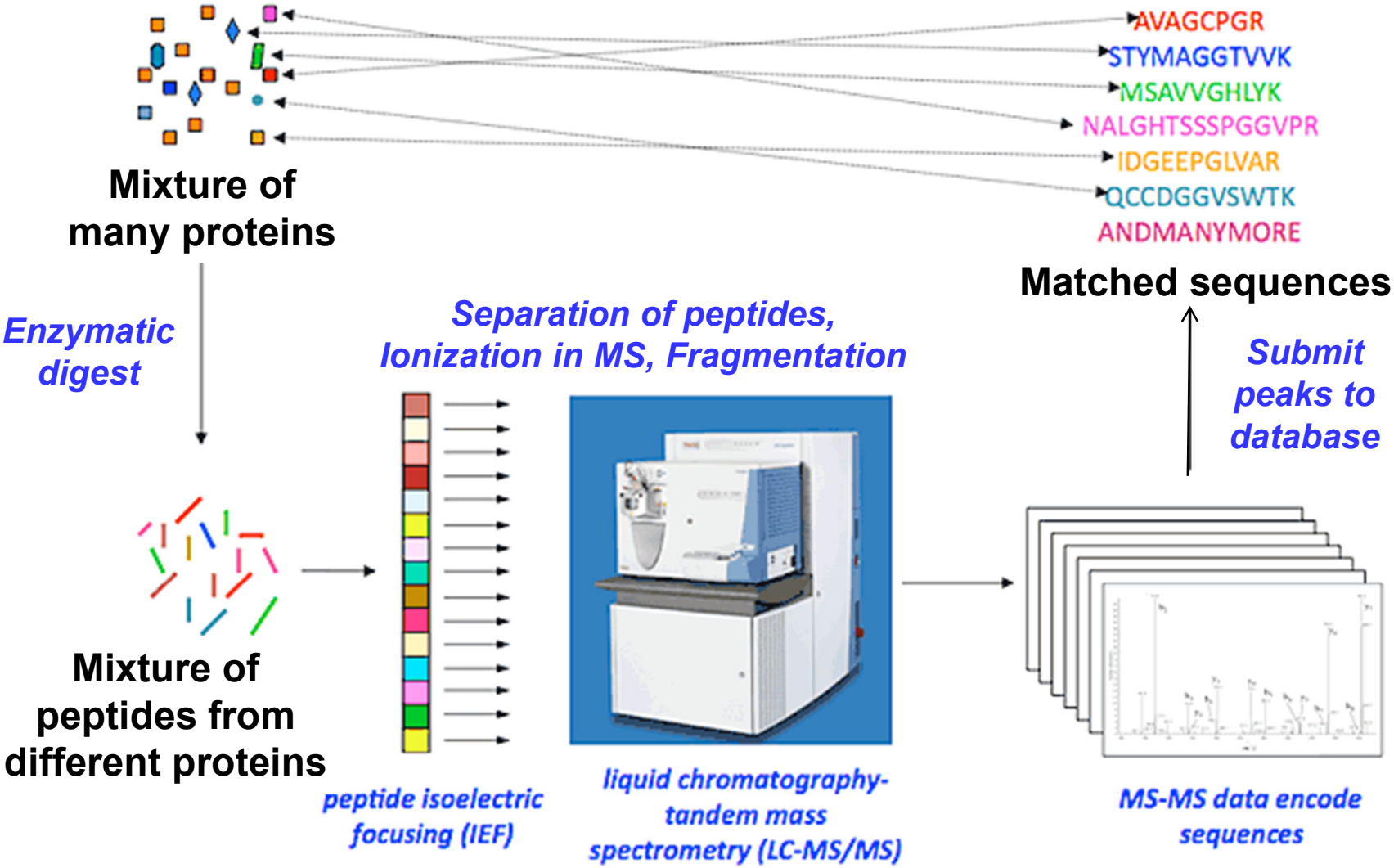
There are different approaches for using mass spectrometry to sequence a protein



Top-Down Proteomics

- Ionize *whole* protein(s), trap in the spectrometer, and measure m/z
- Use the instrument to select one m/z peak and fragment the protein (e.g. by collision)
- Measure m/z ratios of the fragments and use a database to match the peaks to known sequences
- OR Select a peak and fragment again, then match to sequence (Selection and fragmentation may be repeated over and over)

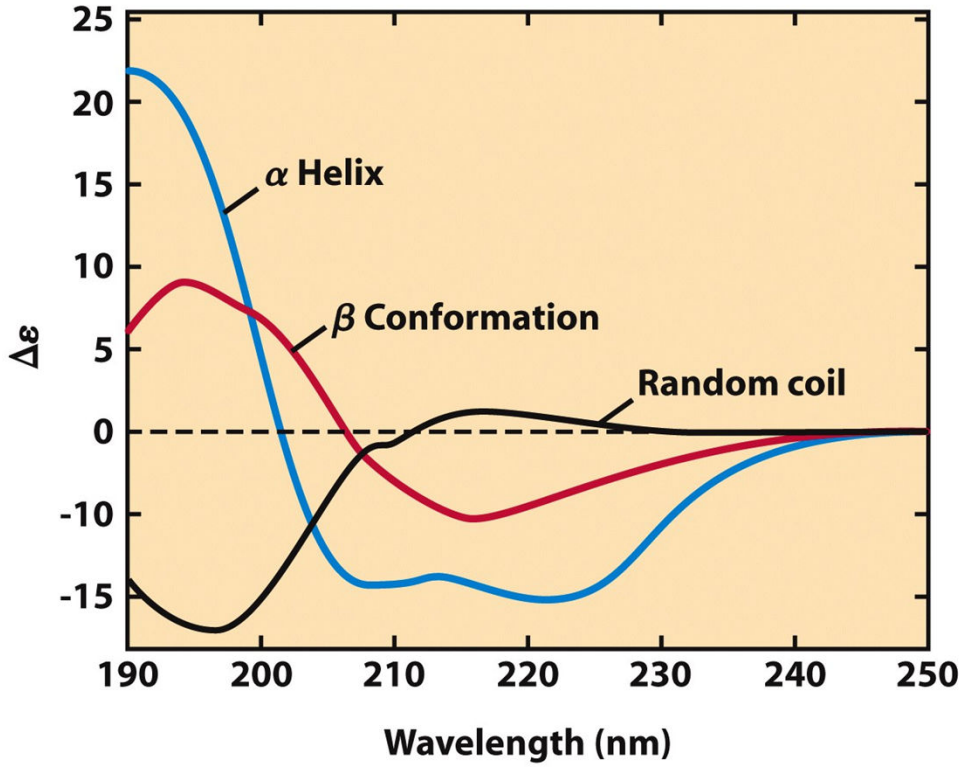
In shotgun proteomics, mass spec. is used to sequence mixtures of proteins



Methods for determining protein structure

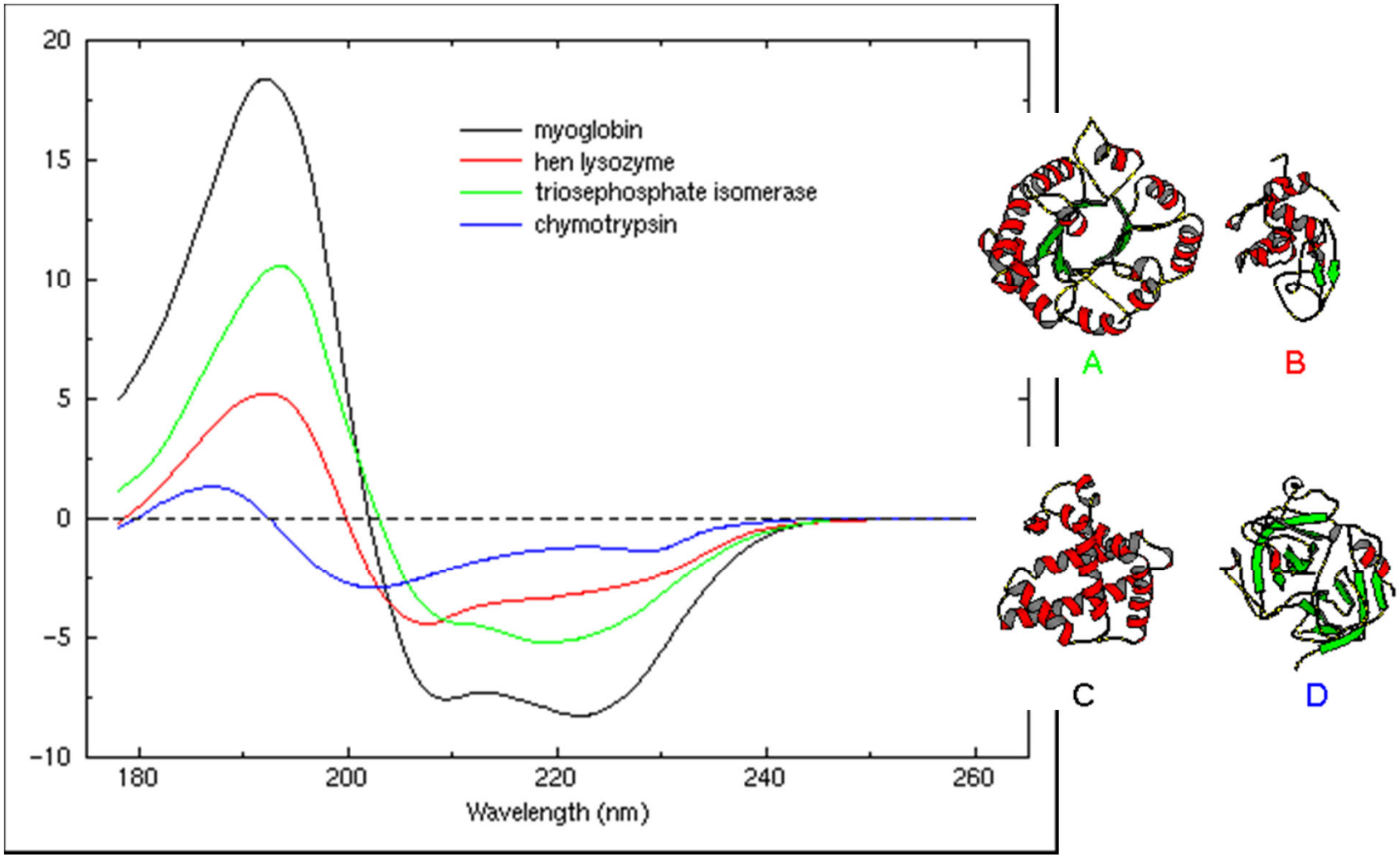
- Sequence:
  - Edman degradation
  - Mass spectrometry
- Secondary structure:
  - Circular Dichroism
  - FTIR
- Tertiary, quaternary structure:
  - NMR
  - X-ray crystallography

Circular dichroism (CD) measures amide absorption of circularly polarized UV light

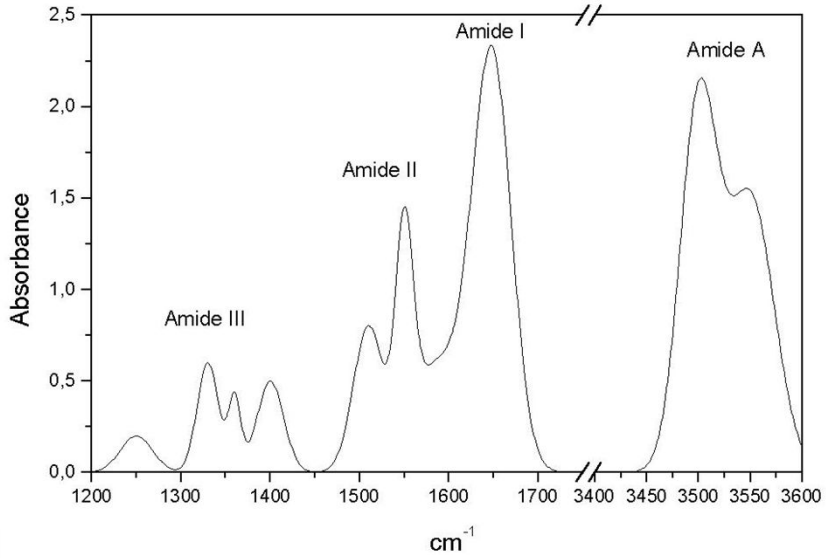


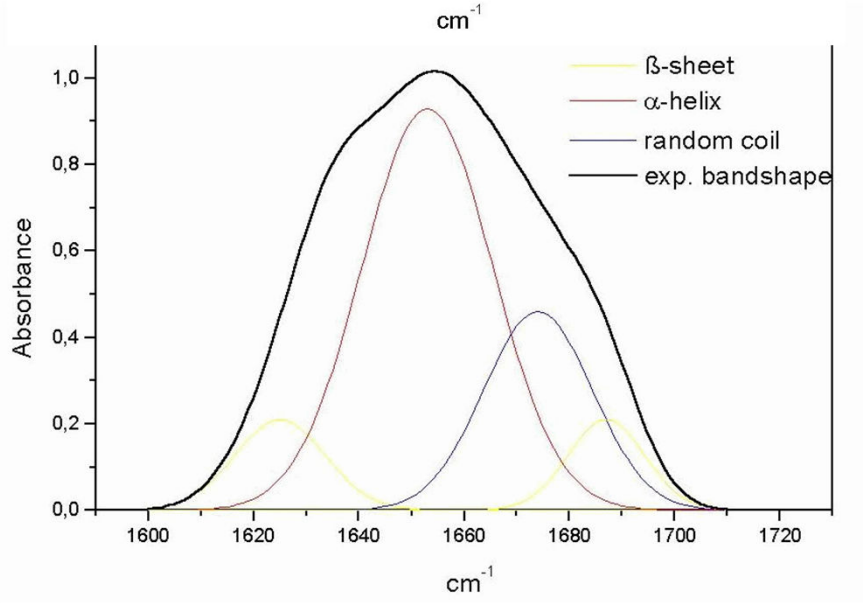
- Ellipticity ($\Delta\varepsilon$) is the difference in absorption of left-handed and right-handed circularly polarized light
- Different secondary structures show different patterns of ellipticity
- Protein's CD spectrum is 'deconvoluted' to estimate fractional contribution of helix, sheet, turn, and coil

Proteins with different compositions of 2° structure give different CD spectra



Fourier transform infrared (FTIR) spectra show amide absorption of infrared light



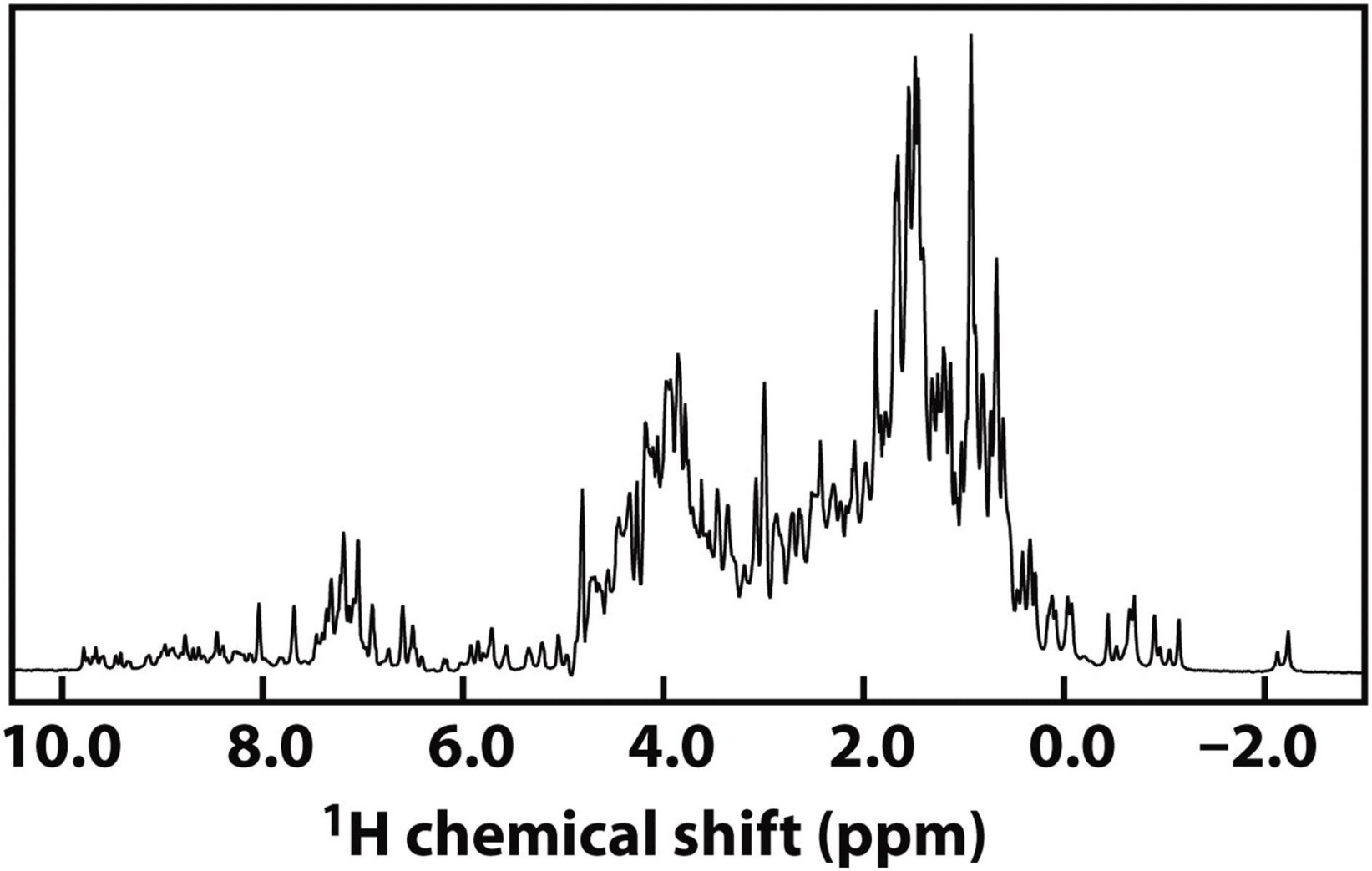


- Peak frequencies show bond stretching and bending, which vary with protein conformation
- C=O stretching frequency of amide I band correlates with secondary structure
- Protein's FTIR spectrum is 'deconvoluted' to estimate fractional contribution of helix, sheet, and coil

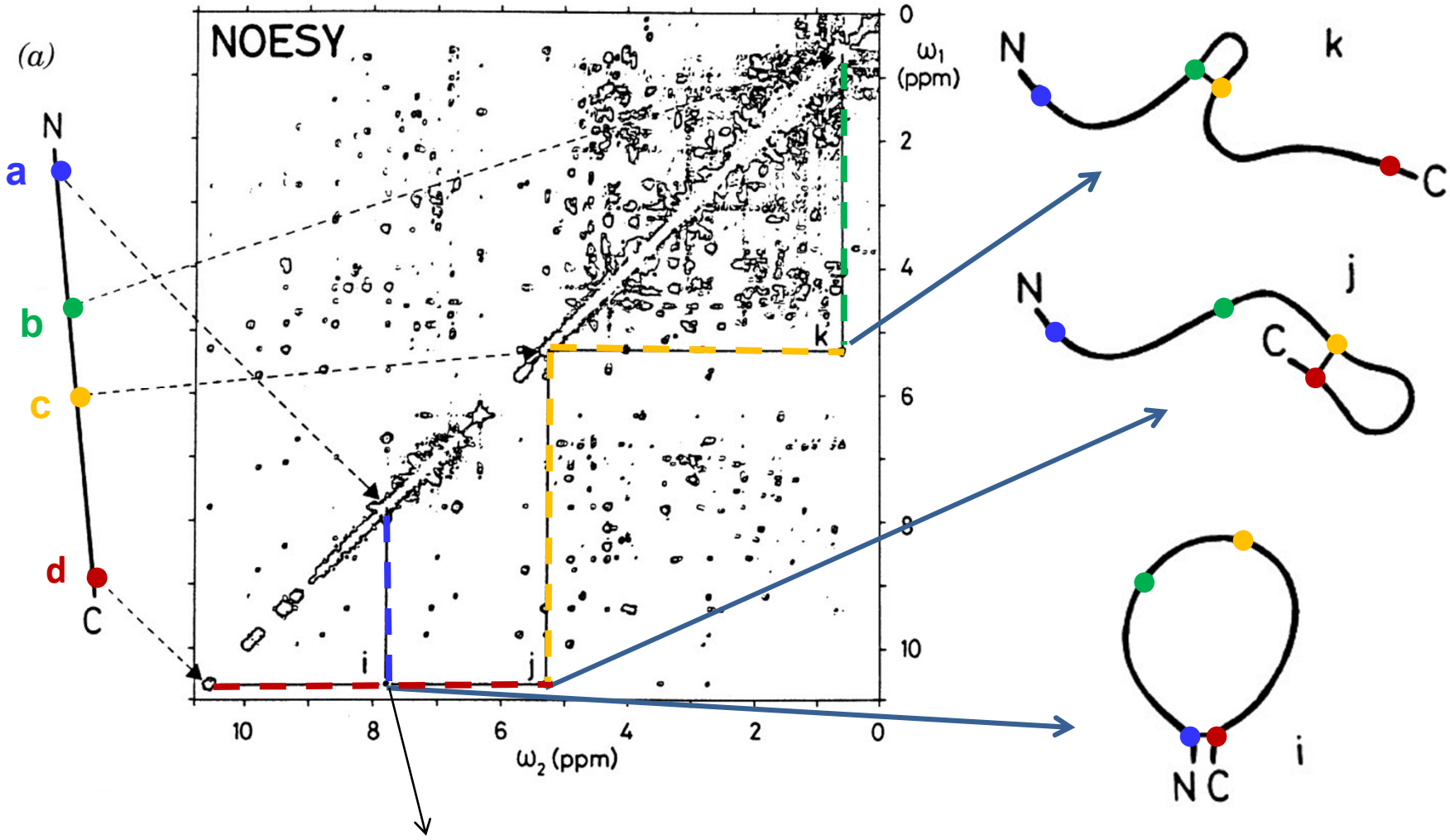
Methods for determining protein structure

- Sequence:
  - Edman degradation
  - Mass spectrometry
- Secondary structure:
  - Circular Dichroism
  - FTIR
- Tertiary, quaternary structure:
  - NMR
  - X-ray crystallography

Proteins have too many protons to be resolved by one-dimensional NMR

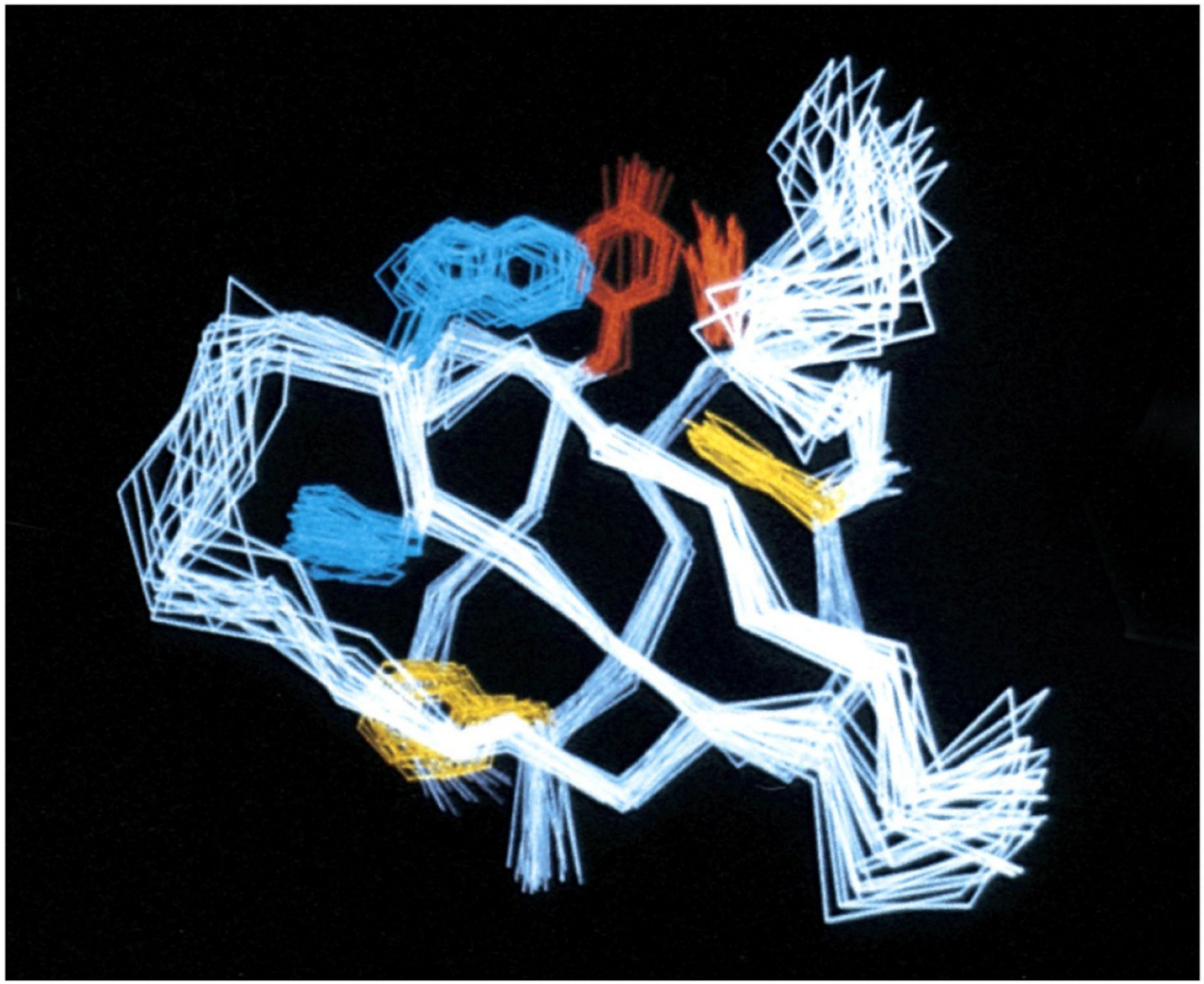


2D NMR separates proton peaks and can reveal approximate distances between nearby atoms



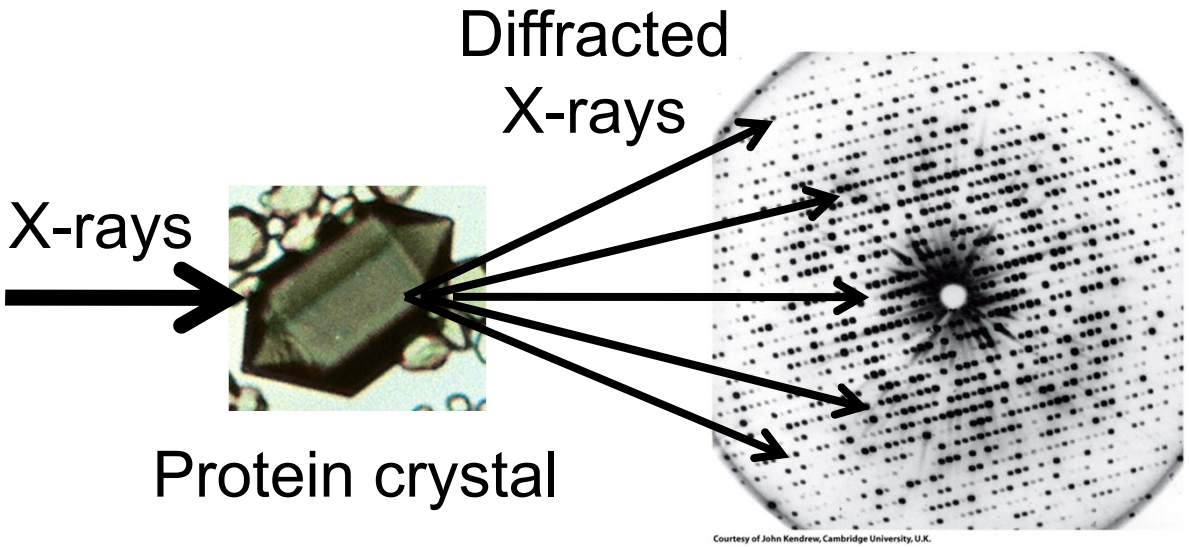
Cross-peaks indicate protons are within 5Å of each other

NMR-derived distance constraints are used to calculate likely protein conformations

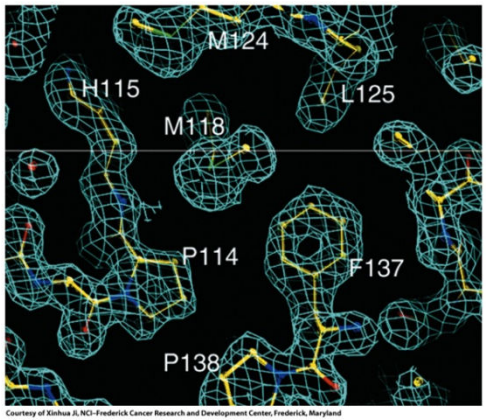


Courtesy of Stuart Schreiber, Harvard University

X-ray crystallography reveals the layout of repeating electron density

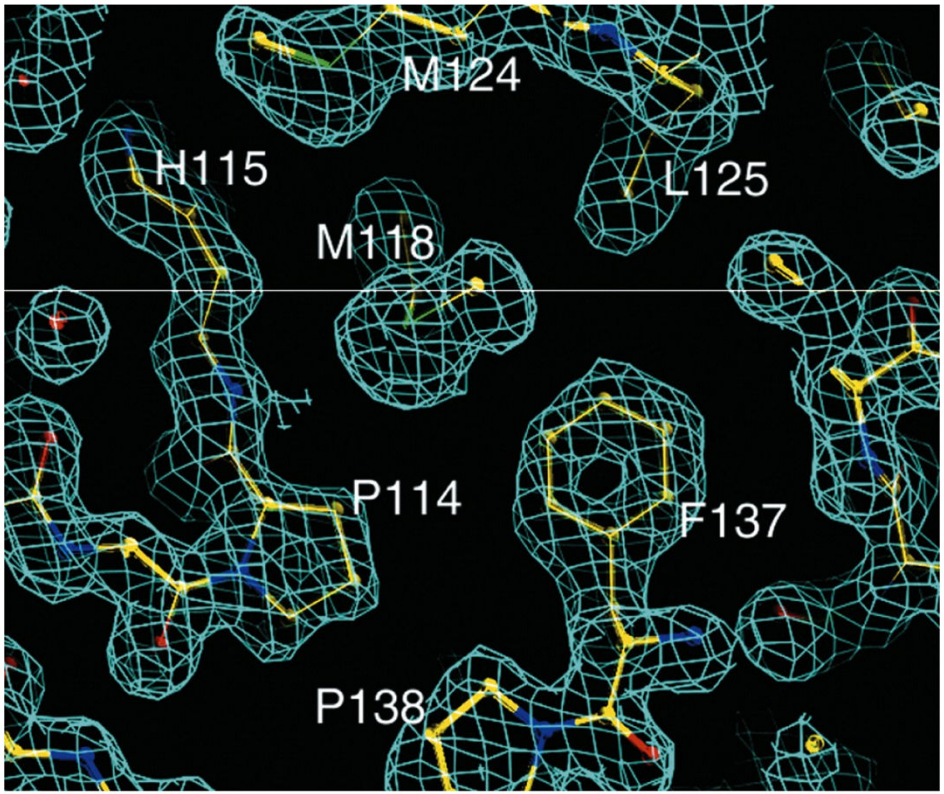


Data processing



Electron density map

Electron density map allows for positioning of protein atoms, revealing structure



Courtesy of Xinhua Ji, NCI–Frederick Cancer Research and Development Center, Frederick, Maryland