Chem 352 - Lecture 3
Part I: Amino Acids and Protein Primary Structure

Question for the Day: Approximately how much mass is required to make just one molecule for each of the possible polypeptides with a length of 100 amino acid residues, which are made from the 20 naturally occurring L-amino acids?

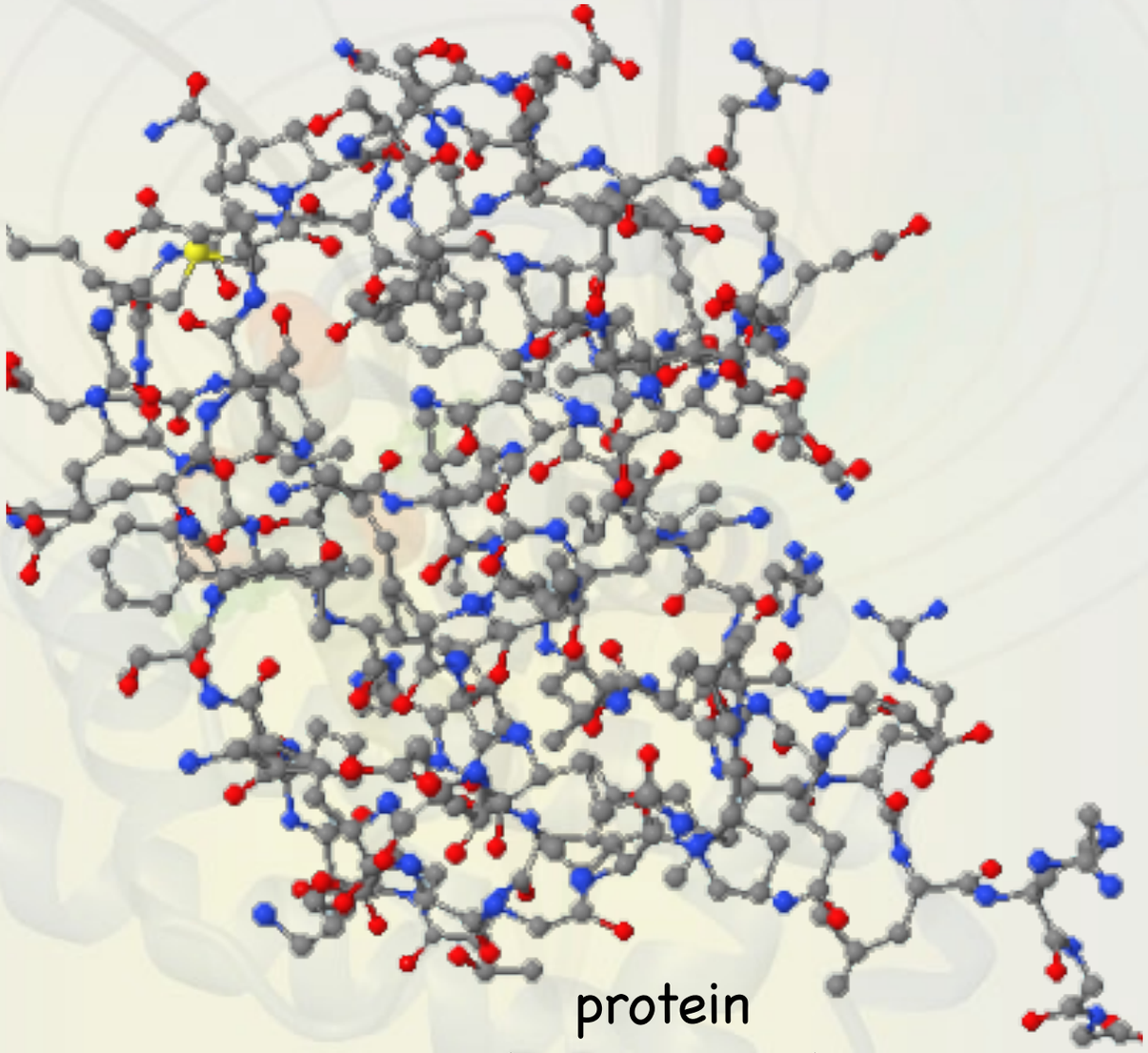
Introduction to Proteins

Proteins are the workhorses of a living cell.

- Biological catalysts (enzymes)
- Storage and transport
- Cytoskeleton
- Cellular regulation
- Hormones
- Antibodies
- ...

Introduction to Proteins

Proteins are polymers of amino acids (polypeptides) that often have a complicated 3-D structures.

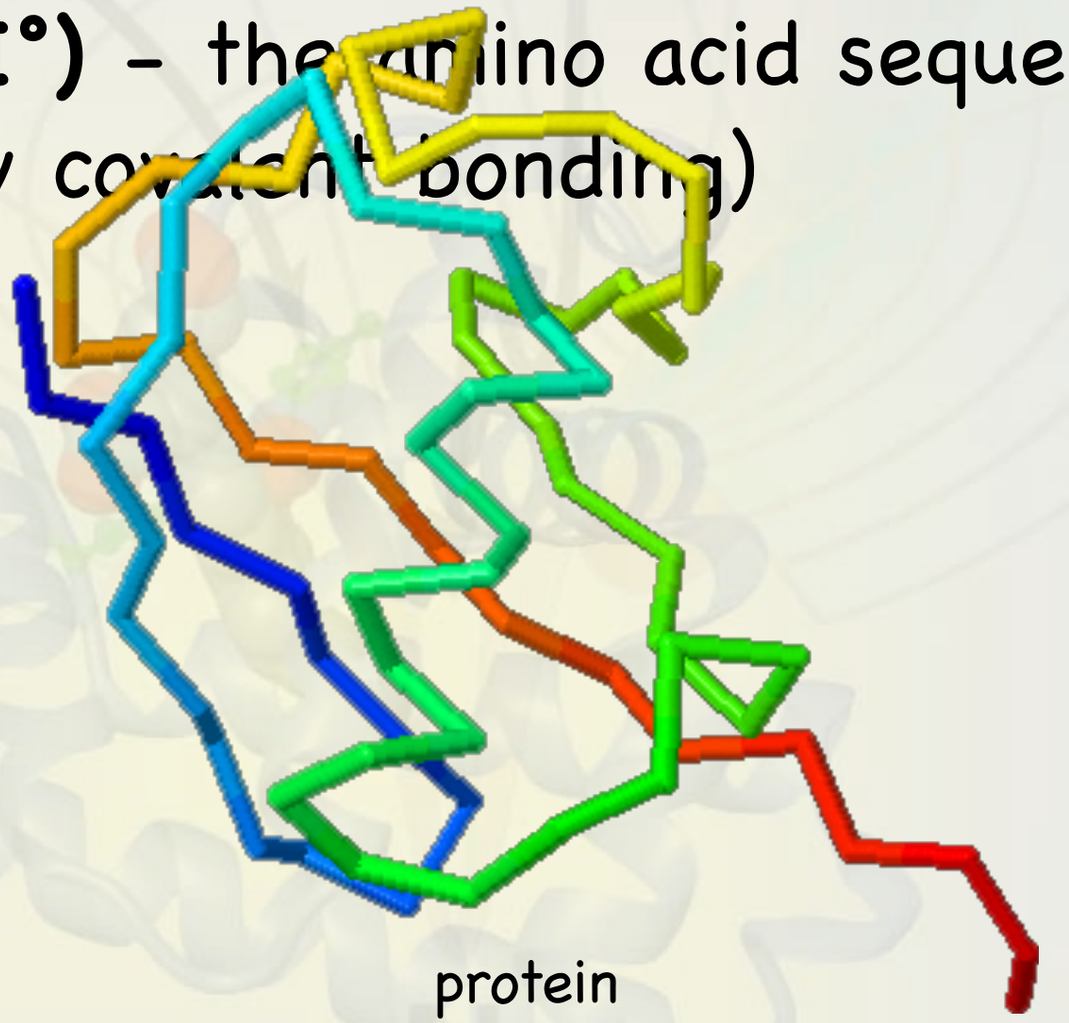


protein
(polypeptide)

Introduction to Proteins

Proteins can have up to four different levels of structure:

- **Primary (I°)** - the amino acid sequence (defined by covalent bonding)



protein
(polypeptide)

Introduction to Proteins

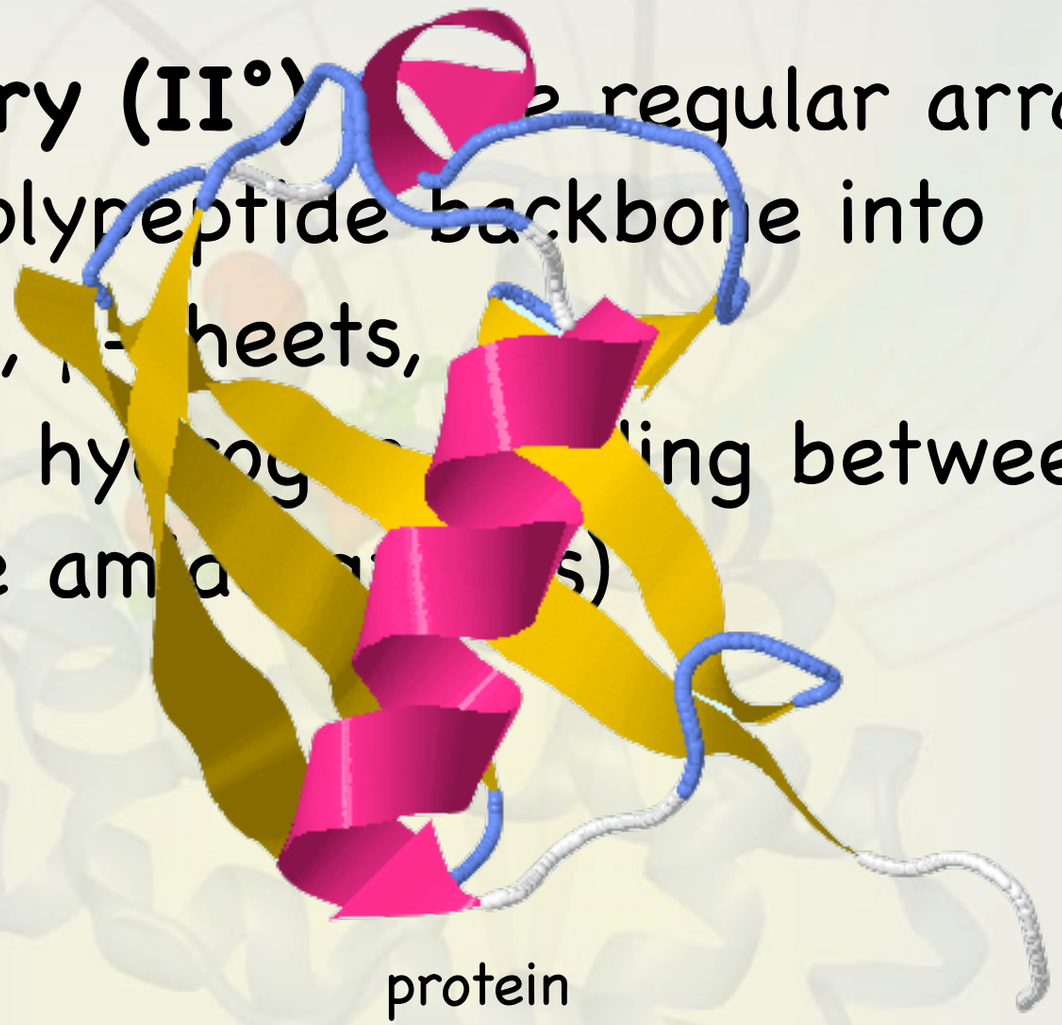
Proteins have different levels of structure:

- **Secondary (II°)** - the regular arrangements of the polypeptide backbone into α-helices, β-sheets, ... (features hydrogen bonding between the backbone amide groups)

Introduction to Proteins

Proteins have different levels of structure:

- **Secondary (II°)** – regular arrangements of the polypeptide backbone into α-helices, β-sheets, (features hydrogen bonding between the backbone amides)



protein
(polypeptide)

Introduction to Proteins

Proteins have different levels of structure:

- **Secondary (II°)** - the regular arrangements of the polypeptide backbone into α-helices, β-sheets, ... (features hydrogen bonding between the backbone amide groups)

Introduction to Proteins

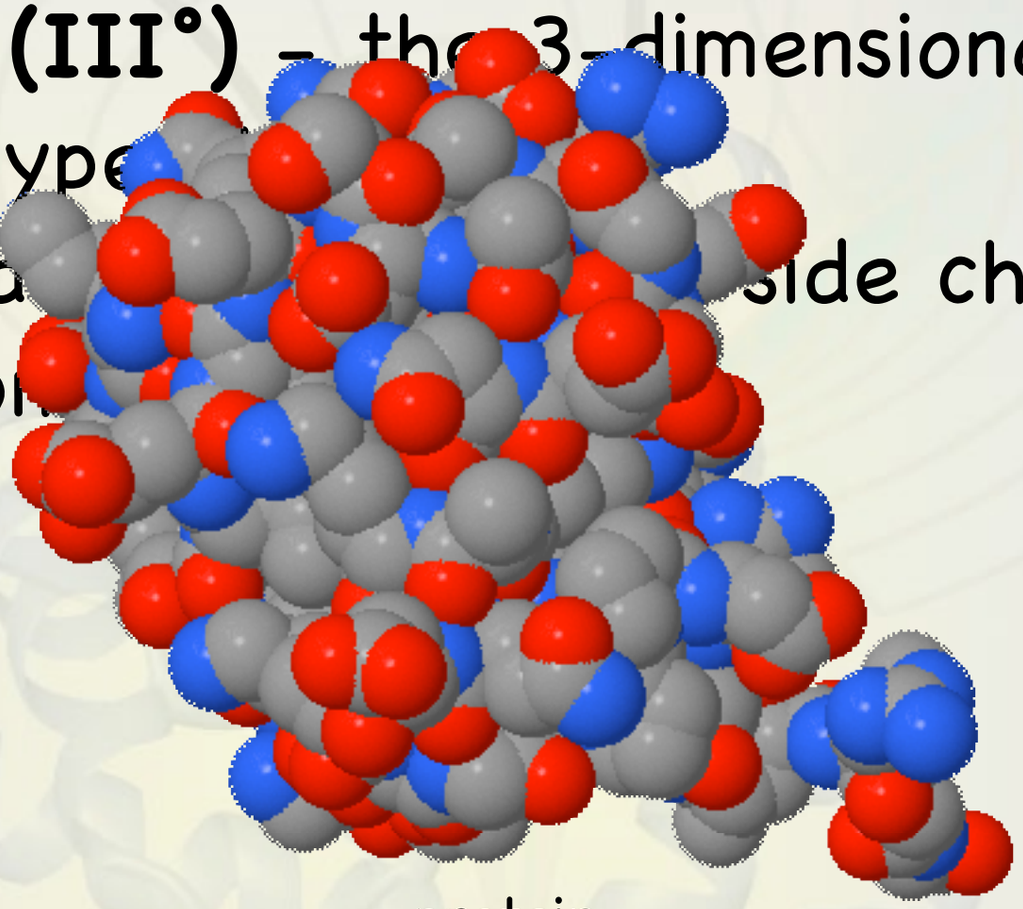
Proteins have different levels of structure:

- **Tertiary (III°)** - the 3-dimensional fold of a single polypeptide
  (stabilized by non-covalent side chain interactions)

Introduction to Proteins

Proteins have different levels of structure:

- **Tertiary (III°)** – the 3-dimensional fold of a single polype
  (stabilized side chain
  interaction



protein
(polypeptide)

Introduction to Proteins

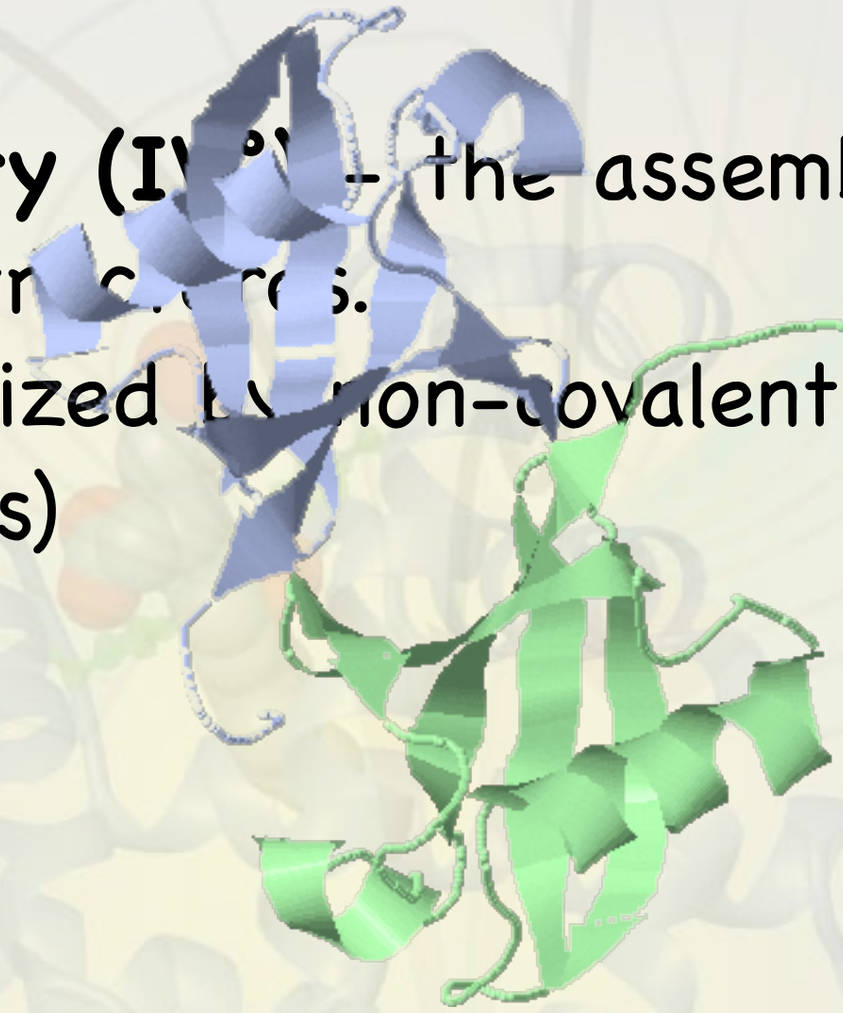
Proteins have different levels of structure:

- **Tertiary (III°)** - the 3-dimensional fold of a single polypeptide
  (stabilized by non-covalent side chain interactions)

Introduction to Proteins

Proteins have different levels of structure:

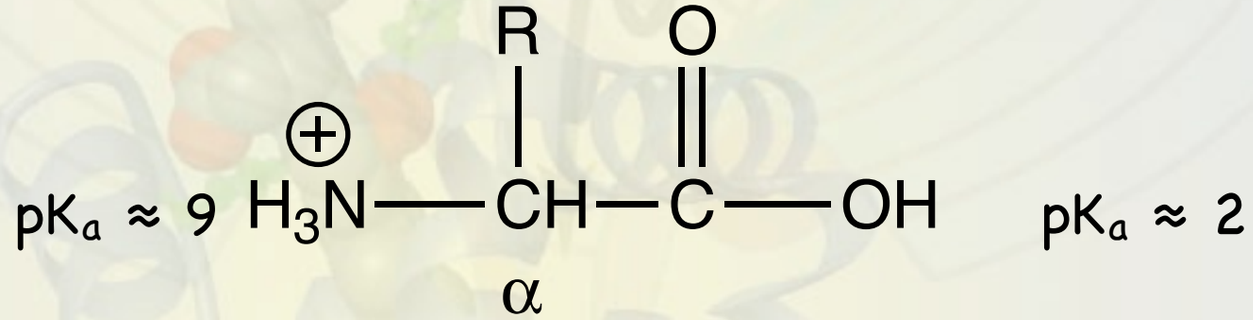
- **Quaternary (IV°)** – the assembly of multiple tertiary structures.
  (also stabilized by non-covalent side chain interactions)



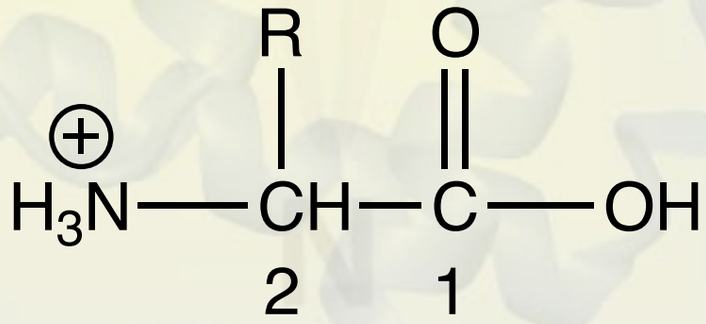
protein with two subunits
(2 polypeptides)

The Amino Acids

- Proteins are made from polymers of α-amino acids.
  - These polymers are called **polypeptides**.
  - The **monomers** are the α-amino acids.



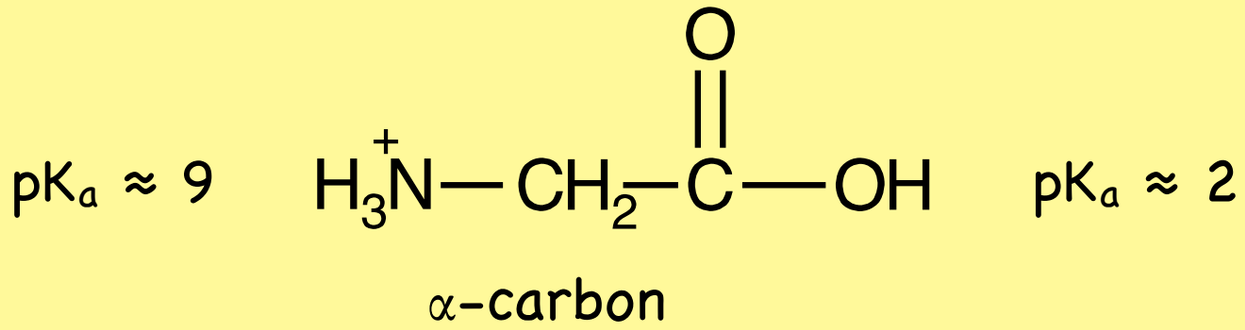
The monomer



The Amino Acids

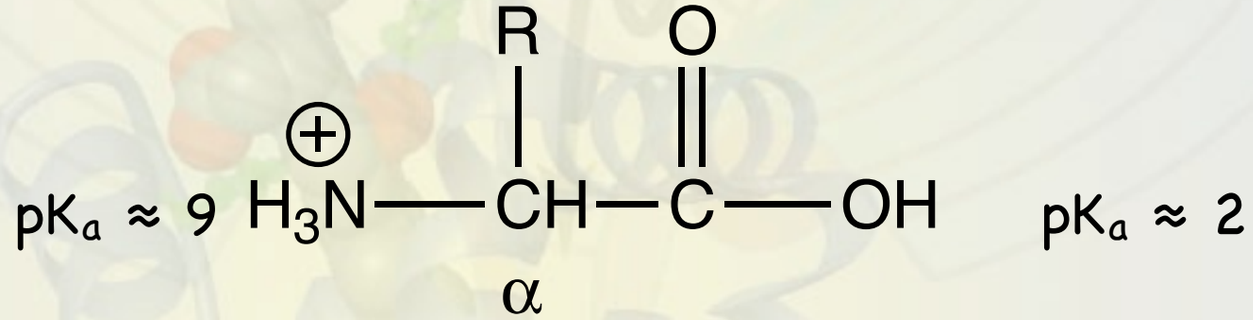
Problem:

Draw the titration curve (pH vs. equivalents) for glycine between pH 0 and pH 12.

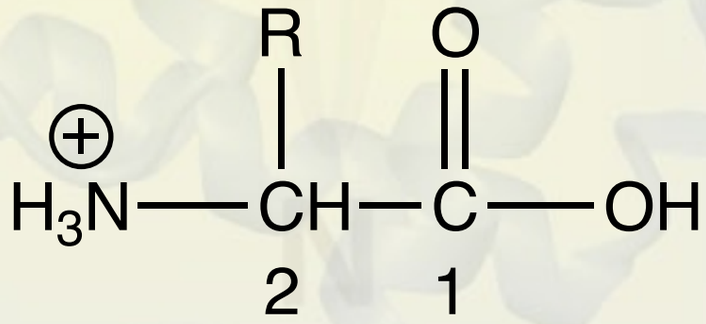


The Amino Acids

- Proteins are made from polymers of α-amino acids.
  - These polymers are called **polypeptides**.
  - The **monomers** are the α-amino acids.

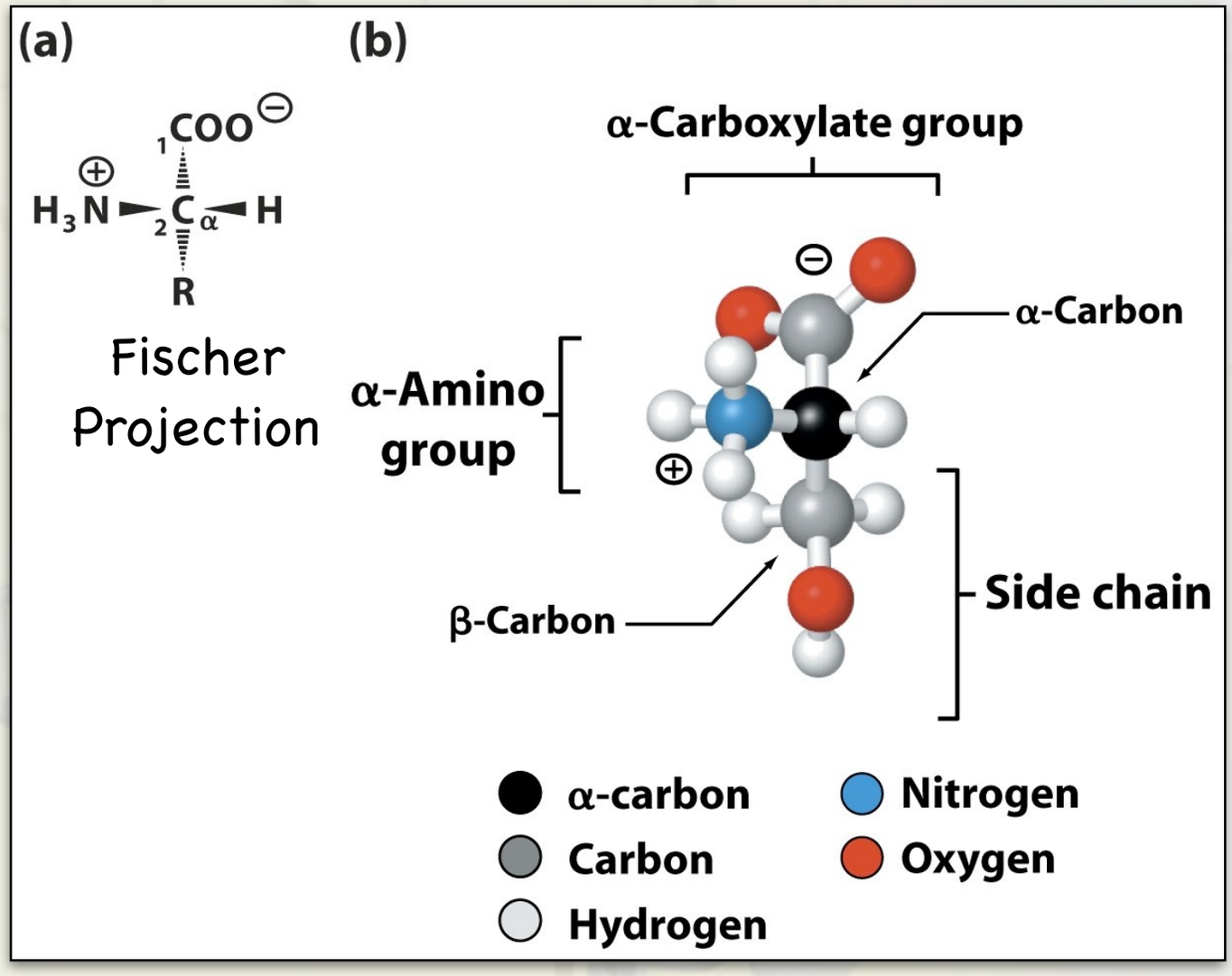


The monomer



The Amino Acids

For 19 out of the 20 common amino acids, the α-carbon is **chiral**.



The Amino Acids

F

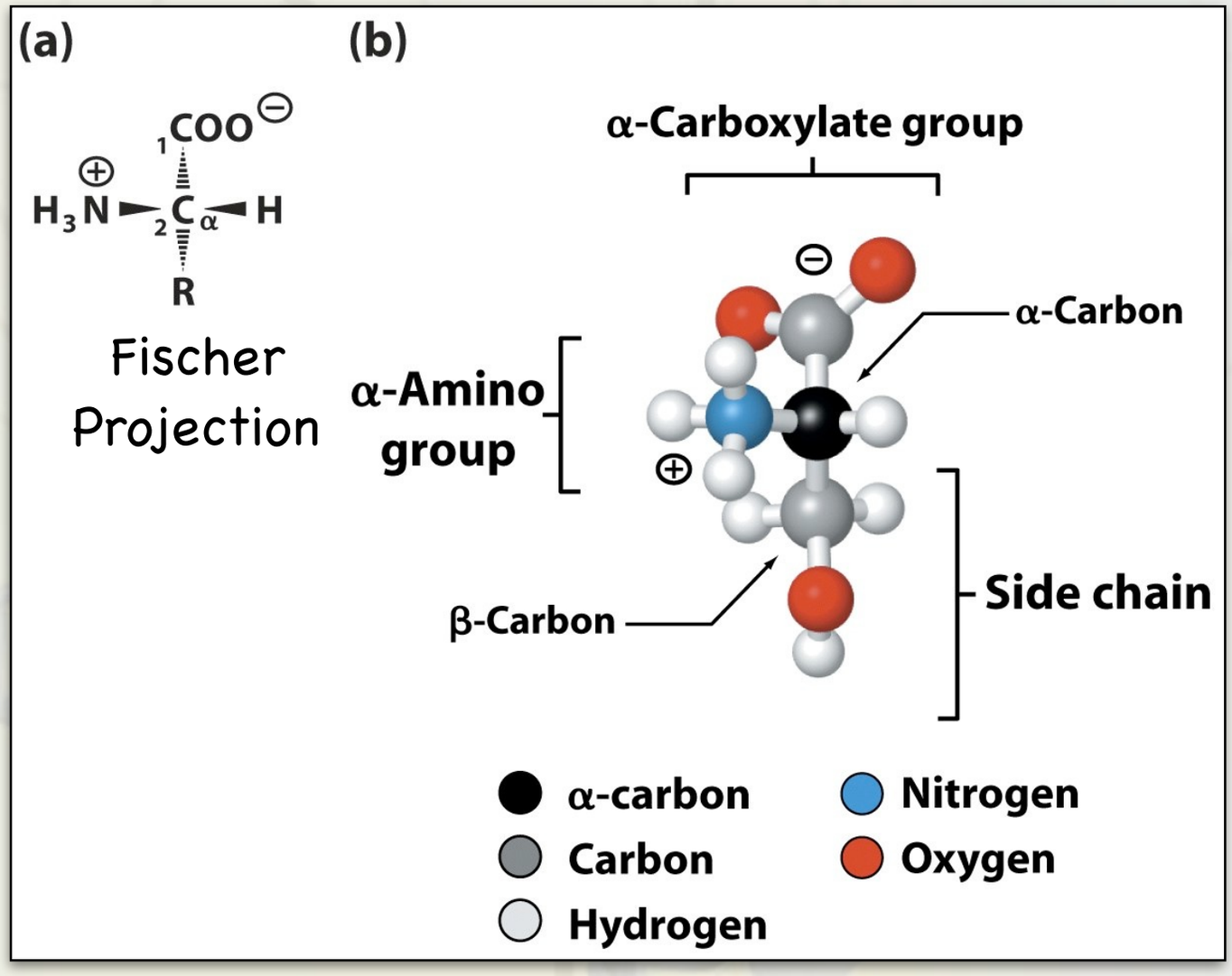
C

Question:

What does it mean to be **chiral**?

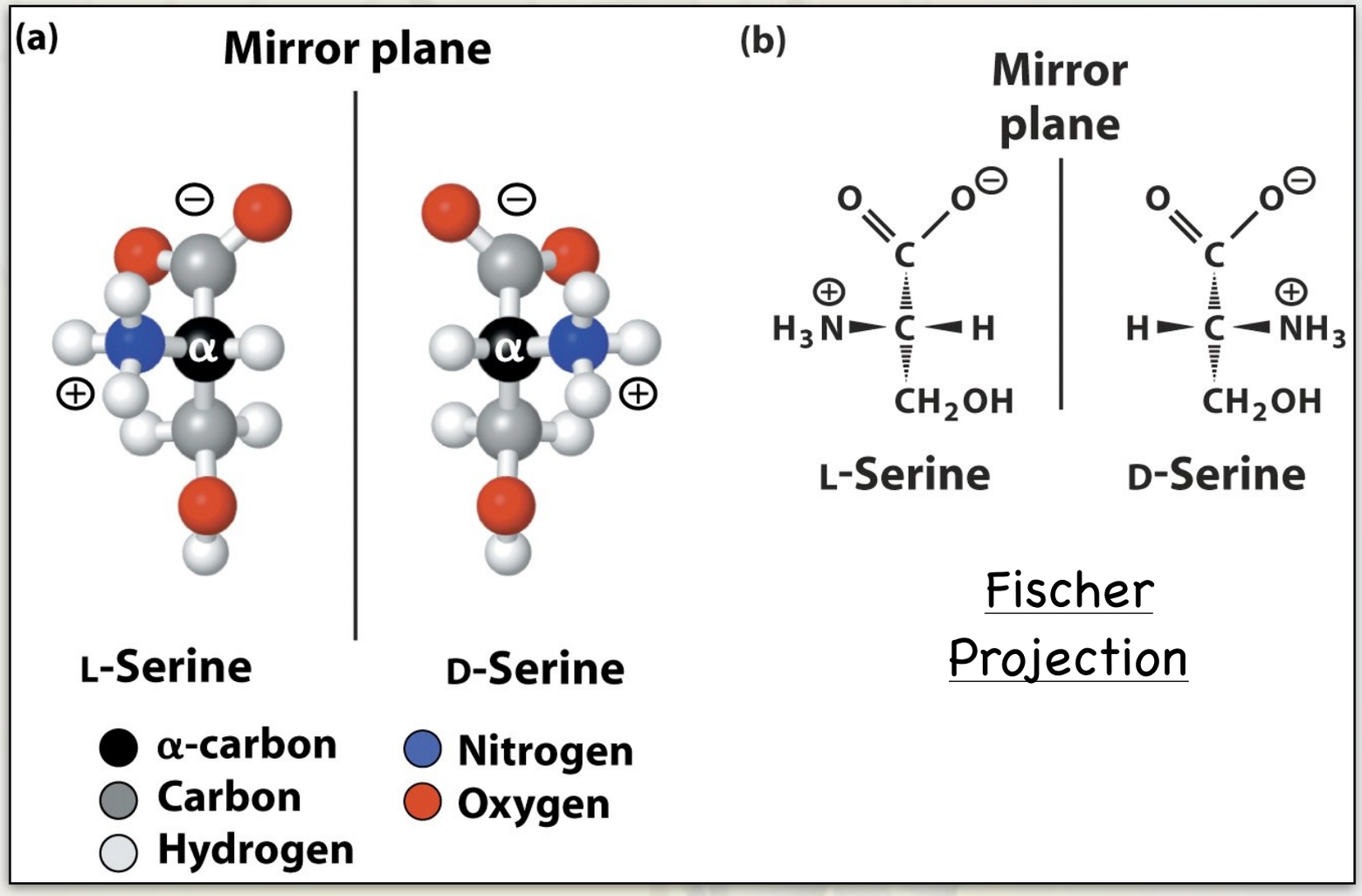
The Amino Acids

For 19 out of the 20 common amino acids, the α-carbon is **chiral**.



The Amino Acids

For 19 out of the 20 common amino acids, the α-carbon is **chiral**.



The Amino Acids

F

C

Question:

Which of the 20 amino acids is not **chiral**?

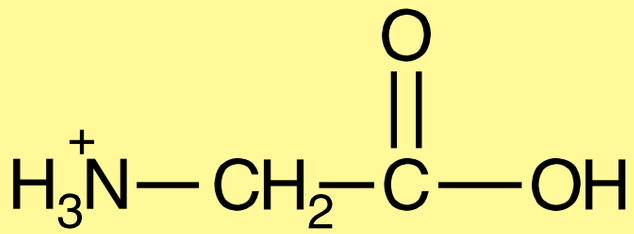
The Amino Acids

F

C

Question:

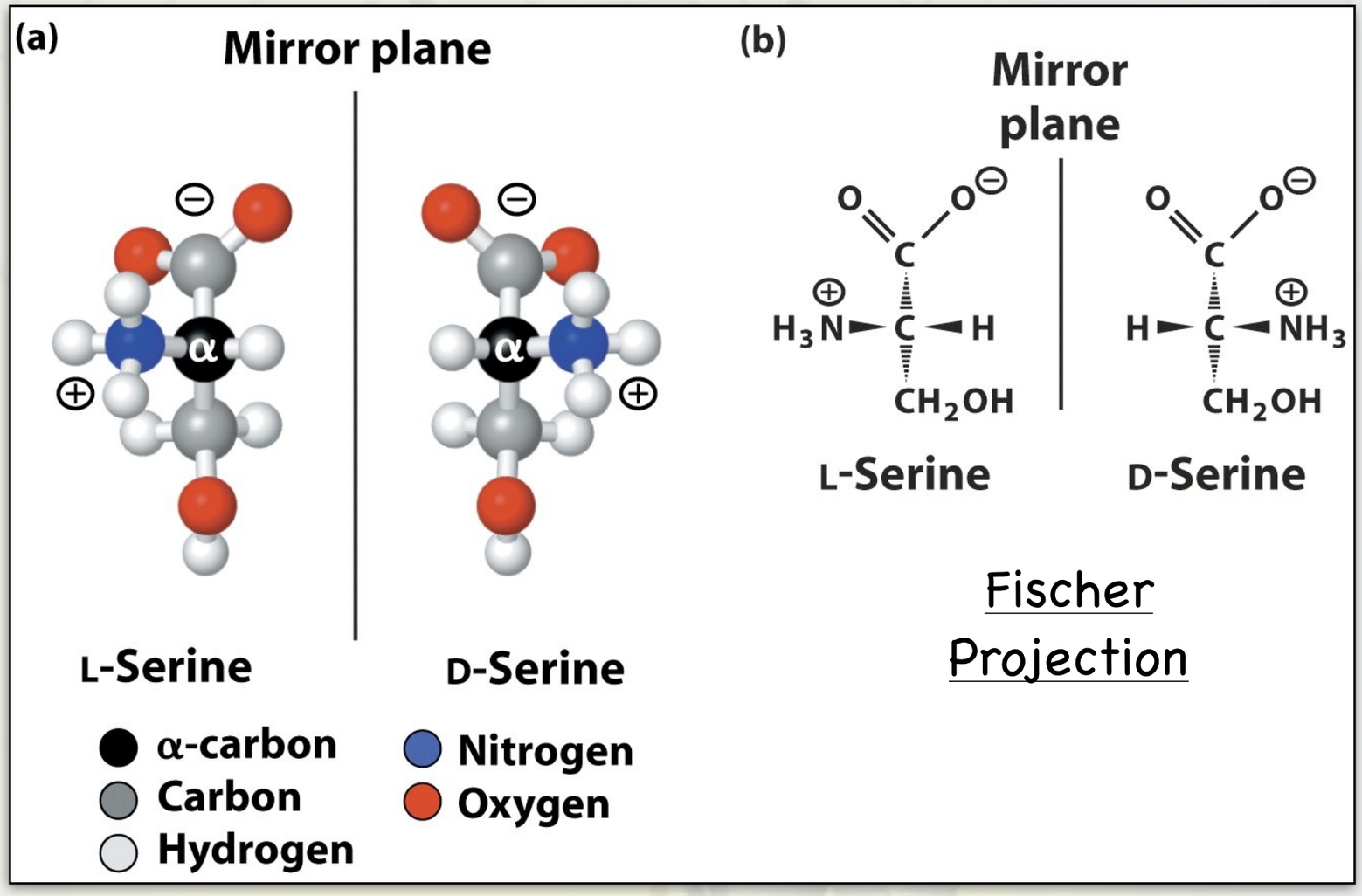
Which of the 20 amino acids is not **chiral**?



Glycine

The Amino Acids

For 19 out of the 20 common amino acids, the α-carbon is **chiral**.



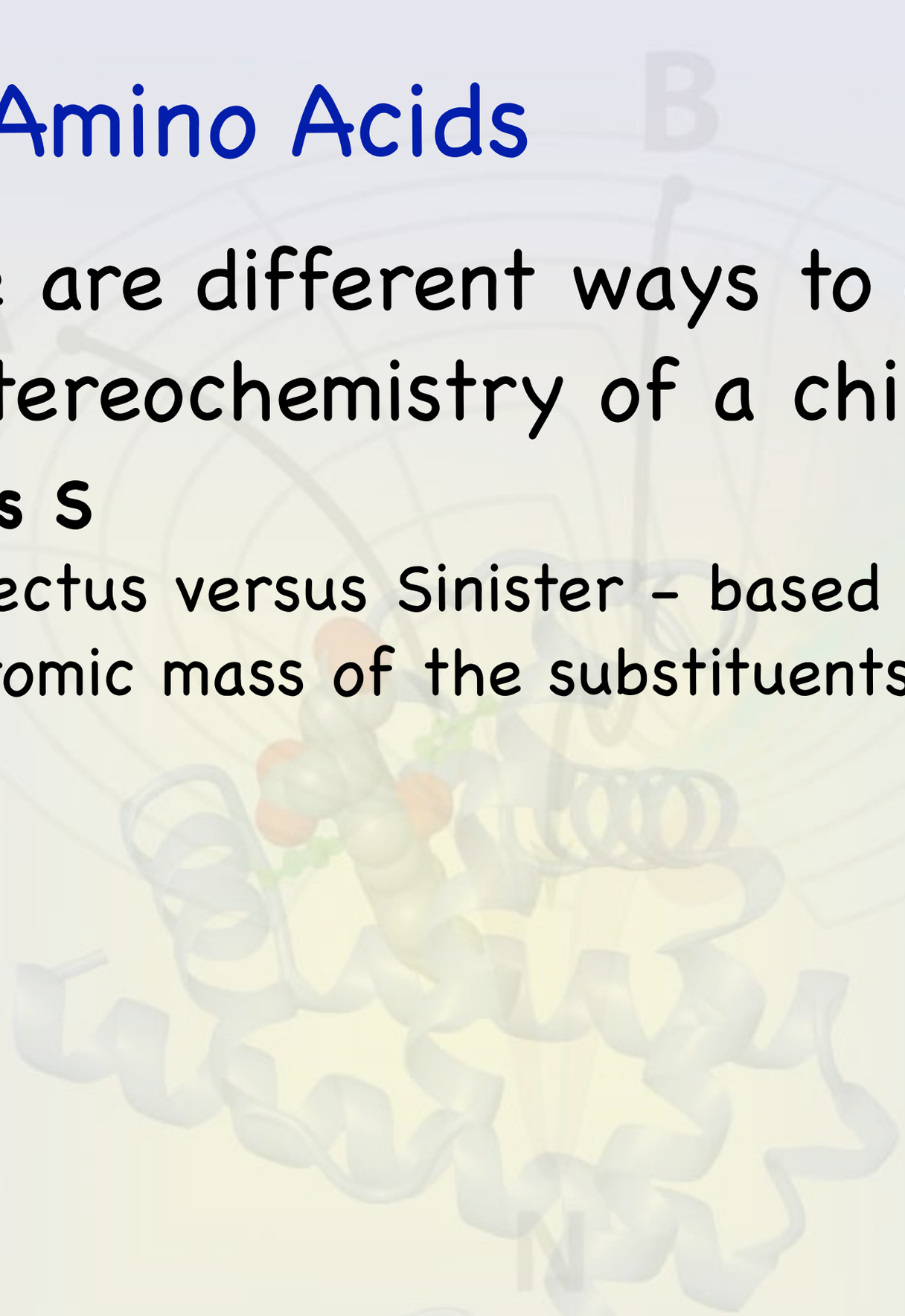
The Amino Acids

There are different ways to designate the stereochemistry of a chiral center:

The Amino Acids

There are different ways to designate the stereochemistry of a chiral center:

- **R vs S**
  - Rectus versus Sinister - based on the atomic mass of the substituents



The Amino Acids

There are different ways to designate the stereochemistry of a chiral center:

- **R vs S**
  - Rectus versus Sinister - based on the atomic mass of the substituents
- **d (+) vs l (-)**
  - dextrorotatory versus levorotatory - based on the bending of plane polarized light

The Amino Acids

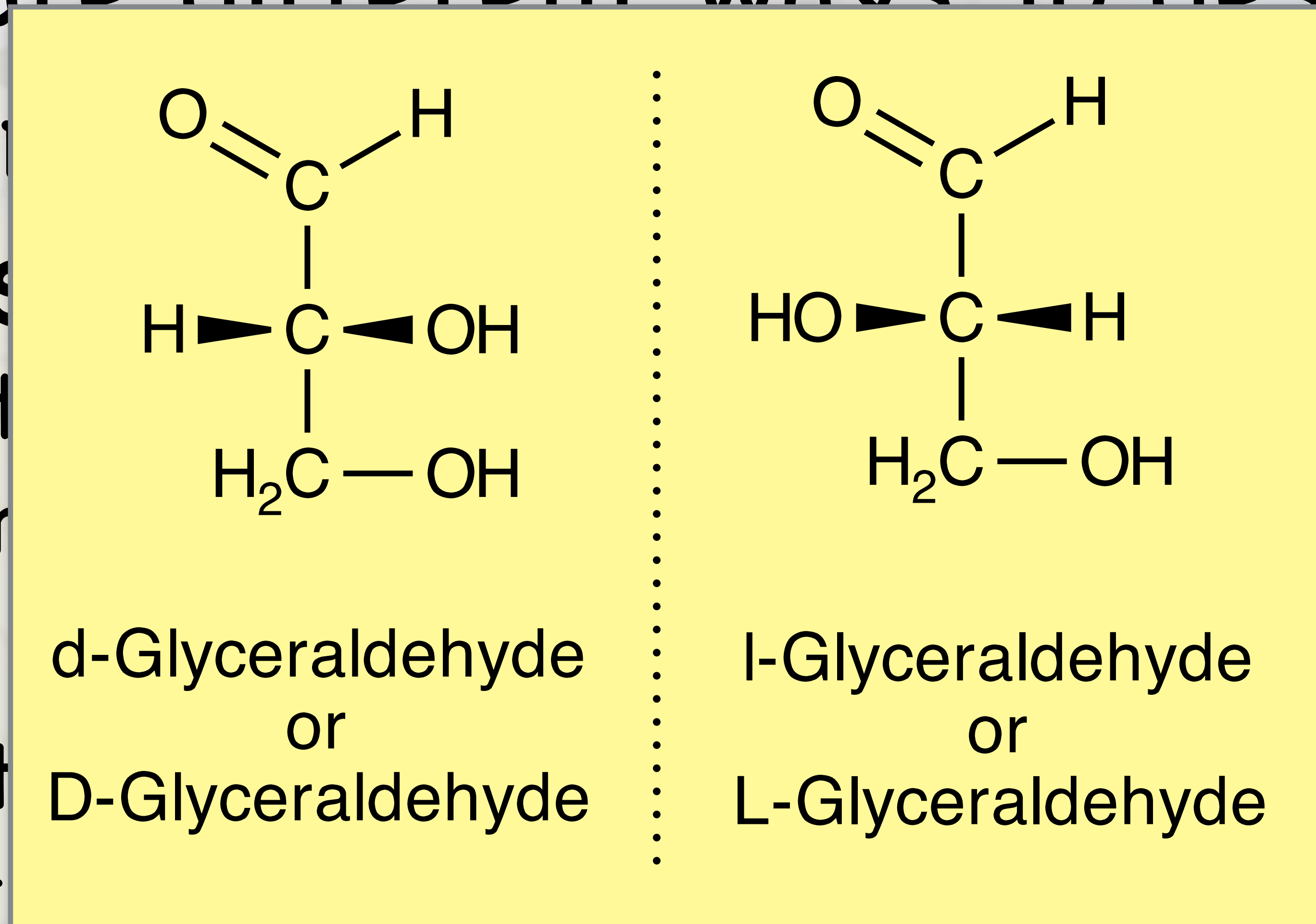
There are different ways to designate the stereochemistry of a chiral center:

- **R vs S**
  - Rectus versus Sinister - based on the atomic mass of the substituents
- **d (+) vs l (-)**
  - dextrorotatory versus levorotatory - based on the bending of plane polarized light
- **D vs L**
  - Based on how glyceraldehyde bends plane polarized light

The Amino Acids

There are different ways to designate the ste[...] center:

- **R vs S**
  - Rect[...]he atom[...]
- **d (+)**
  - dext[...] based on t[...]ght



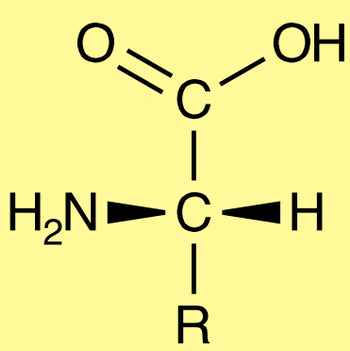
- **D vs L**
  - Based on how glyceraldehyde bends plane polarized light

The Amino Acids

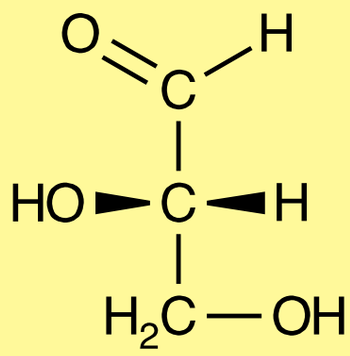
There are different ways to designate the stereochemistry of a chiral center:

- **R vs S**
  - Rectus versus Sinister - based on the atomic mass of the substituents
- **d (+) vs l (-)**
  - dextrorotatory versus levorotatory - based on the bending of plane polarized light
- **D vs L**
  - Based on how glyceraldehyde bends plane polarized light

The Amino Acids



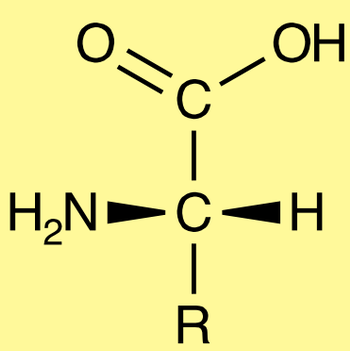
L-Amino acid
some are l-
some are d-



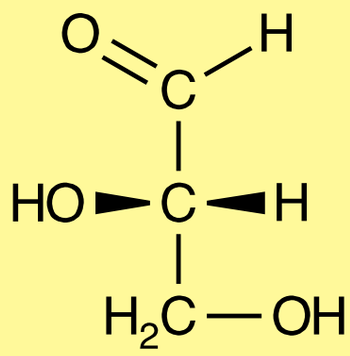
L-Glyceraldehyde
or
l-Glyceraldehyde

The Amino Acids

All of the amino acids used to to make proteins are L-amino acids



L-Amino acid
some are l-
some are d-



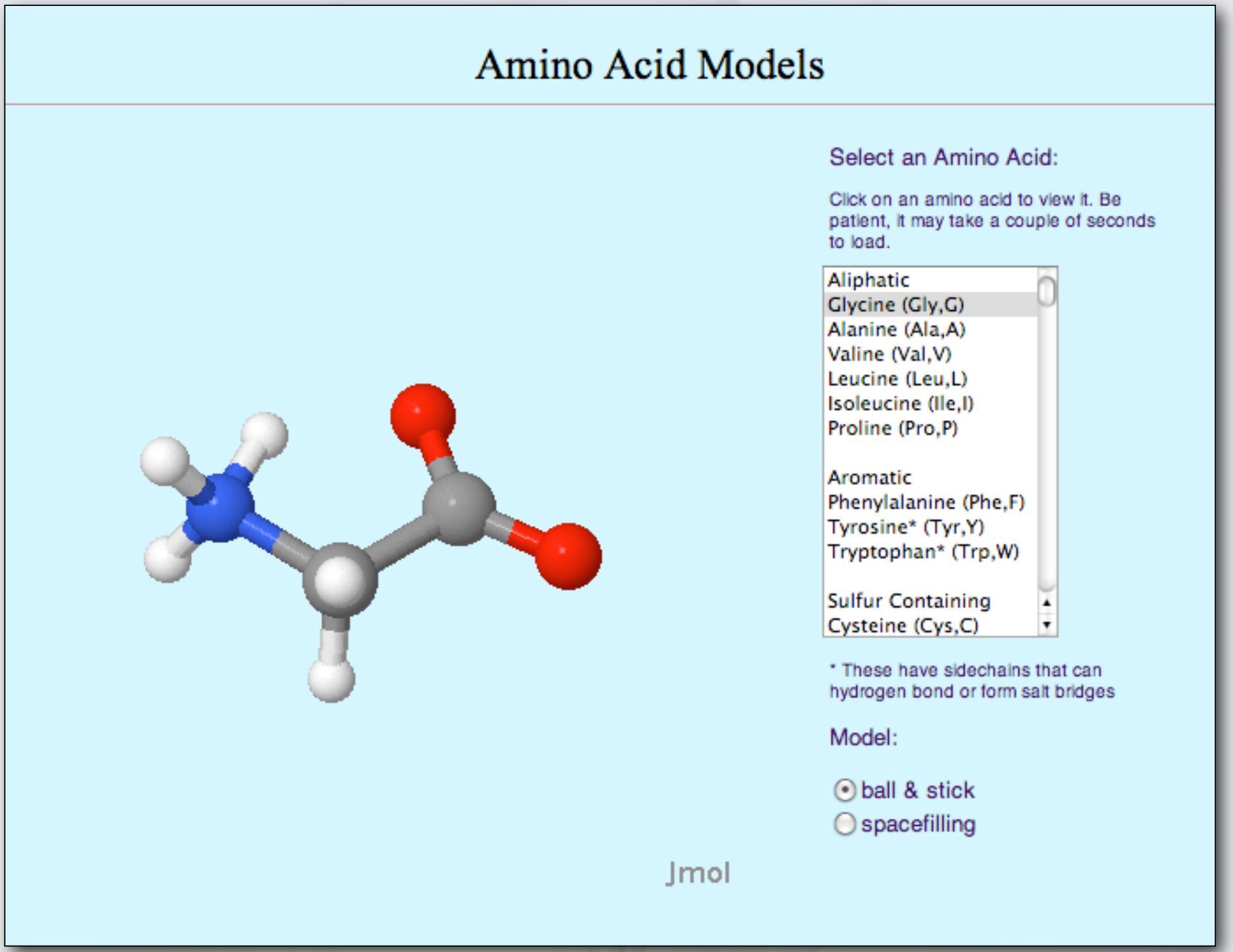
L-Glyceraldehyde
or
l-Glyceraldehyde

The Amino Acids

There is considerable variety in the chemical and physical properties of the the 20 different amino acid side chains.

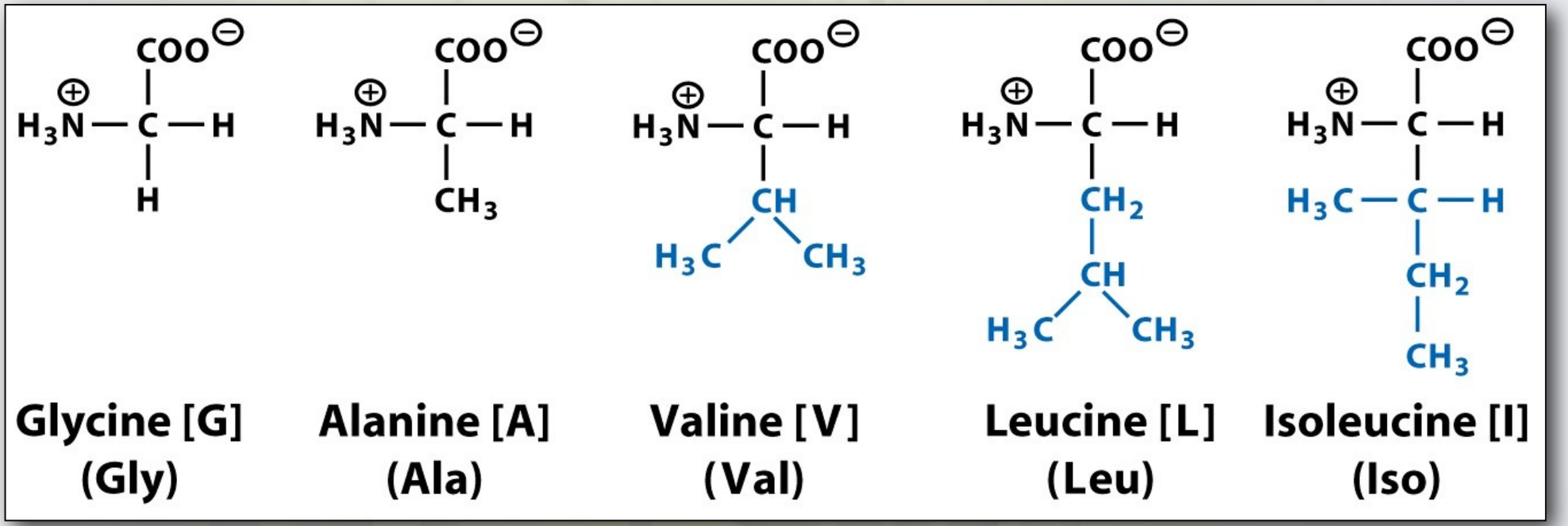
- Aliphatic (saturated hydrocarbon) (G,A,V,L,I,P)
- Aromatic (F,Y,W)
- Sulfur-containing (C,M)
- Alcohols (S,T)
- Bases (K,R,H)
- Acids (D,E)
- Amides (N,Q)

The Amino Acids



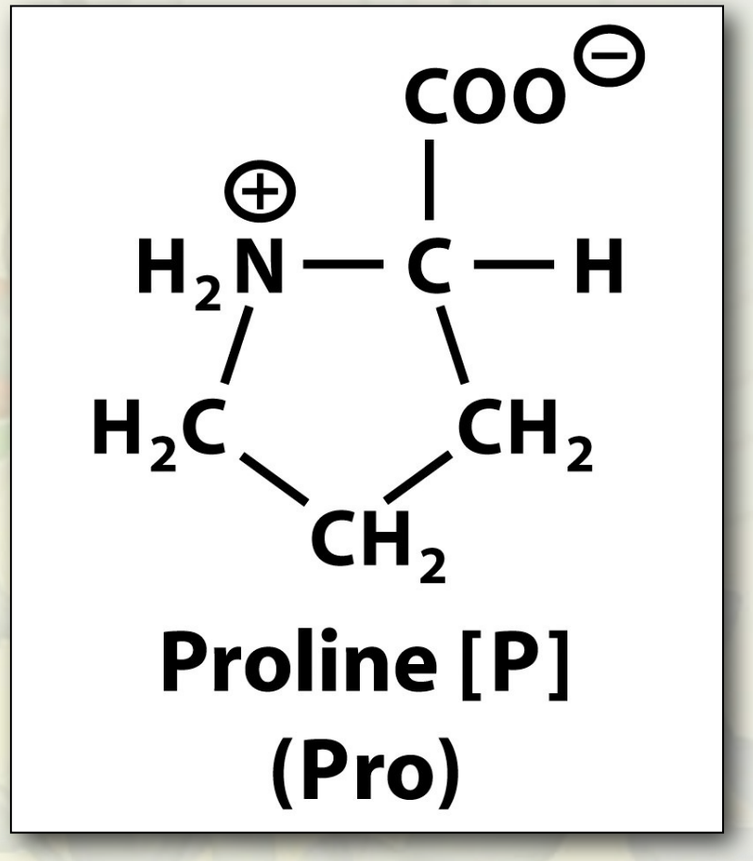
The Amino Acids

- Aliphatic (saturated hydrocarbon) (G,A,V,L,I,P)



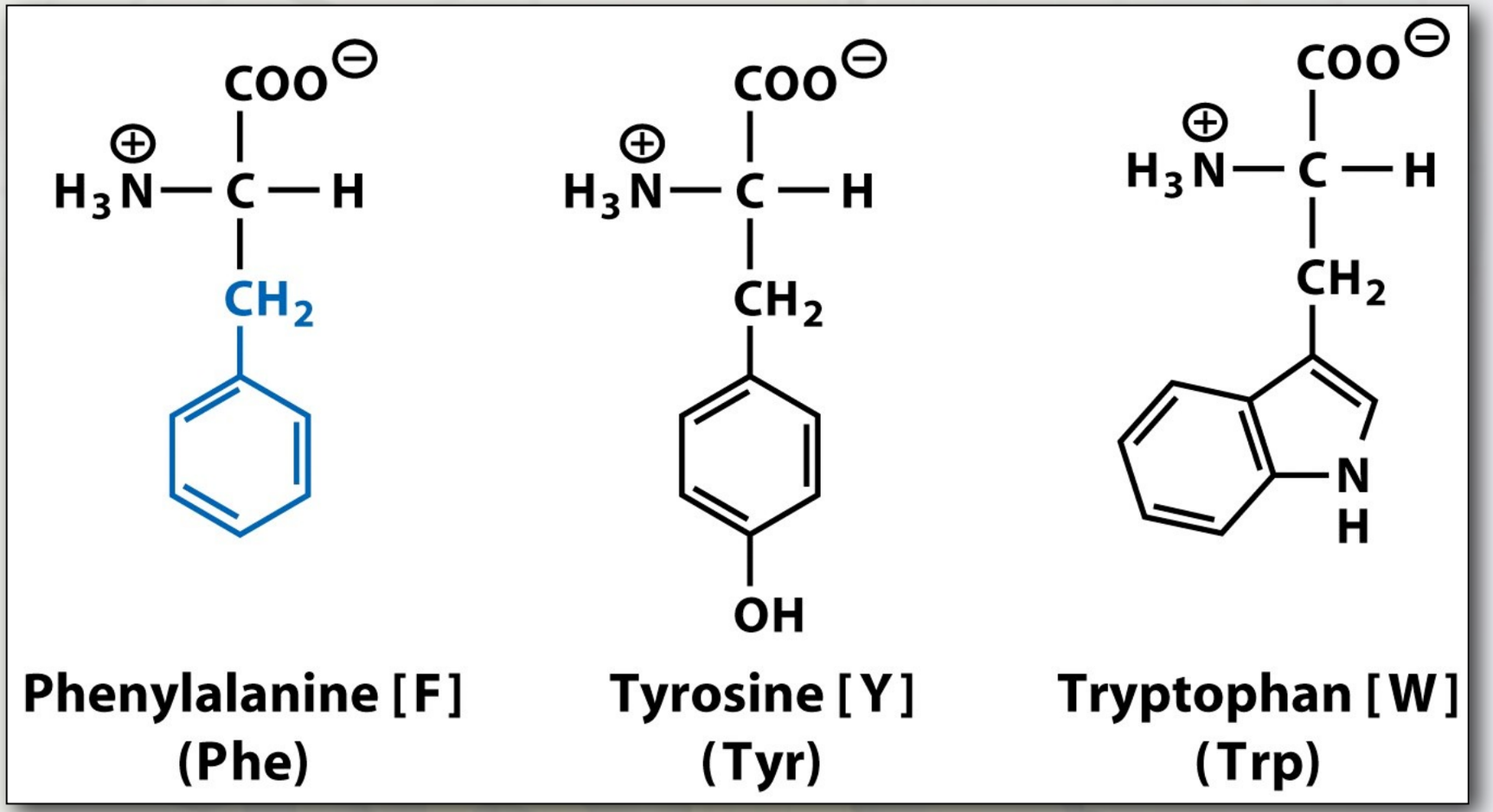
The Amino Acids

- Aliphatic (saturated hydrocarbon) (G,A,V,L,I,P)



The Amino Acids

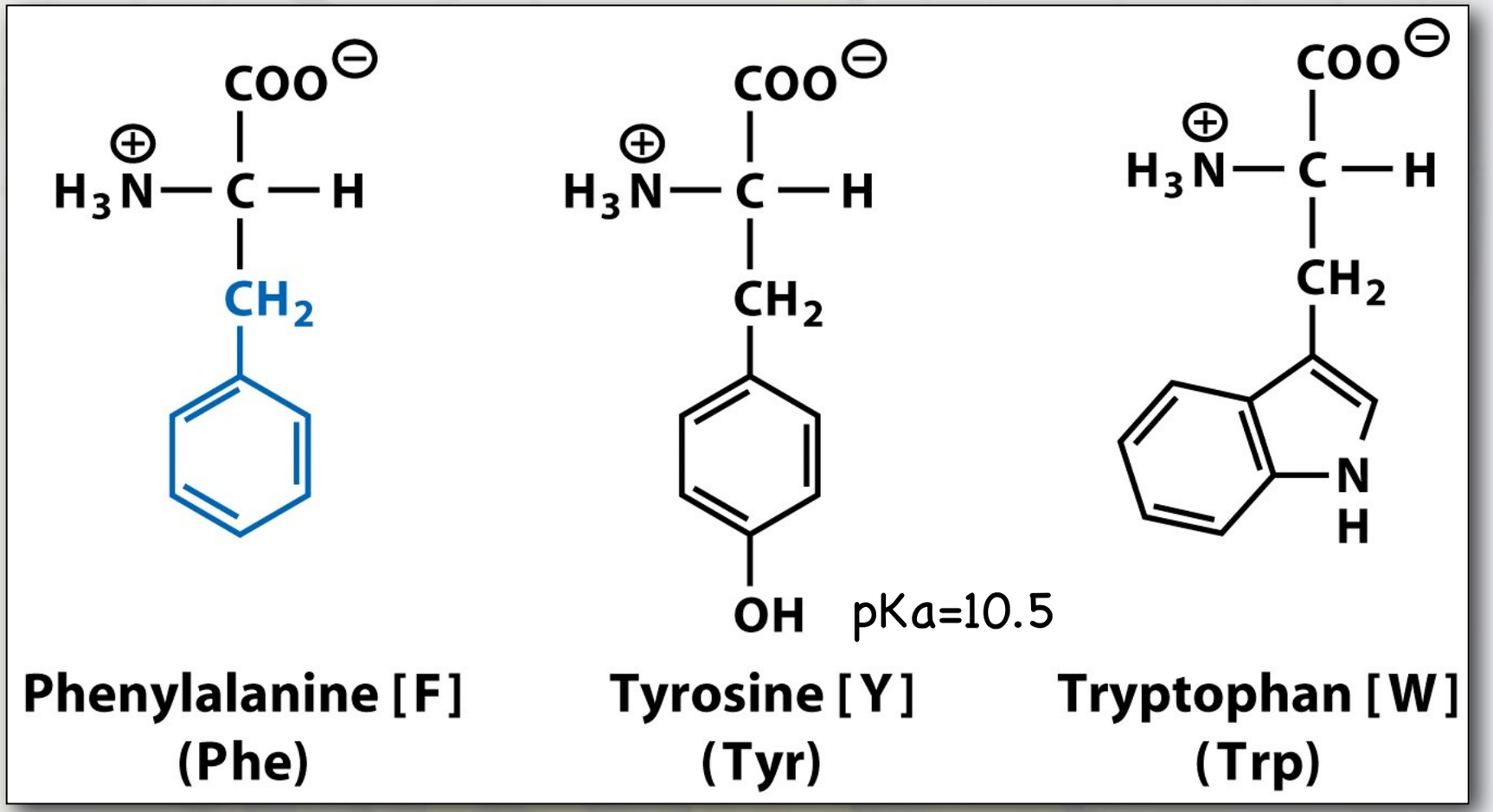
- Aromatic (F,Y,W)



Phenylalanine [F] (Phe) | **Tyrosine [Y] (Tyr)** | **Tryptophan [W] (Trp)**

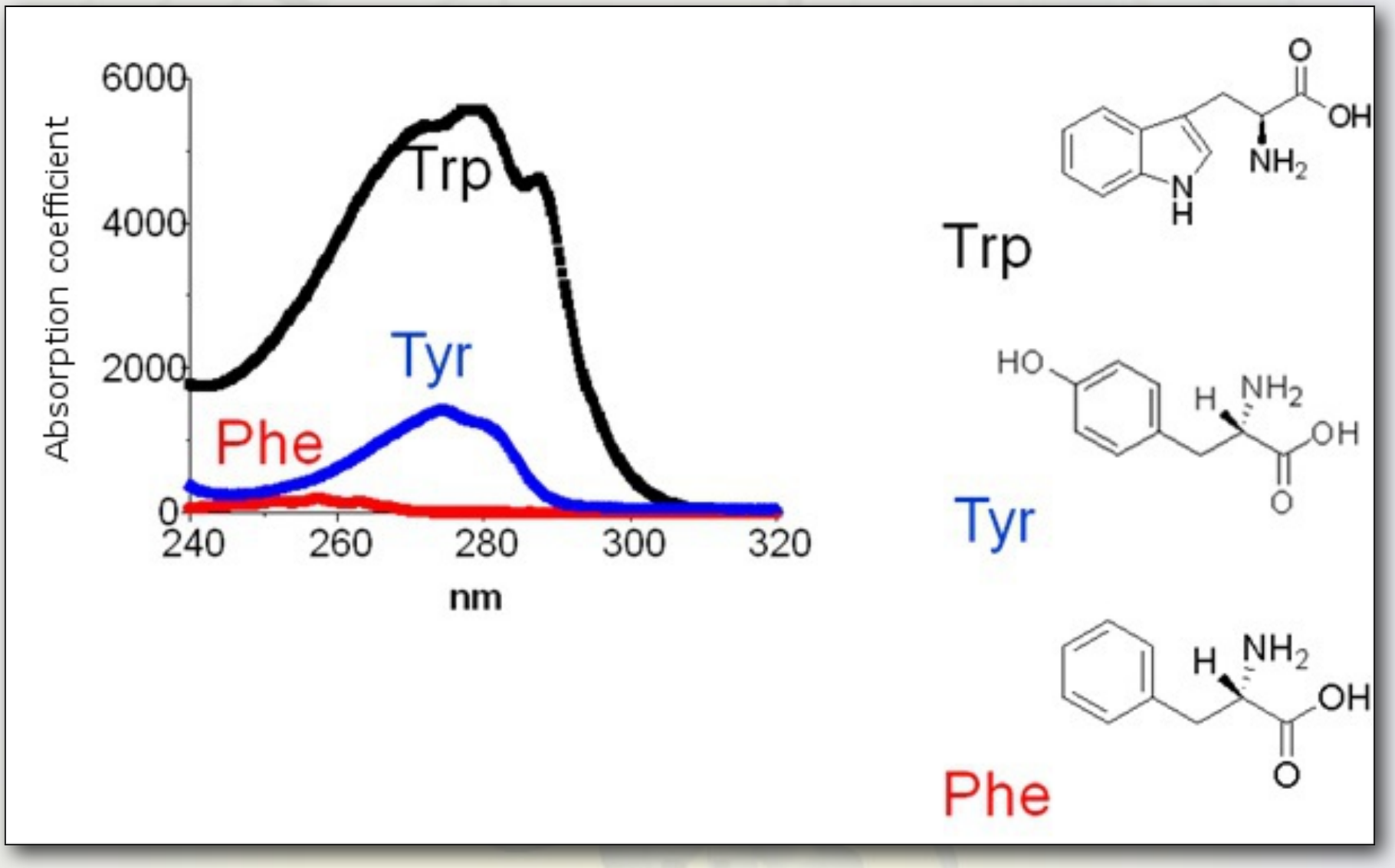
The Amino Acids

- Aromatic (F,Y,W)



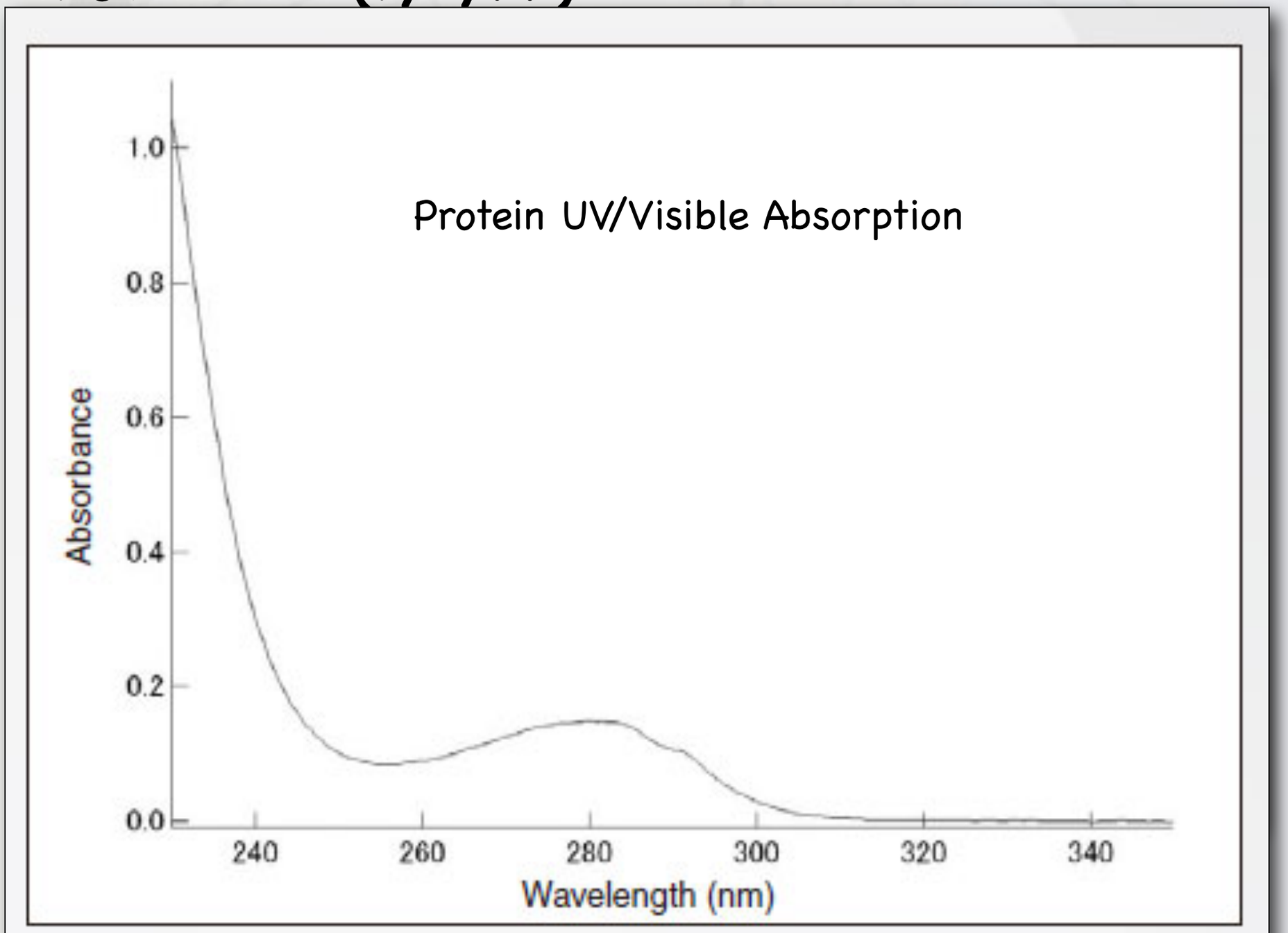
The Amino Acids

- Aromatic (F,Y,W)



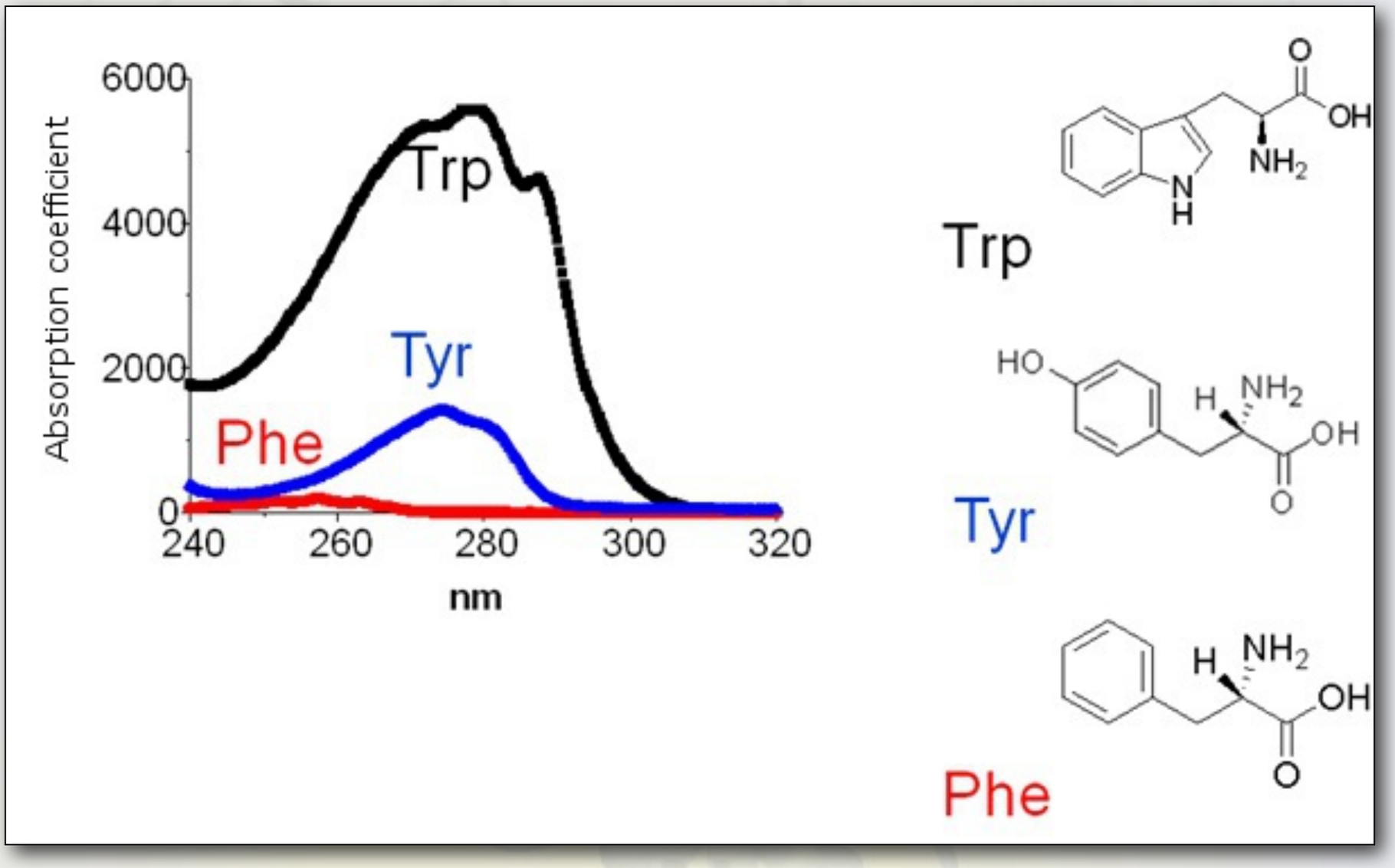
The Amino Acids

- Aromatic (F,Y,W)



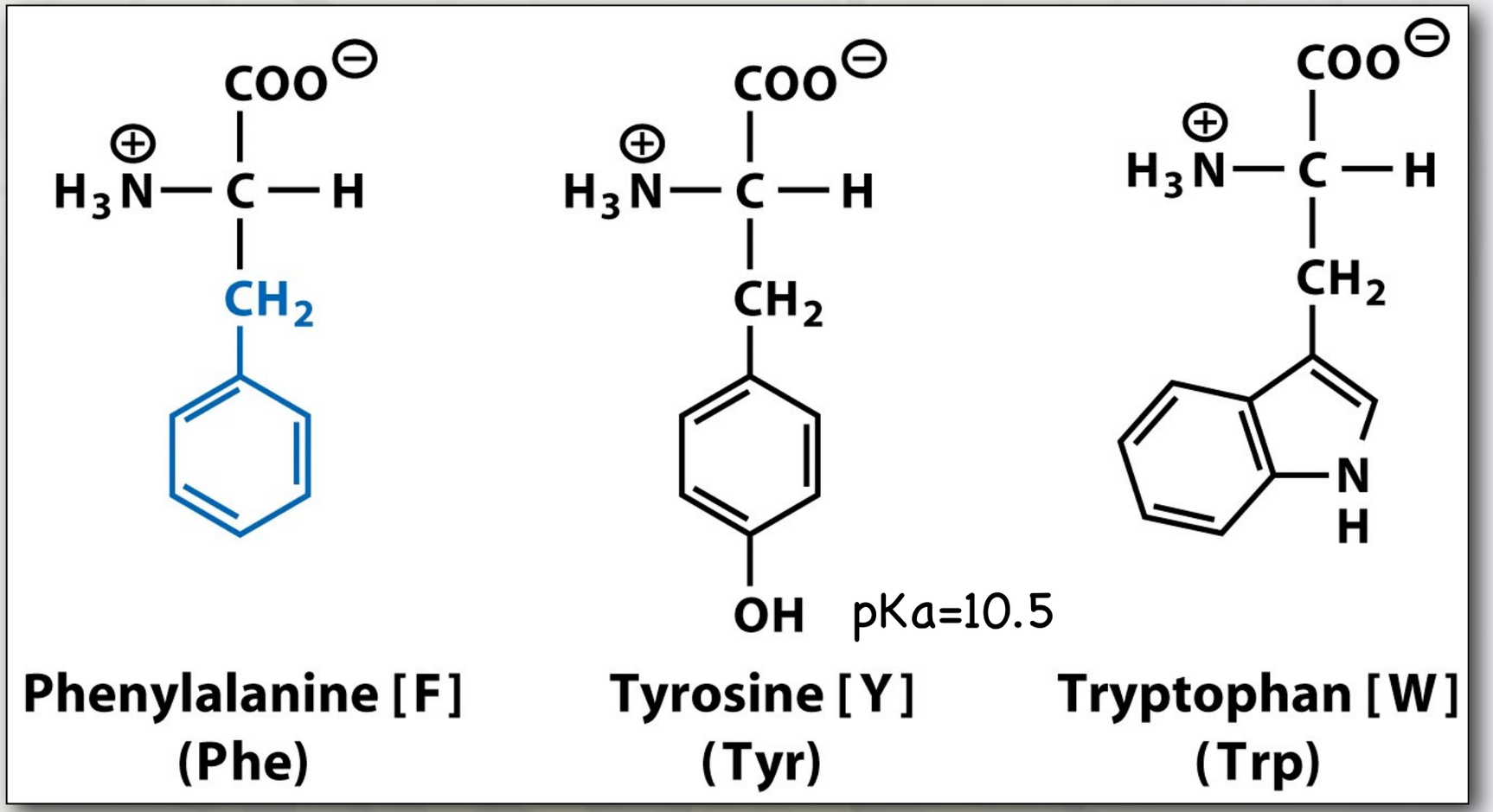
The Amino Acids

- Aromatic (F,Y,W)



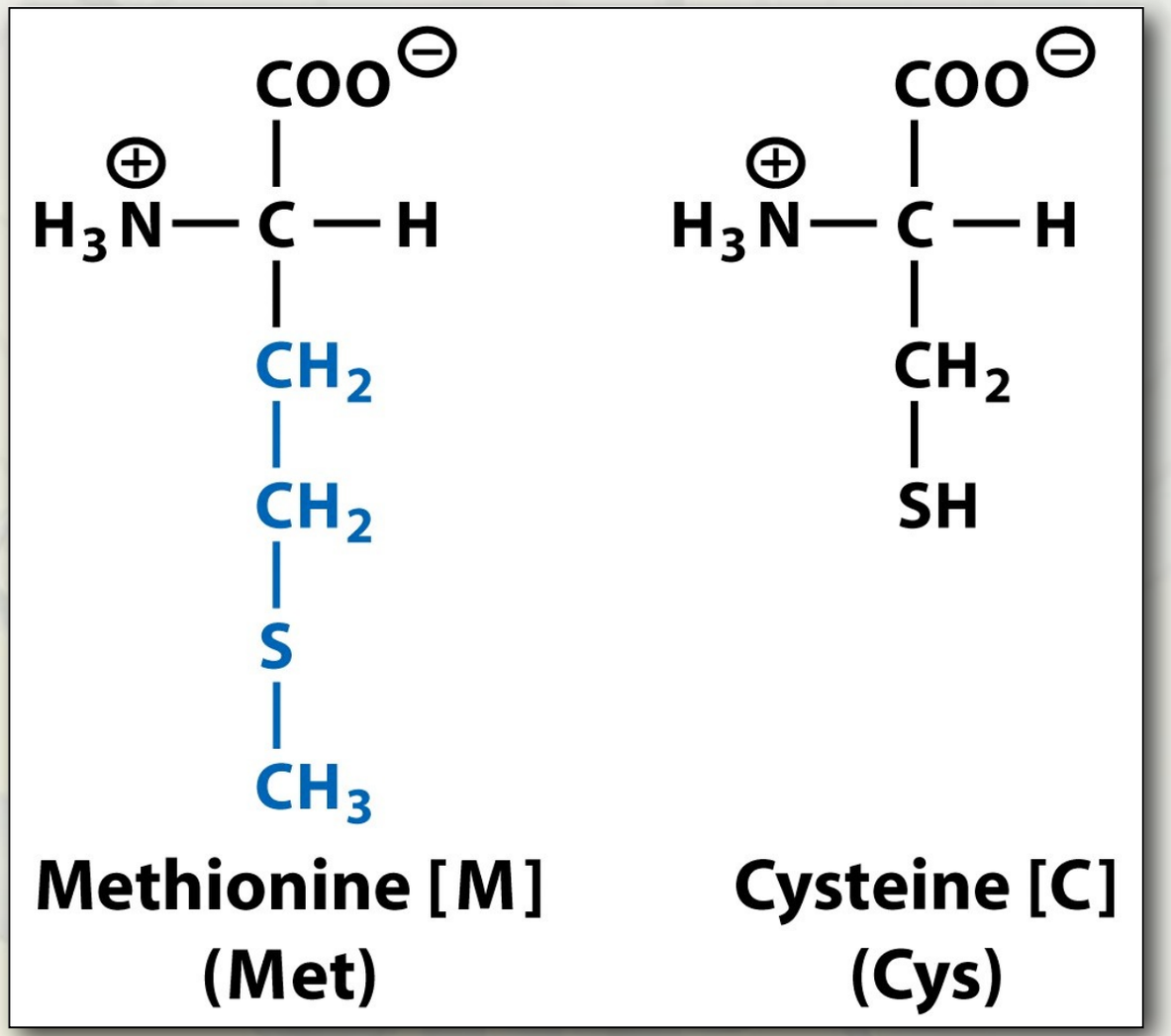
The Amino Acids

- Aromatic (F,Y,W)



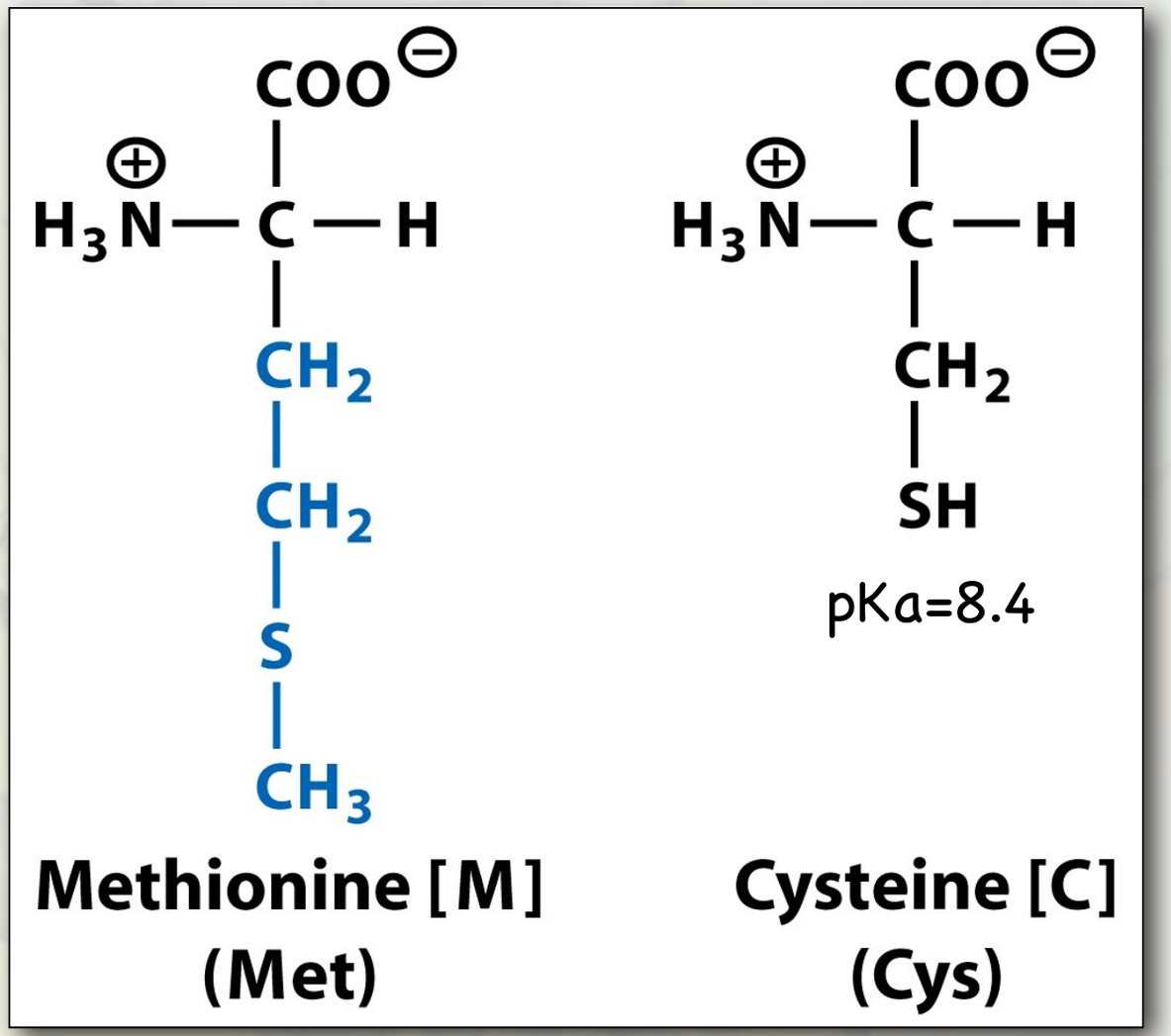
The Amino Acids

- Sulfur-containing (C,M)



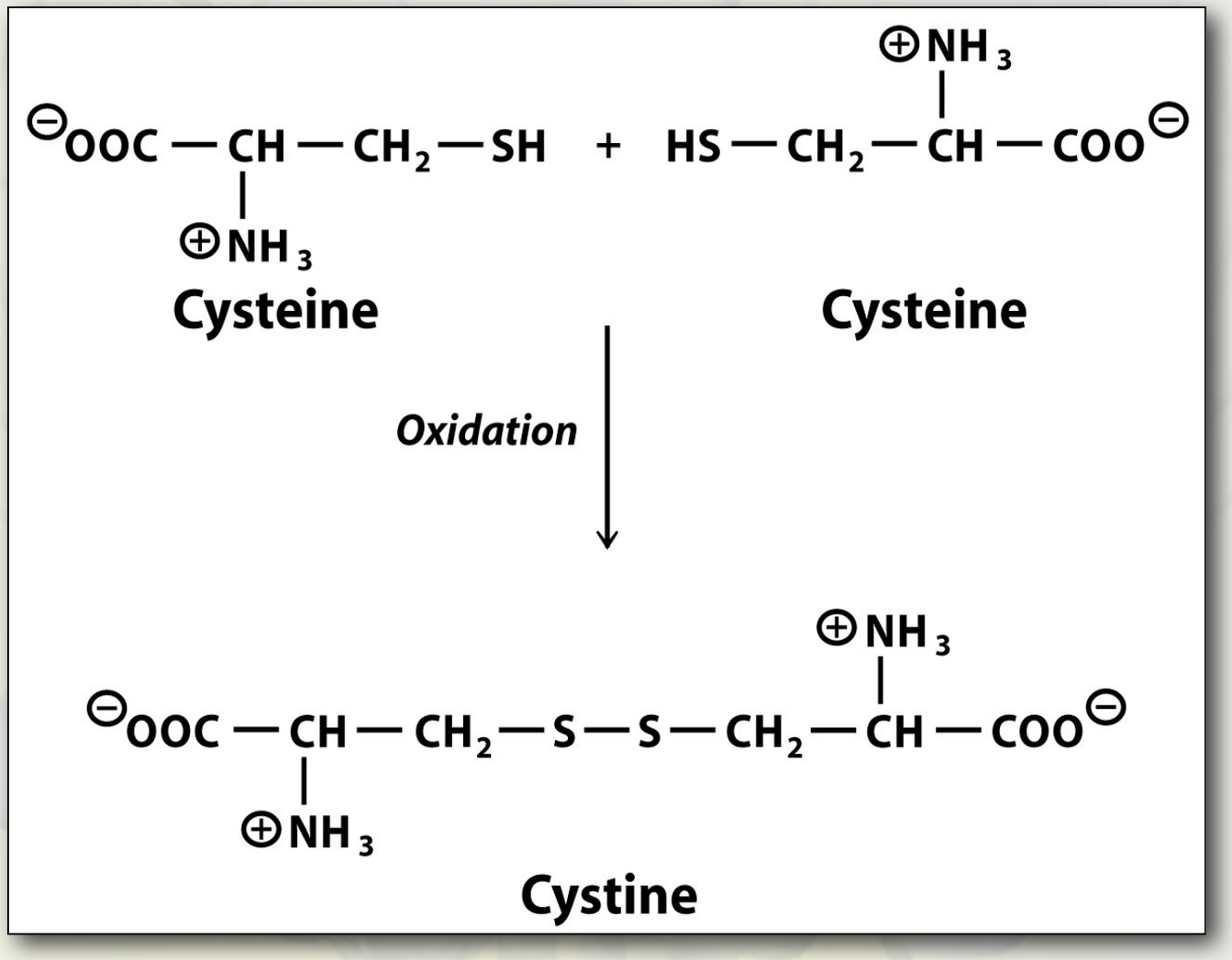
The Amino Acids

- Sulfur-containing (C,M)



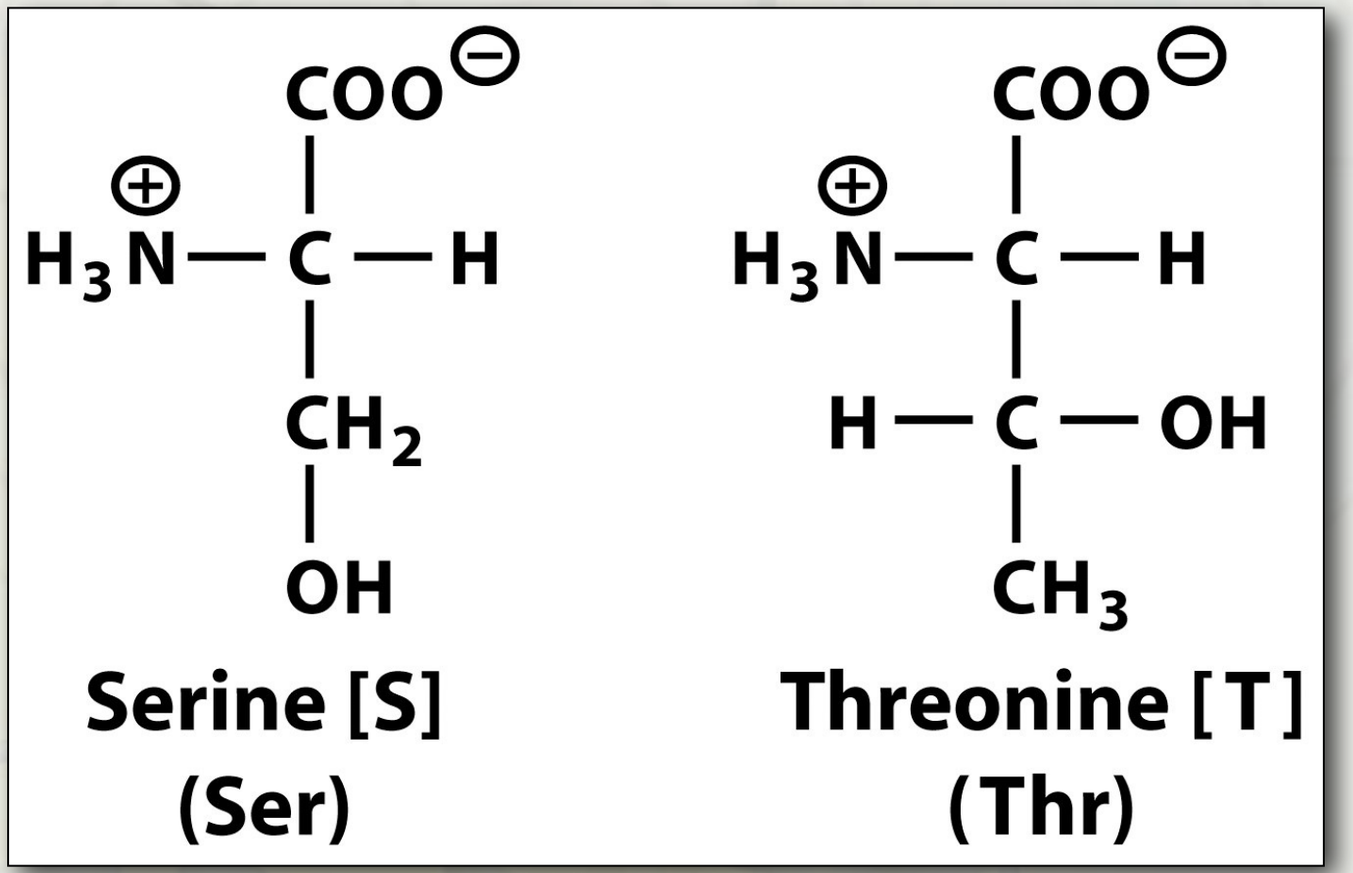
The Amino Acids

- Sulfur-containing (C,M)



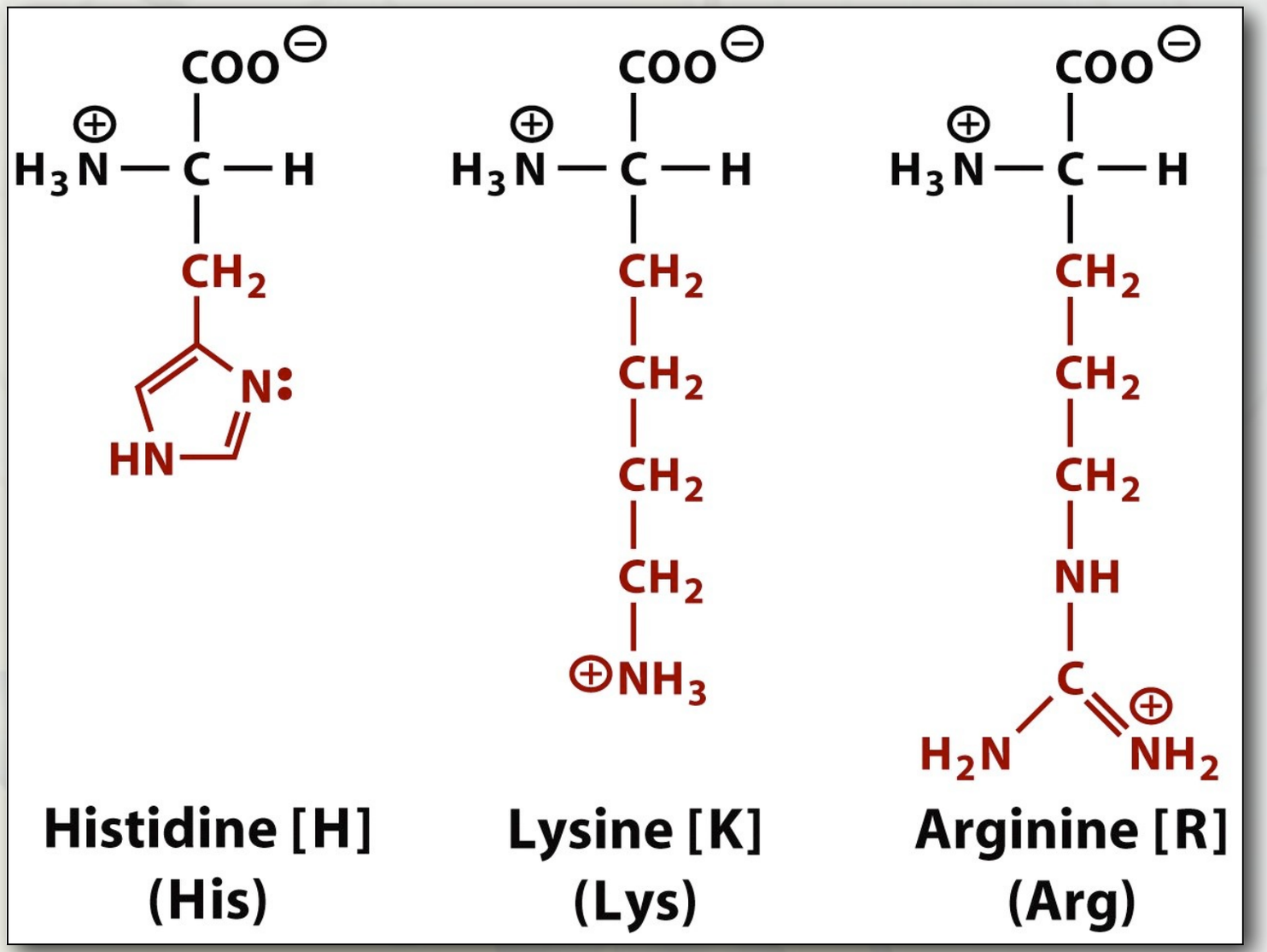
The Amino Acids

- Alcohols (S,T)



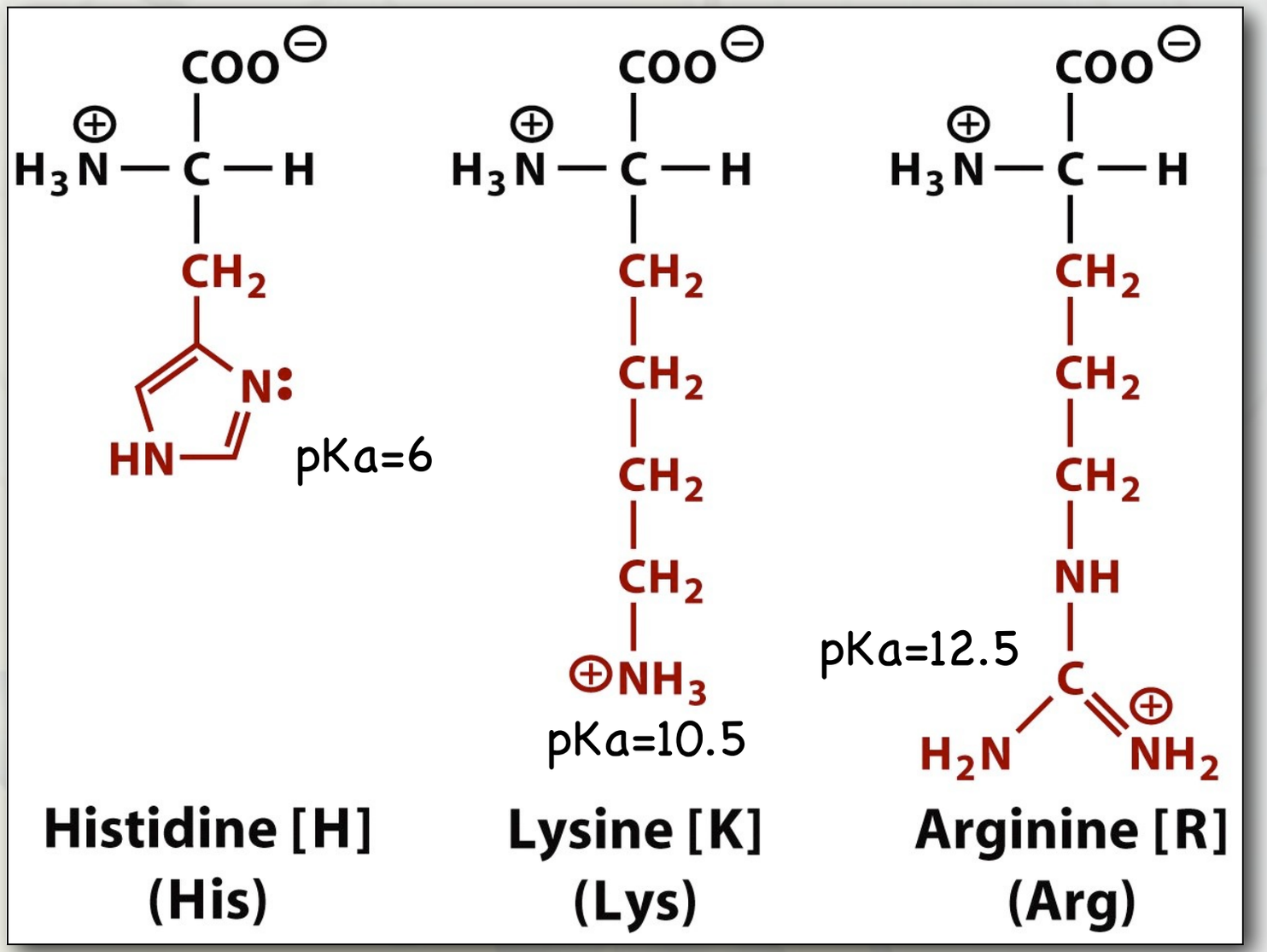
The Amino Acids

- Bases (H,K,R)



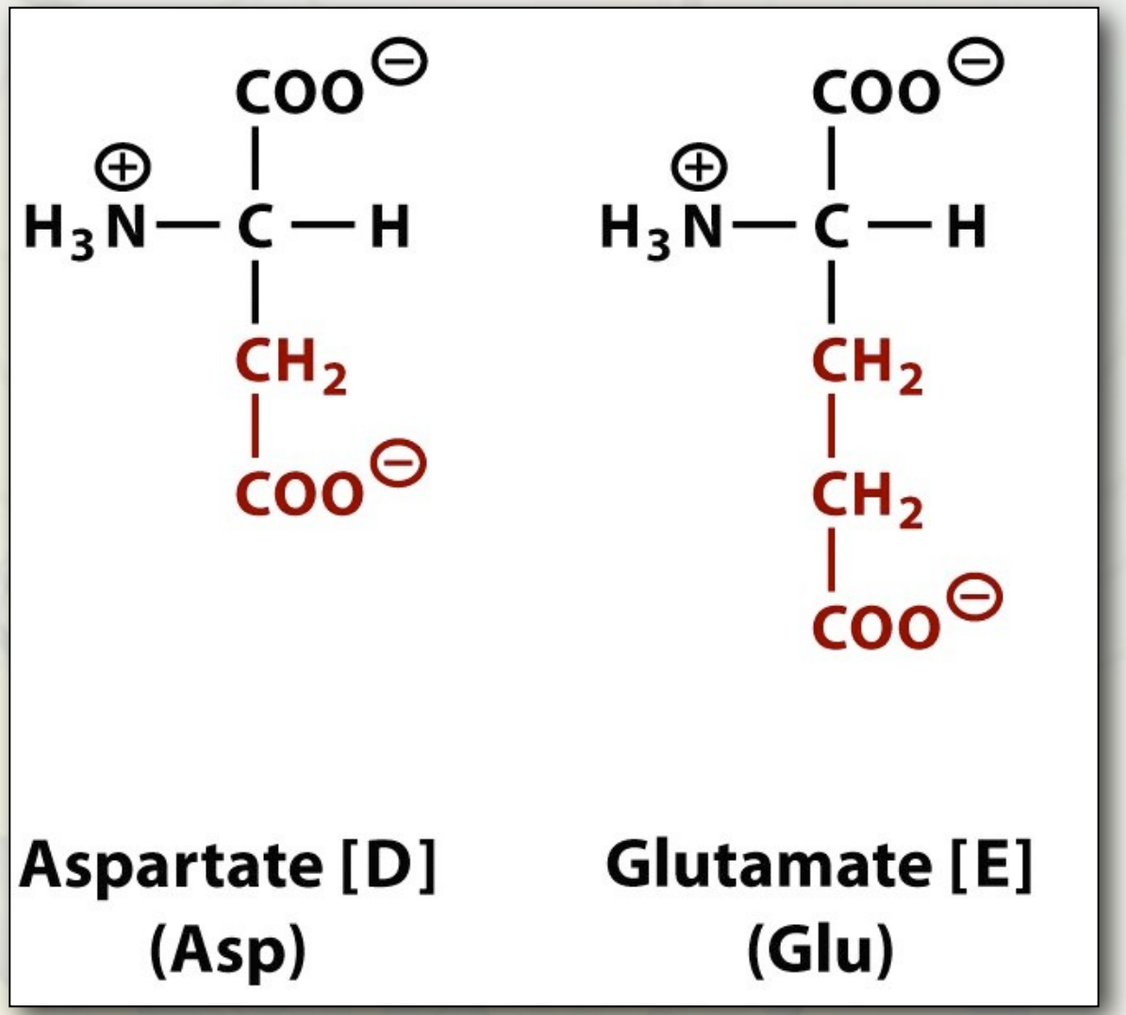
The Amino Acids

- Bases (H,K,R)



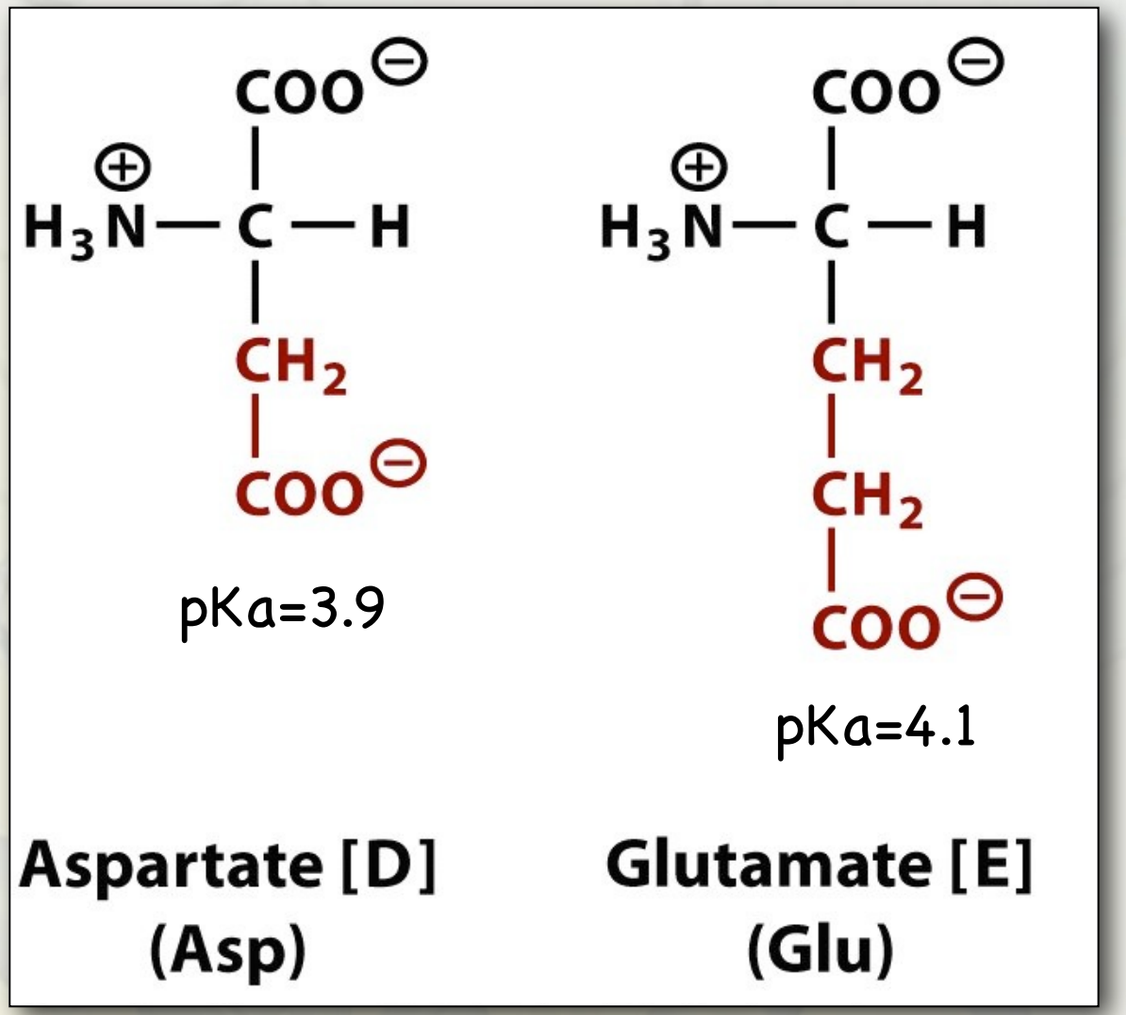
The Amino Acids

- Acids (D,E)



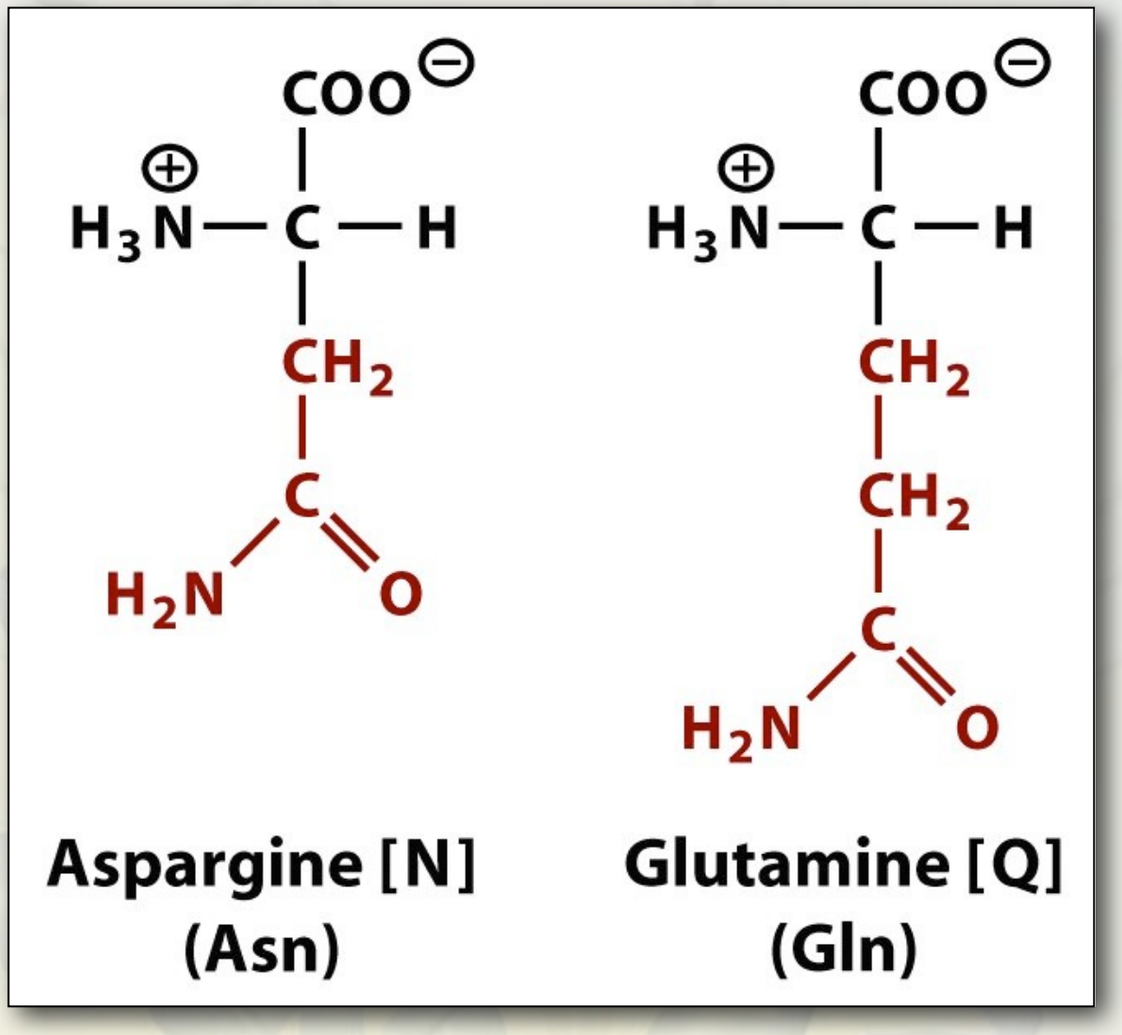
The Amino Acids

- Acids (D,E)



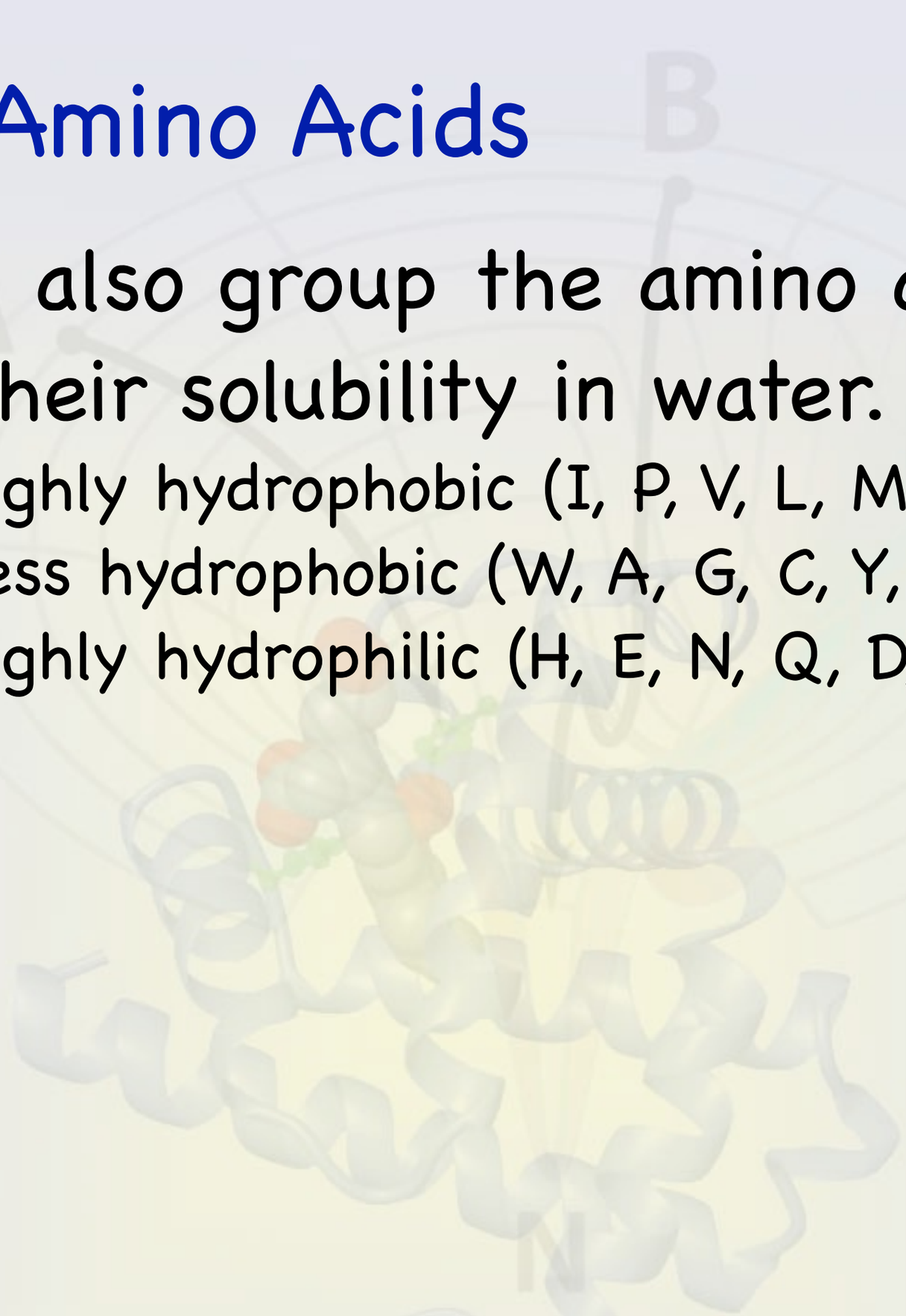
The Amino Acids

- Amides (N,Q)



The Amino Acids

- Can also group the amino acids based on their solubility in water.
  - Highly hydrophobic (I, P, V, L, M)
  - Less hydrophobic (W, A, G, C, Y, P, T, S)
  - Highly hydrophilic (H, E, N, Q, D, K, R)



The Amino Acids

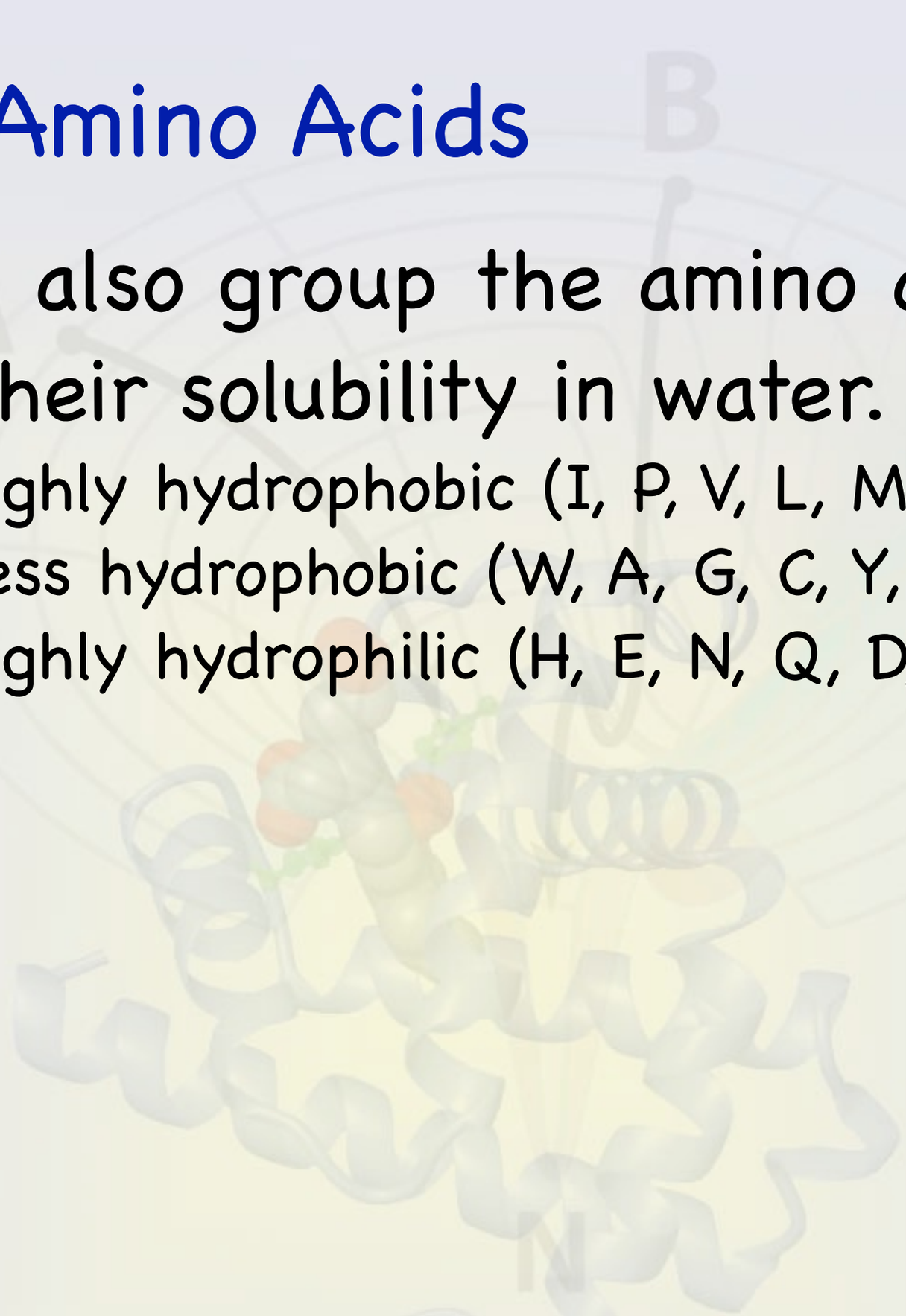
- Can also … acids based on their …
  - Highly h…)
  - Less hy… P, T, S)
  - Highly h… K, R)

TABLE 3.1 Hydropathy scale for amino acid residues

Amino acid	Free-energy change for transfer[a] (kj mol^{-1})
Highly hydrophobic	
Isoleucine	3.1
Phenylalanine	2.5
Valine	2.3
Leucine	2.2
Methionine	1.1
Less hydrophobic	
Tryptophan	1.5[b]
Alanine	1.0
Glycine	0.67
Cysteine	0.17
Tyrosine	0.08
Proline	−0.29
Threonine	−0.75
Serine	−1.1
Highly hydrophilic	
Histidine	−1.7
Glutamate	−2.6
Asparagine	−2.7
Glutamine	−2.9
Aspartate	−3.0
Lysine	−4.6
Arginine	−7.5

The Amino Acids

- Can also group the amino acids based on their solubility in water.
  - Highly hydrophobic (I, P, V, L, M)
  - Less hydrophobic (W, A, G, C, Y, P, T, S)
  - Highly hydrophilic (H, E, N, Q, D, K, R)

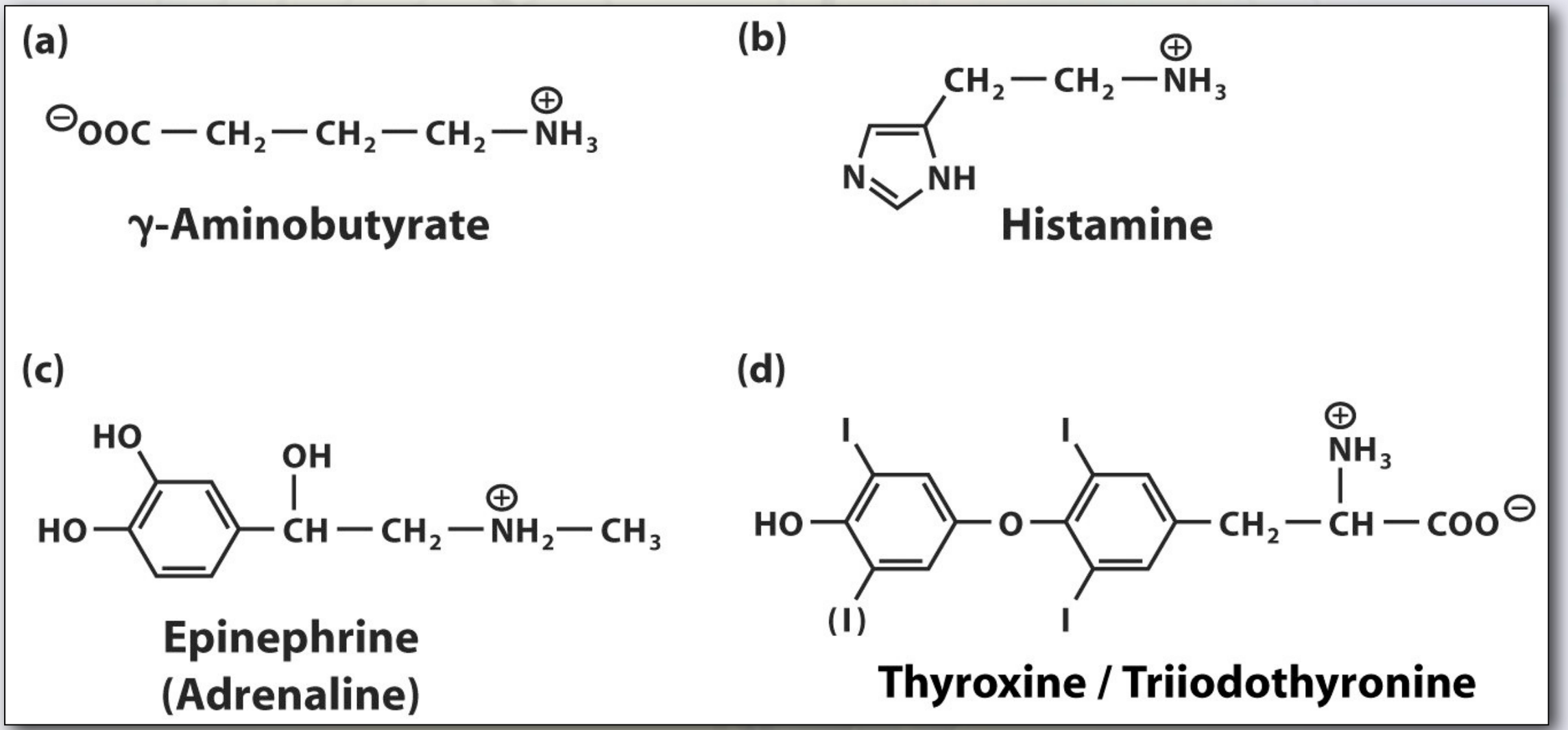


The Amino Acids

- Many important biological molecules are derived from amino acids, e.g.,
  - **histamine** (derived from His) (triggers allergic response)
  - **epinephrine** (derived from Tyr) (hormone that triggers "flight or fight" response)
  - **thyroxine** (derived from Tyr) (thyroid hormone) (one of the few uses for iodine)
  - **GABA** (derived from Glu) (neurotransmitter)

The Amino Acids

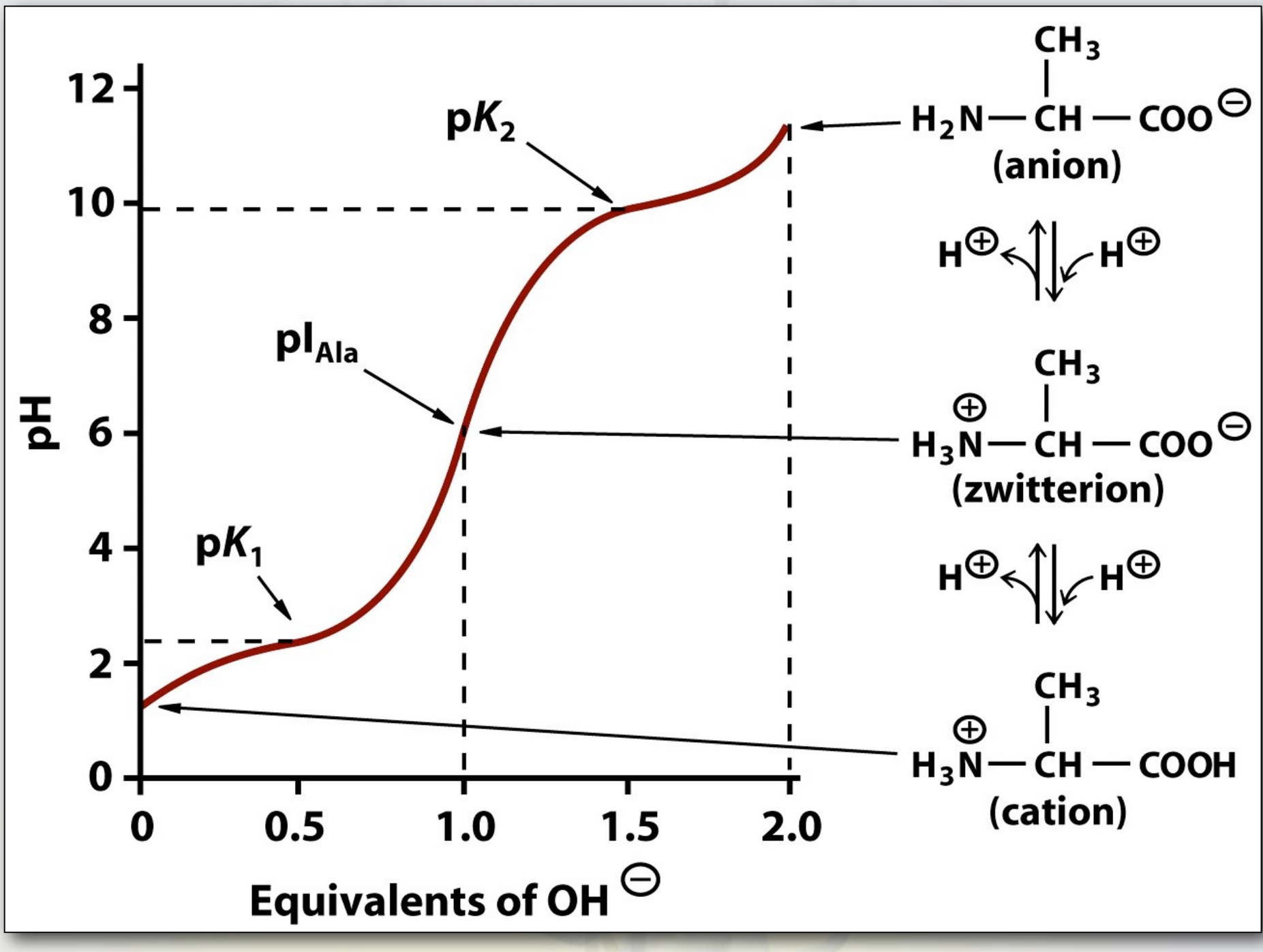
- Many important biological molecules are derived from amino acids, e.g.,



The Amino Acids

- Many important biological molecules are derived from amino acids, e.g.,
  - **histamine** (derived from His) (triggers allergic response)
  - **epinephrine** (derived from Tyr) (hormone that triggers "flight or fight" response)
  - **thyroxine** (derived from Tyr) (thyroid hormone) (one of the few uses for iodine)
  - **GABA** (derived from Glu) (neurotransmitter)

The Amino Acids as Acids & Bases



The Amin s & Bases

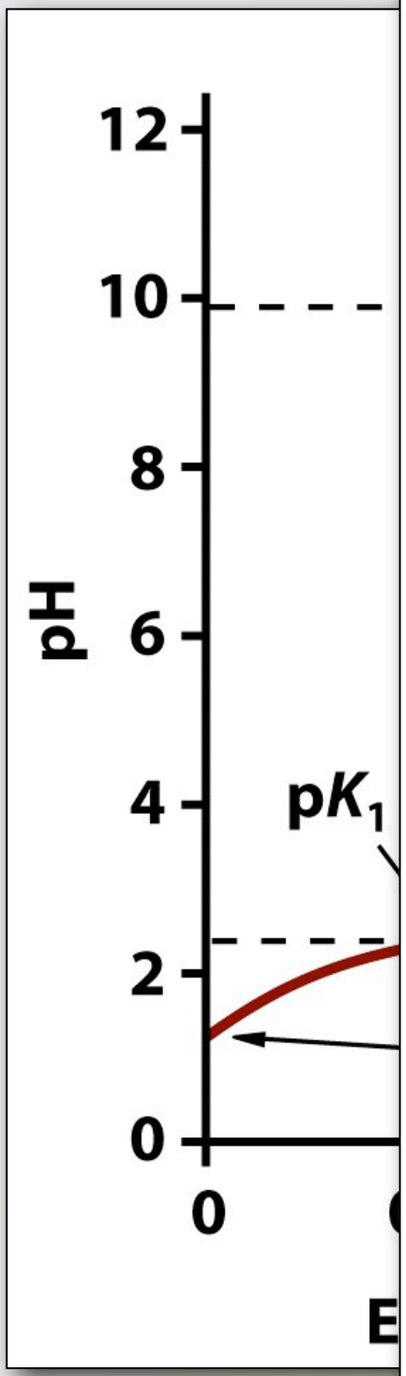
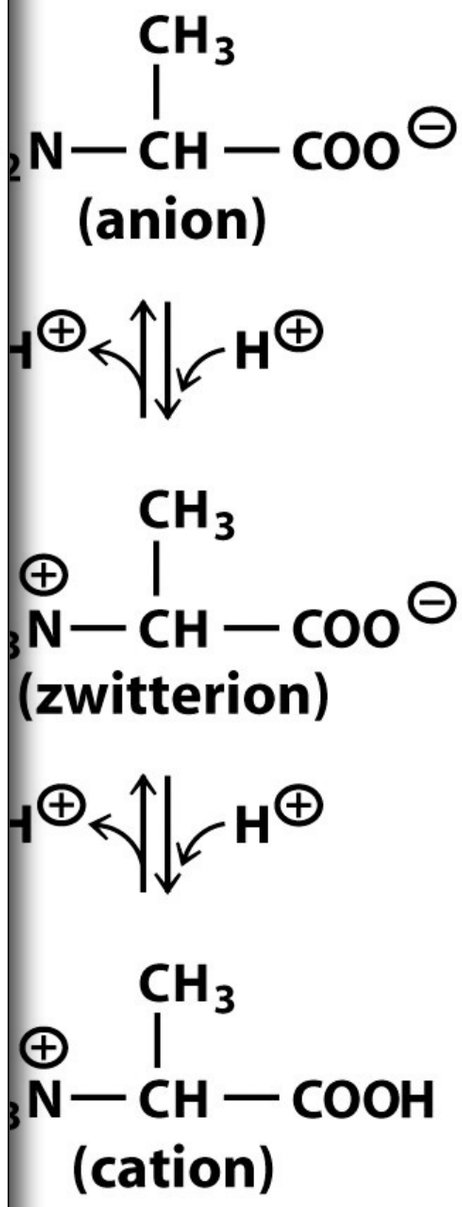
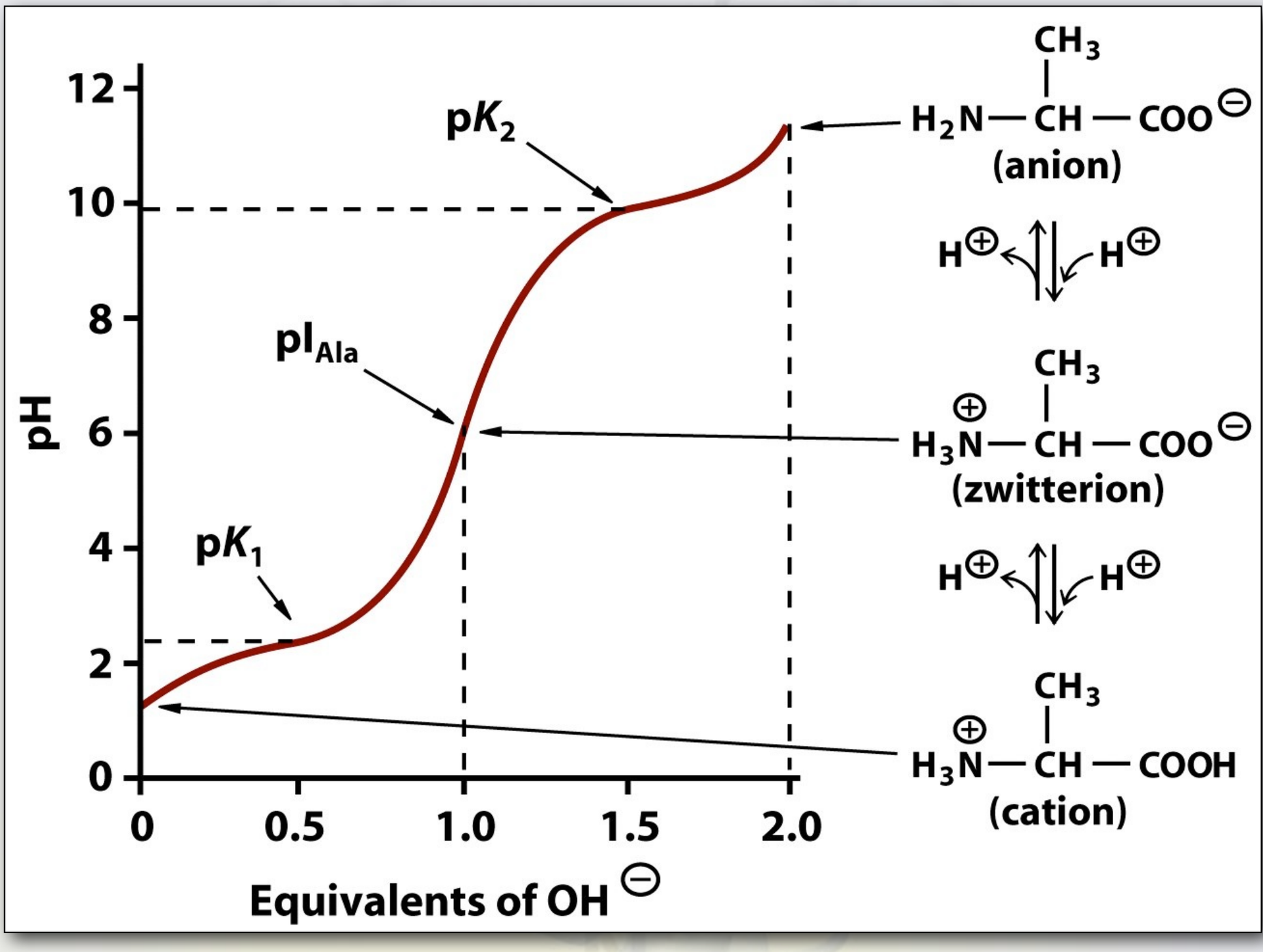


TABLE 3.2 pK_a values of acidic and basic constituents of free amino acids at 25°C

Amino acid	pKa value		
	Carboxyl group	Amino group	Side chain
Glycine	2.4	9.8	
Alanine	2.4	9.9	
Valine	2.3	9.7	
Leucine	2.3	9.7	
Isoleucine	2.3	9.8	
Methionine	2.1	9.3	
Proline	2.0	10.6	
Phenylalanine	2.2	9.3	
Tryptophan	2.5	9.4	
Serine	2.2	9.2	
Threonine	2.1	9.1	
Cysteine	1.9	10.7	8.4
Tyrosine	2.2	9.2	10.5
Asparagine	2.1	8.7	
Glutamine	2.2	9.1	
Aspartic acid	2.0	9.9	3.9
Glutamic acid	2.1	9.5	4.1
Lysine	2.2	9.1	10.5
Arginine	1.8	9.0	12.5
Histidine	1.8	9.3	6.0



The Amino Acids as Acids & Bases

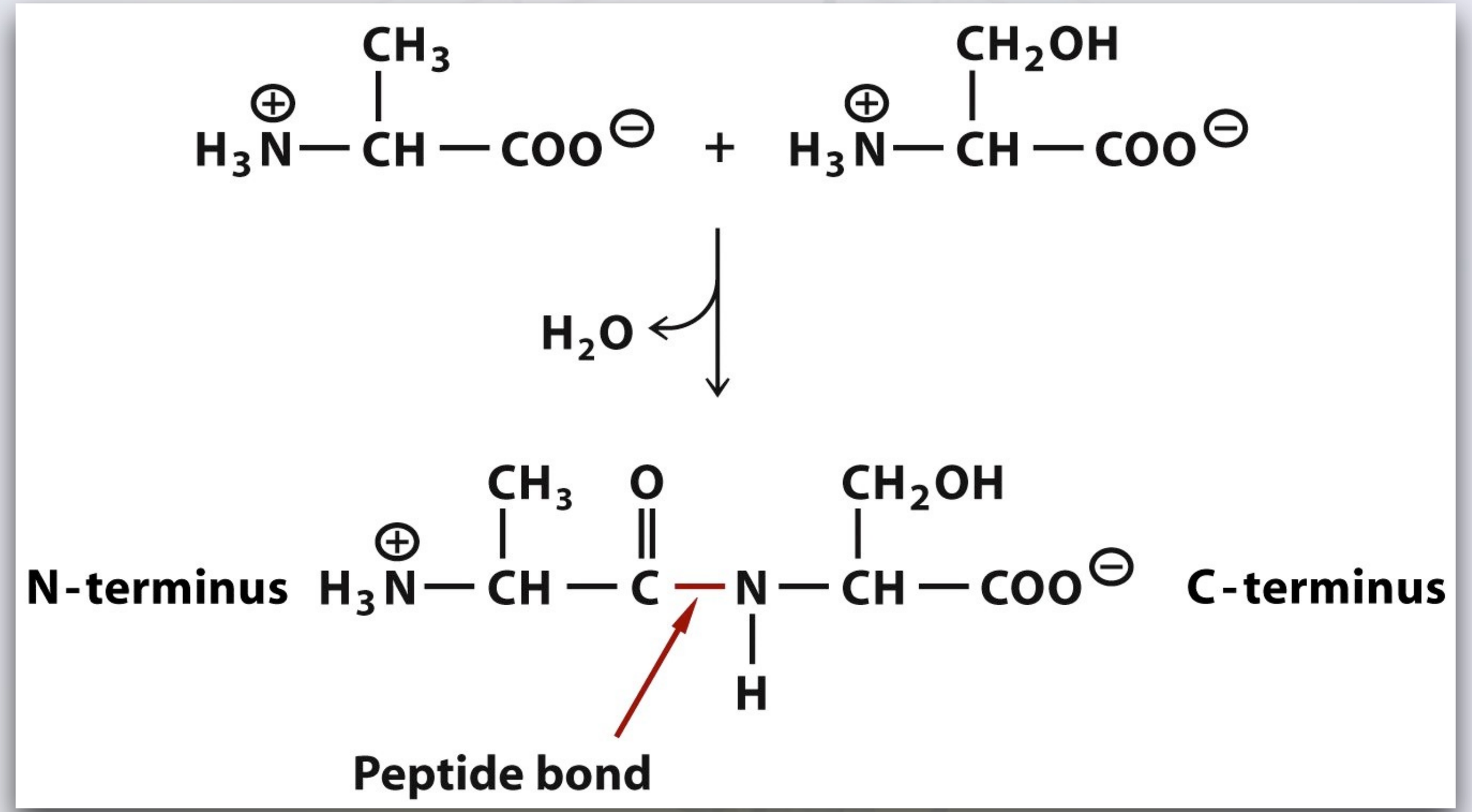


The Peptide Bond

Amino acids are joined together by peptide (amide) bonds.

- A peptide bond is formed from the condensation of an α-amino group from one amino acid with the α-carboxyl group of another amino acid.

The Peptide Bond



The Peptide Bond

Amino acids are joined together by peptide (amide) bonds.

- A peptide bond is formed from the condensation of an α-amino group from one amino acid with the α-carboxyl group of another amino acid.

Peptides

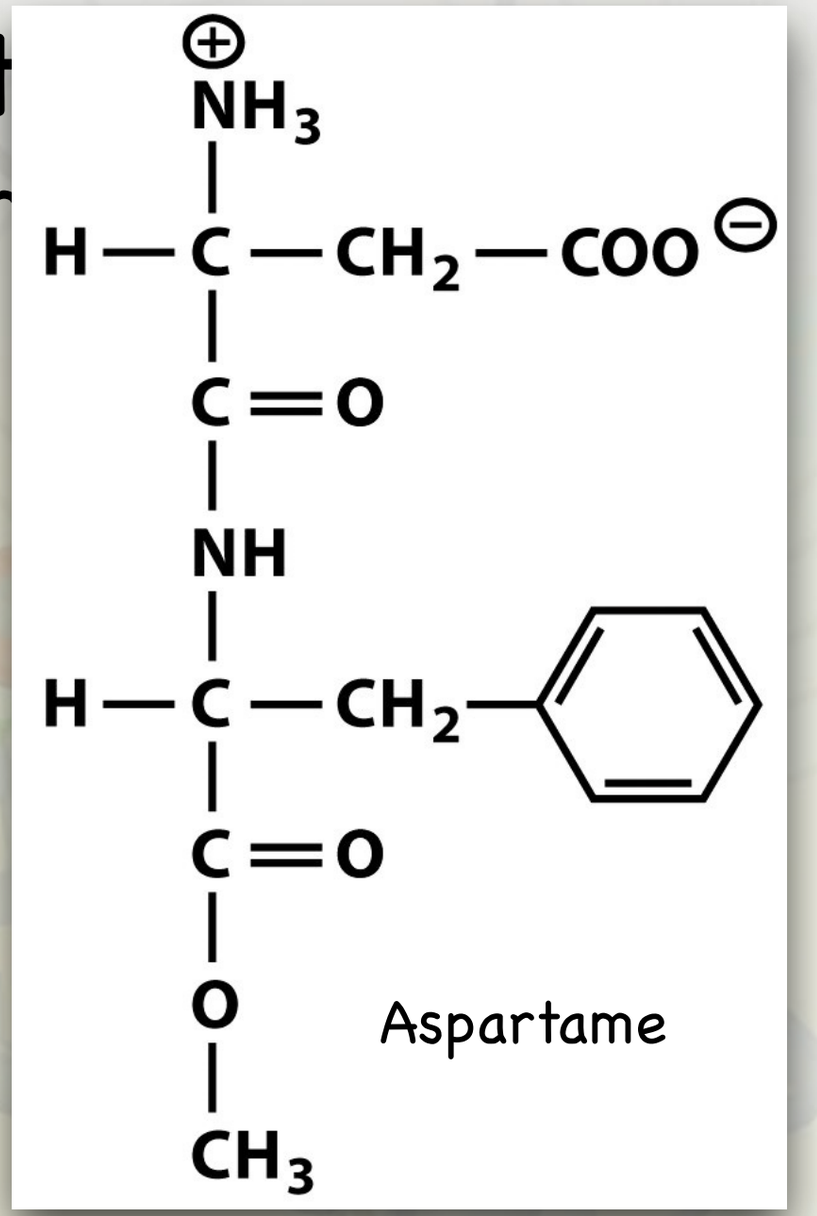
Some small peptide have important biological activities.

- peptide hormones
- toxins
- sweeteners

Peptides

Some small peptide have important biological act

- peptide horm
- toxins
- sweeteners



Aspartame

Peptides

Some small peptide have important biological activities.

- peptide hormones
- toxins
- sweeteners

Peptides

Problem:

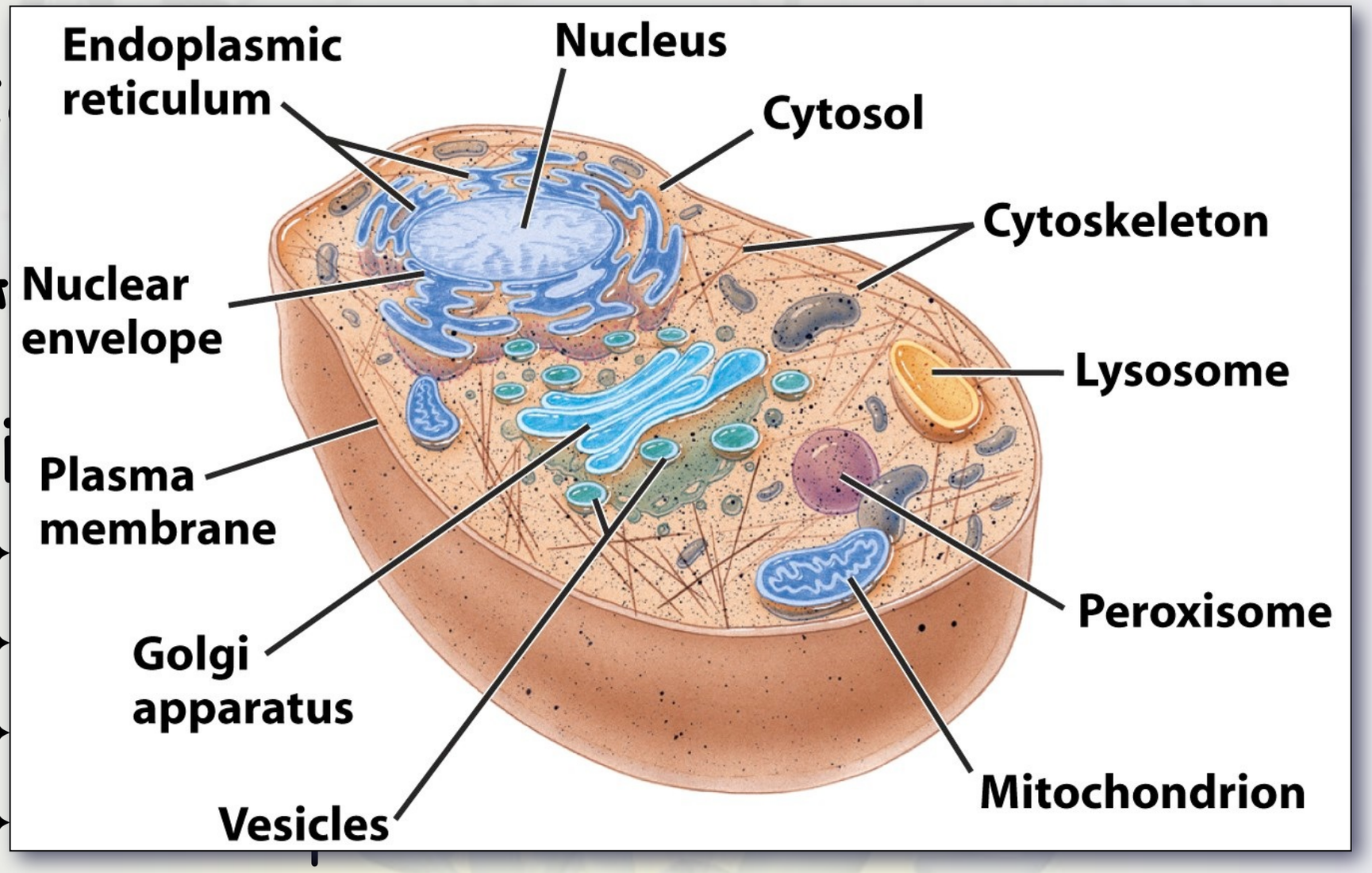
Sketch the titration curve for the dipeptide cysteinylaspartic acid and draw the structures of the predominant species that exists at each of the endpoints. Confirm your answers using Marvinsketch.

Isolation of Peptides and Proteins

- Cell disruption
- Centrifugation
- Ammonium sulfate precipitation
- Liquid chromatography
  - ion exchange
  - size exclusion (gel filtration)
  - affinity chromatography
  - reverse phase.

Isolation of Peptides and Proteins

- Cell disruption
- C
- A
- Li
  - ✦
  - ✦
  - ✦
  - ✦

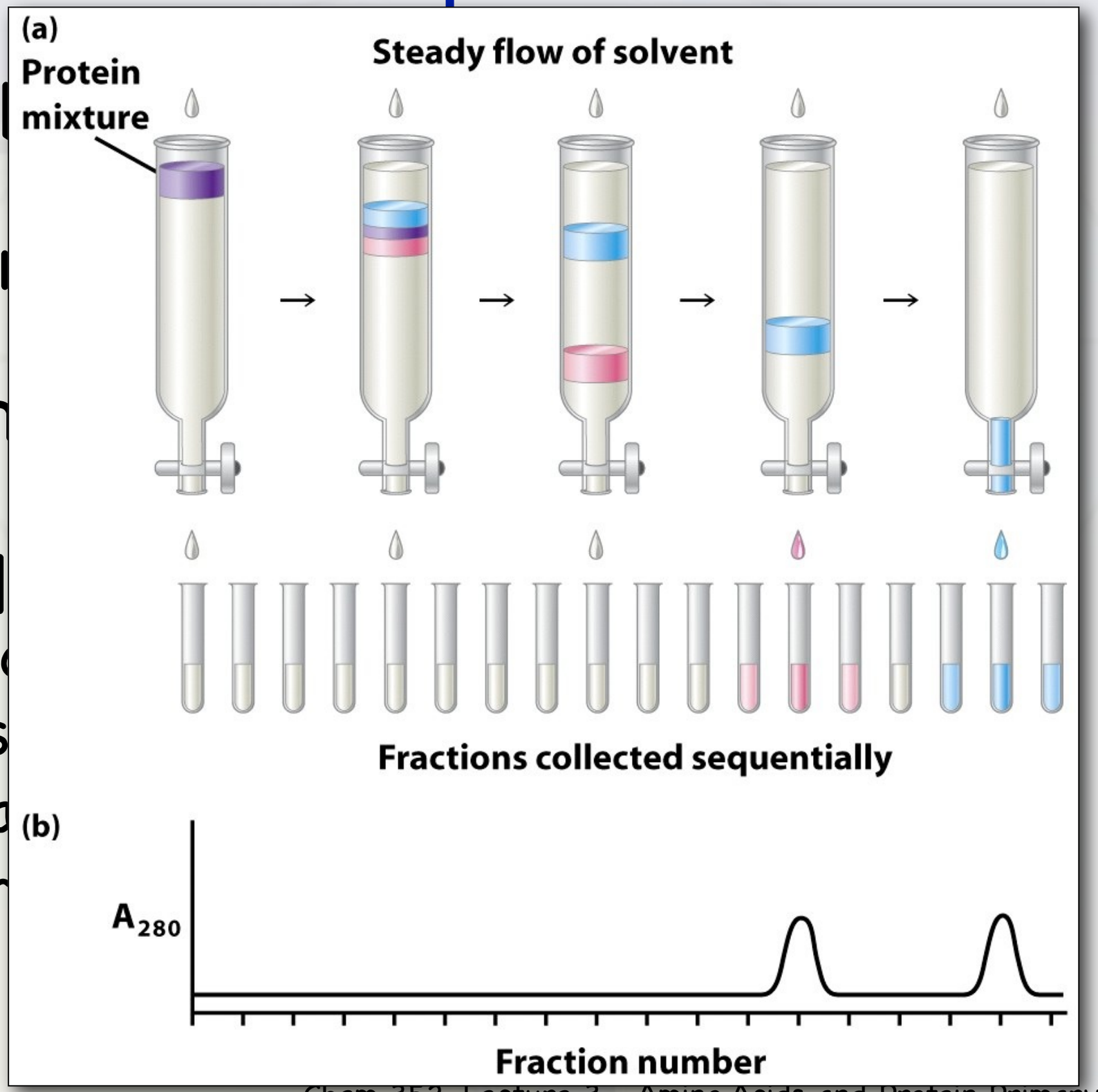


Isolation of Peptides and Proteins

- Cell disruption
- Centrifugation
- Ammonium sulfate precipitation
- Liquid chromatography
  - ion exchange
  - size exclusion (gel filtration)
  - affinity chromatography
  - reverse phase.

Isolation of Peptides and Proteins

- Cel
- Cel
- Am
- Liq
  - i
  - s
  - c
  - r

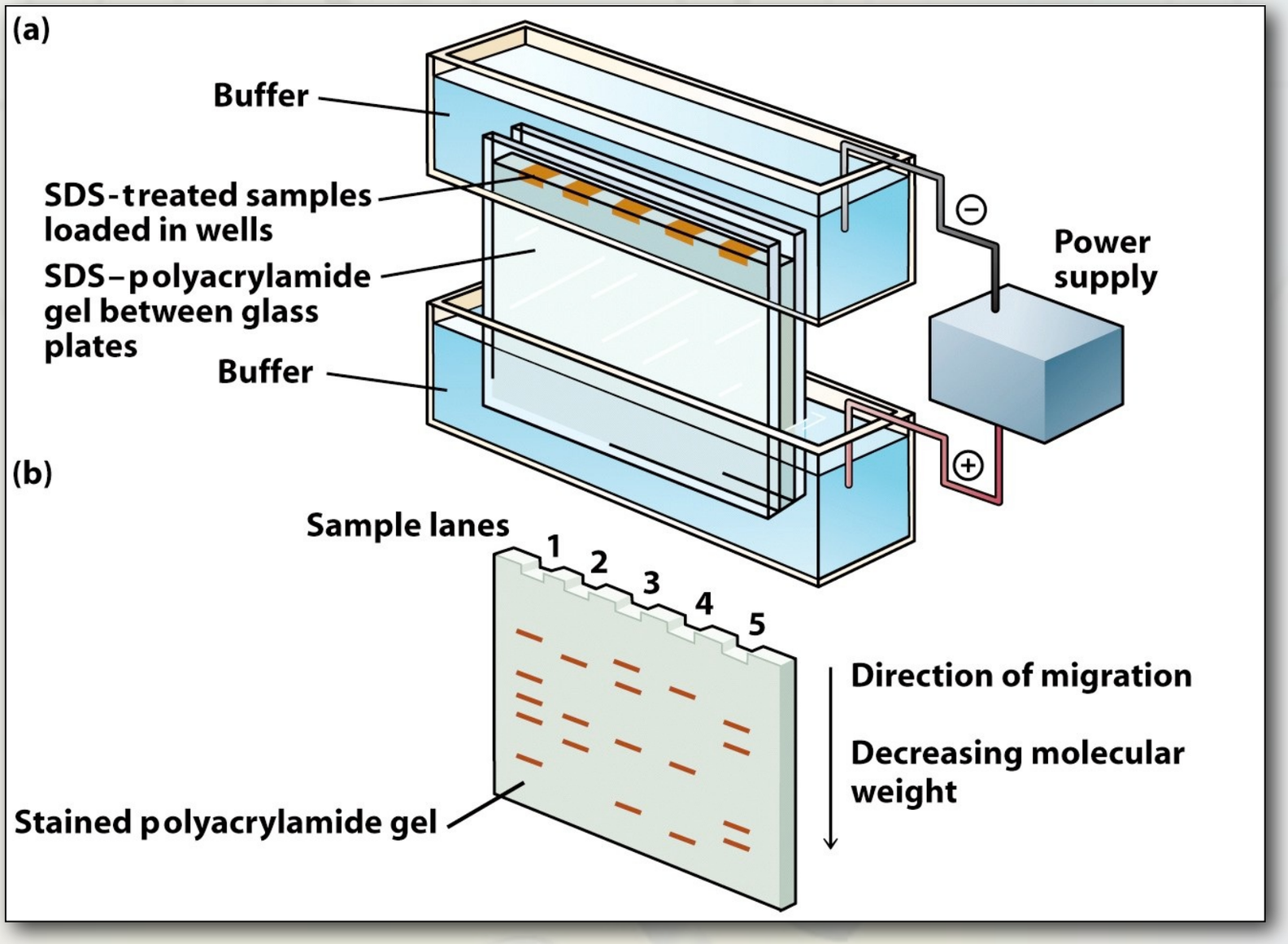


Isolation of Peptides and Proteins

- Cell disruption
- Centrifugation
- Ammonium sulfate precipitation
- Liquid chromatography
  - ion exchange
  - size exclusion (gel filtration)
  - affinity chromatography
  - reverse phase.

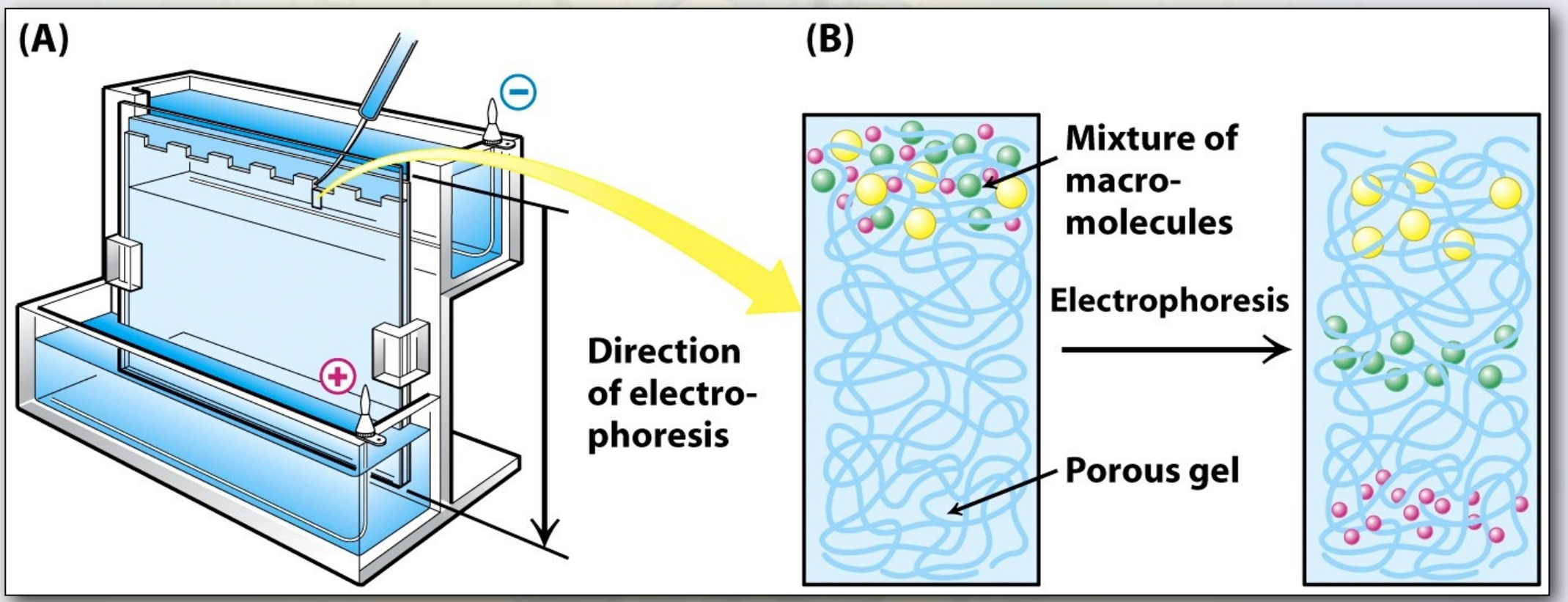
Protein Isolation and Analysis

- SDS-PAGE (SDS-polyacrylamide gel electrophoresis)



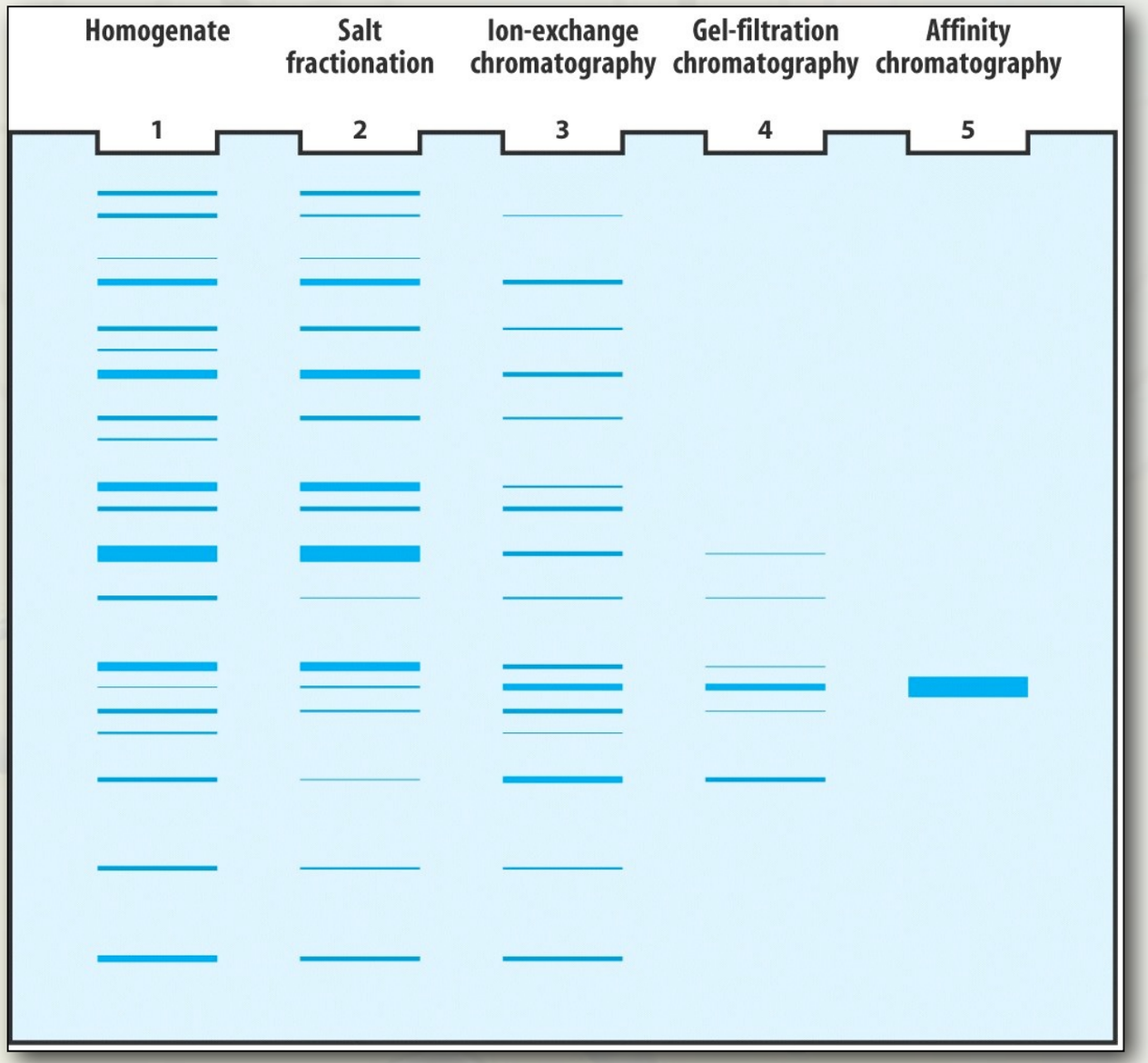
Protein Isolation and Analysis

- SDS-PAGE (SDS-polyacrylamide gel electrophoresis)



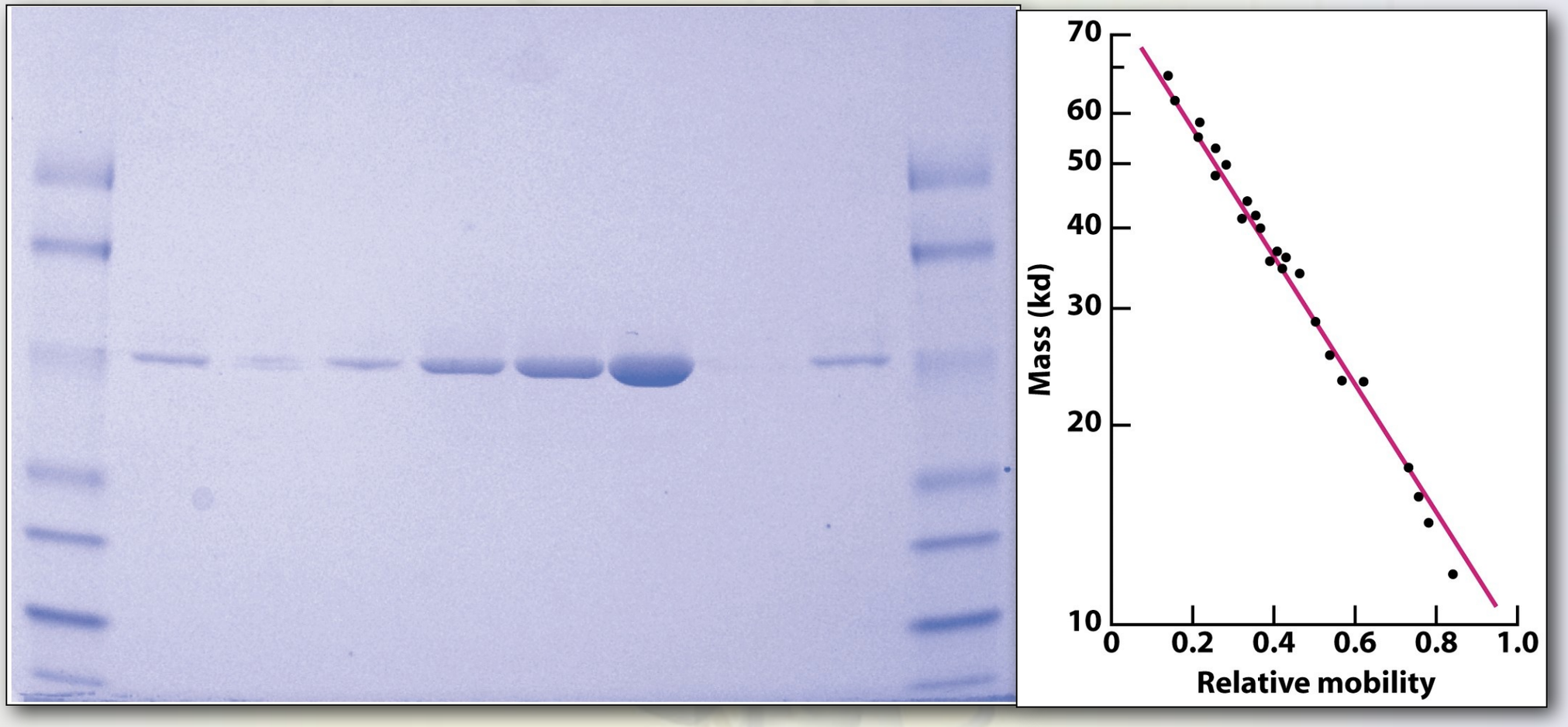
Protein Isolation and Analysis

- SDS-PAGE (SDS-polyacrylamide gel electrophoresis)



Protein Isolation and Analysis

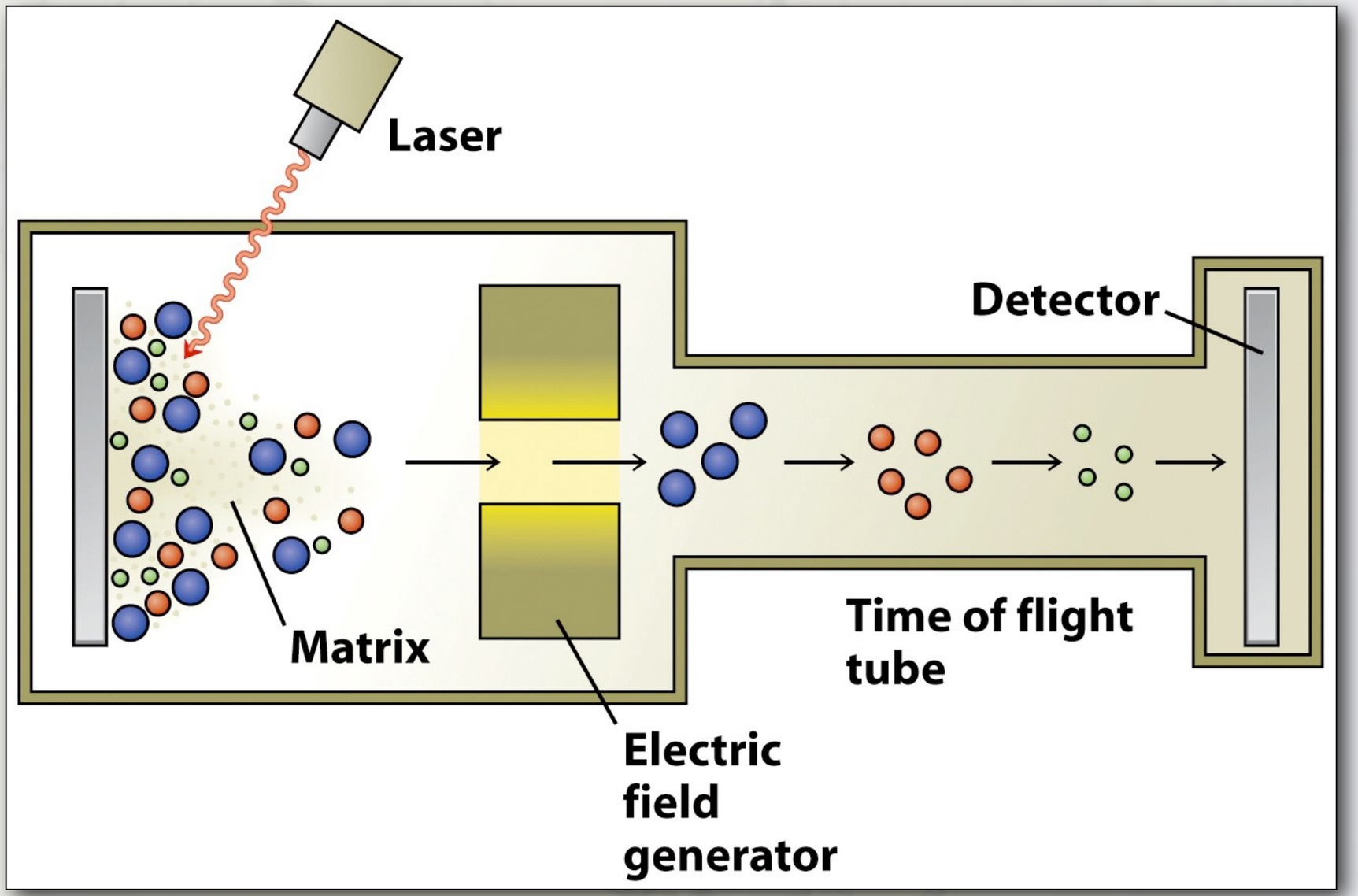
- SDS-PAGE (SDS-polyacrylamide gel electrophoresis)



Protein Characterization

- Mass spectroscopy

MALDI
Ionization



Protein Characterization

- Mass spectroscopy

ESI
Ionization

Protein Characterization

- Mass spectroscopy

ESI
Ionization

John B. Fenn (1917–)

Koichi Tanaka (1959–)

MALDI
Ionization

Nobel Prize in Chemistry, 2002

Protein Characterization

- Mass spectroscopy

ESI
Ionization

John B. Fenn (1917- 2010)

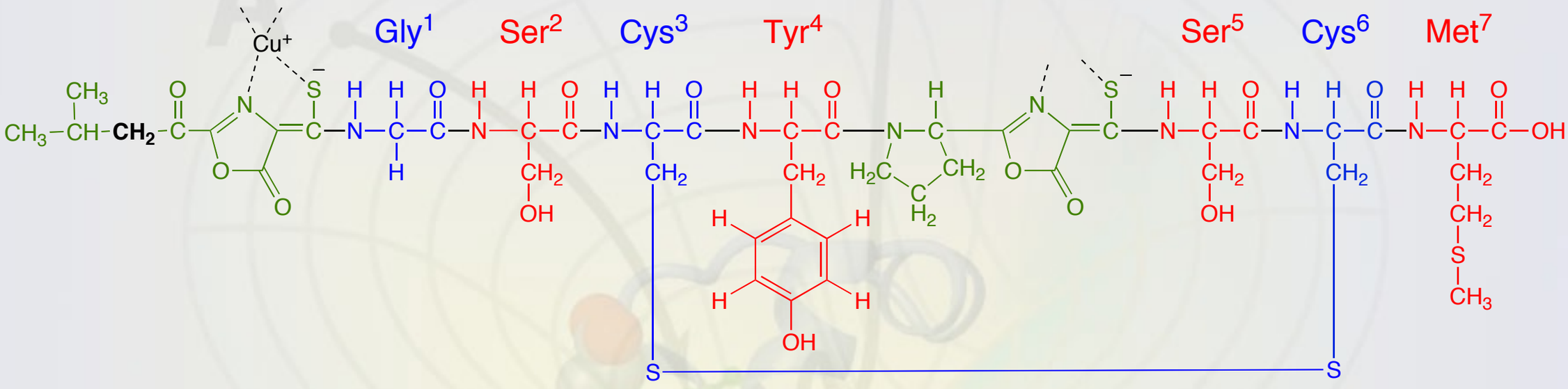
MALDI
Ionization

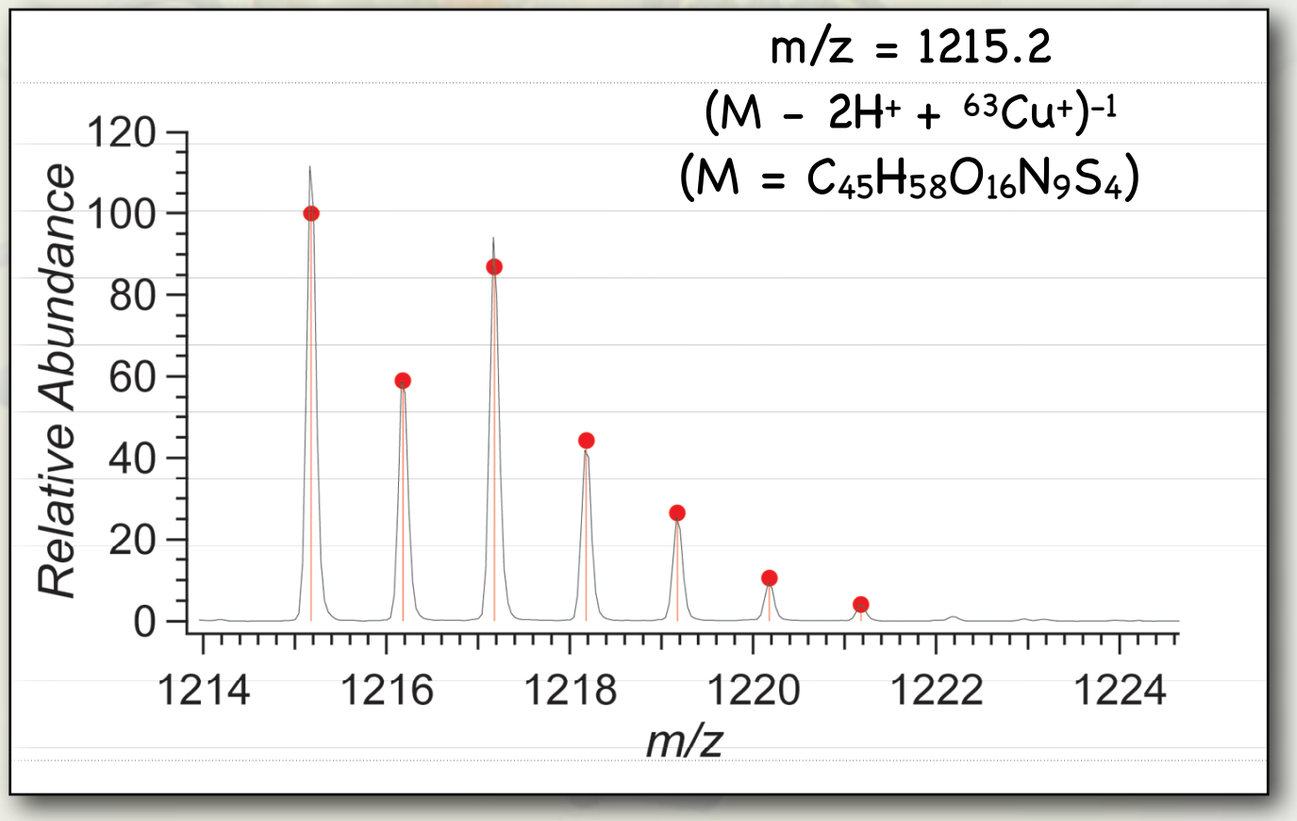
Koichi Tanaka (1959–)

Nobel Prize in Chemistry, 2002

m/z = 1215.1781
$(M - 2H^+ + {}^{63}Cu^+)^{-1}$

Methanobactin

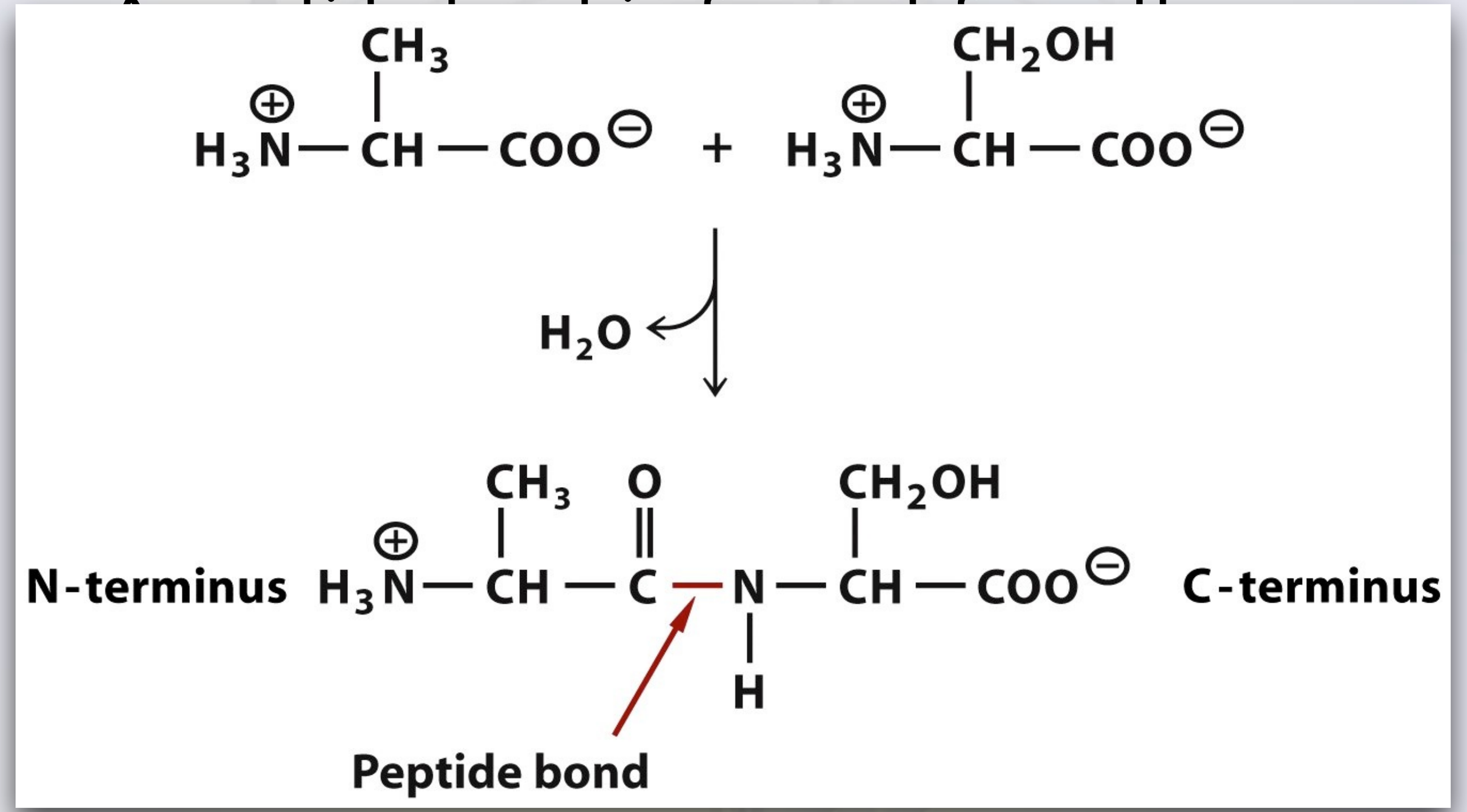




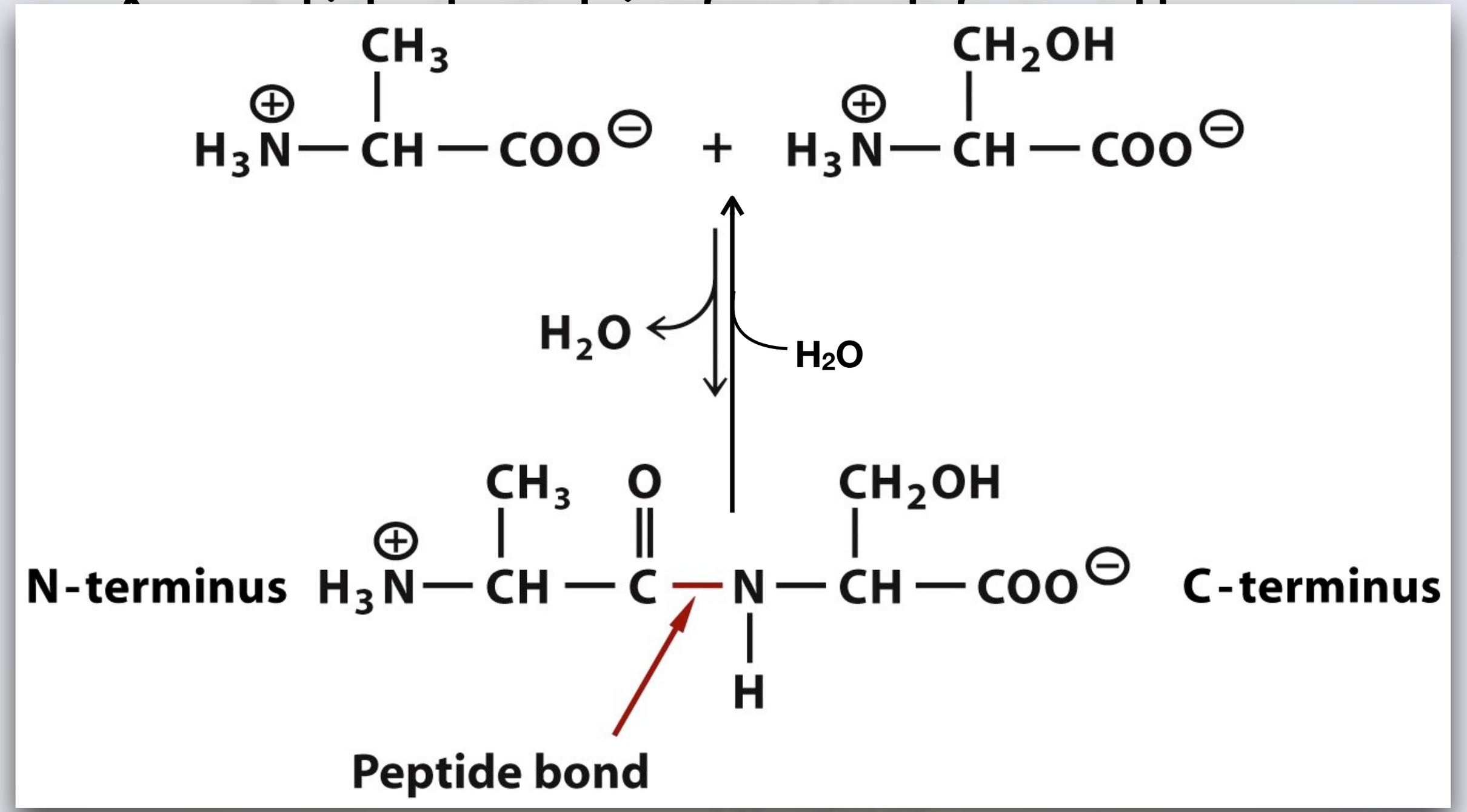
Protein Primary Structure

A peptide bond is formed from the condensation of an α-amino group from one amino acid with the α-carboxyl group of another amino acid.

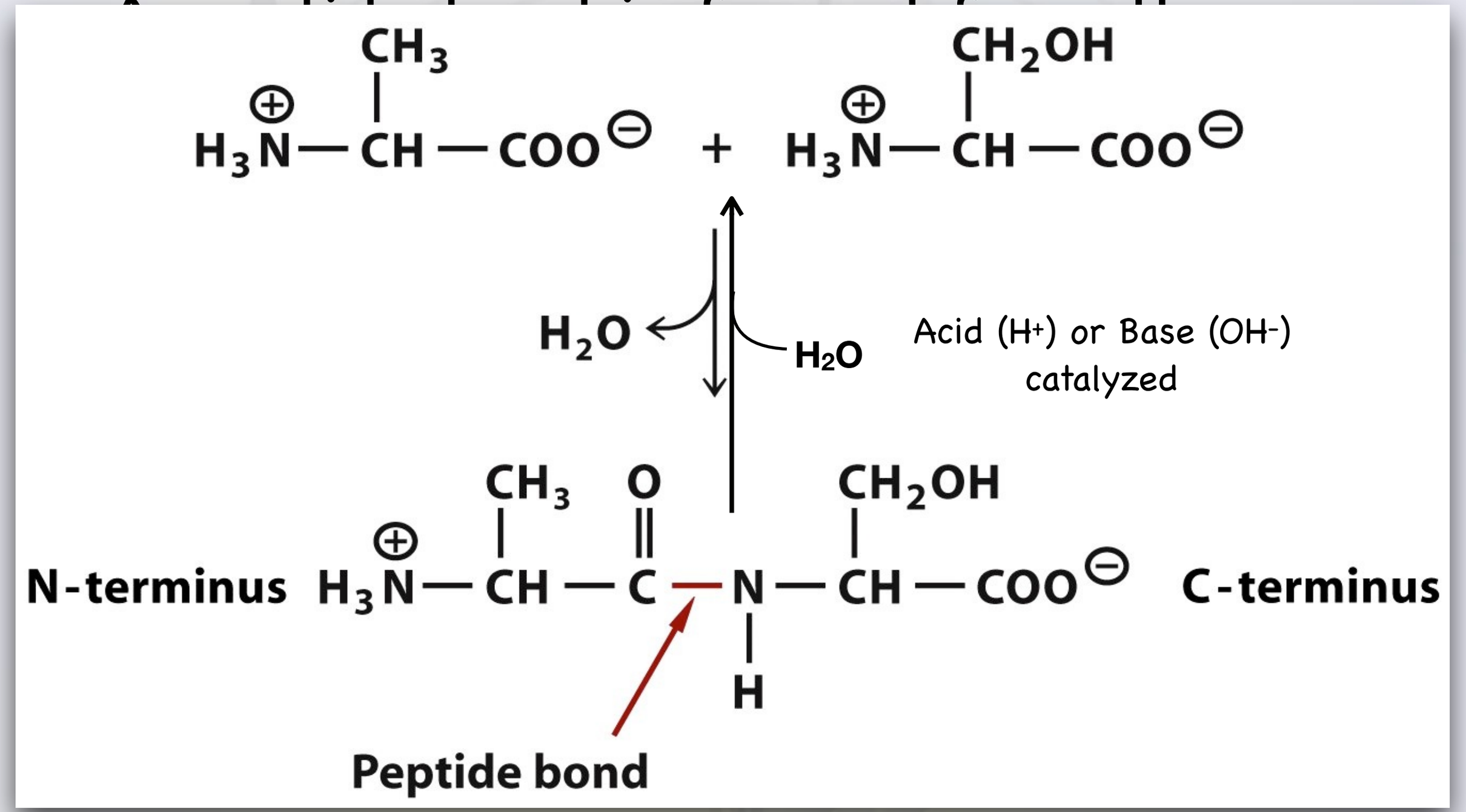
Protein Primary Structure



Protein Primary Structure

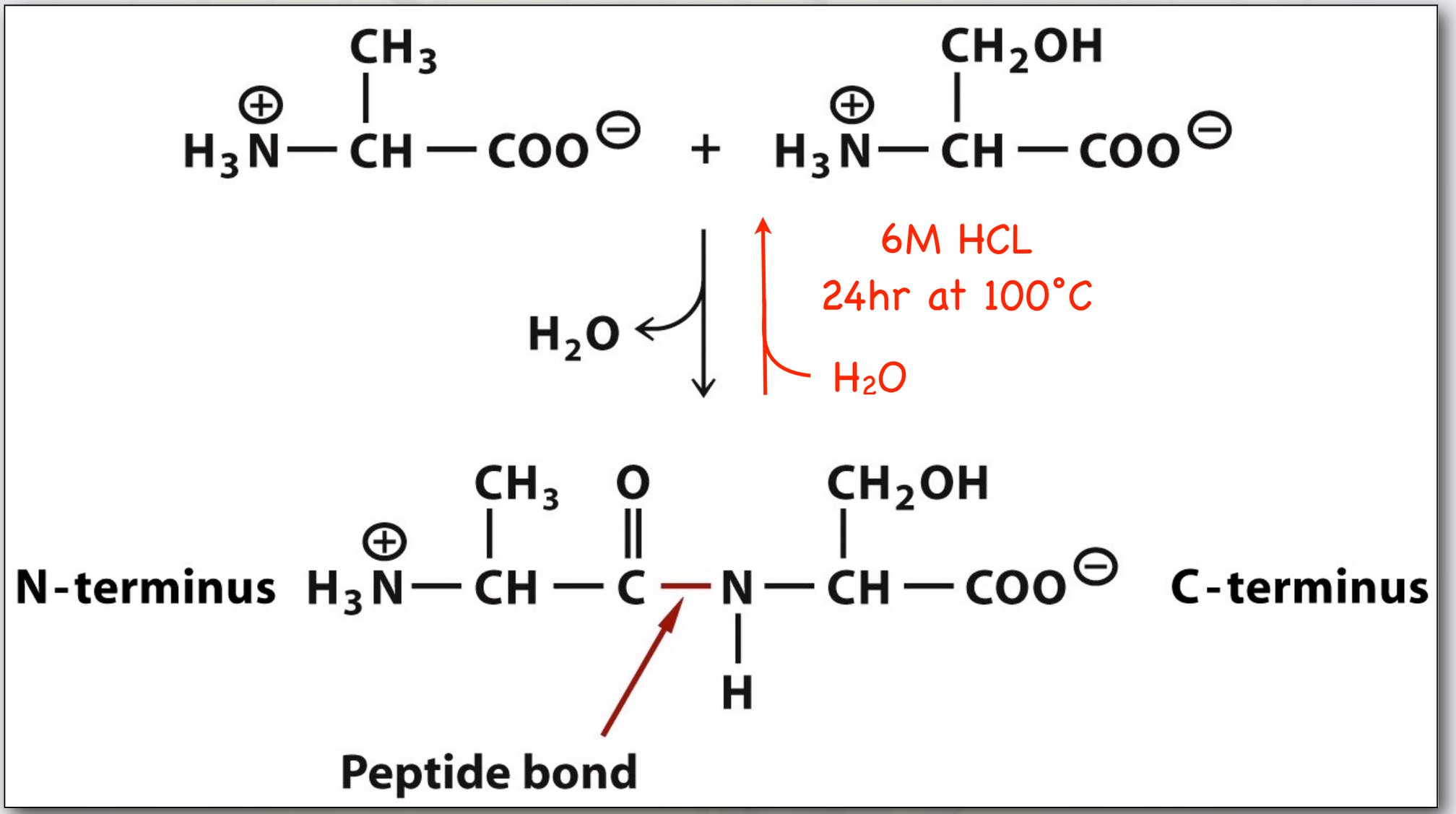


Protein Primary Structure



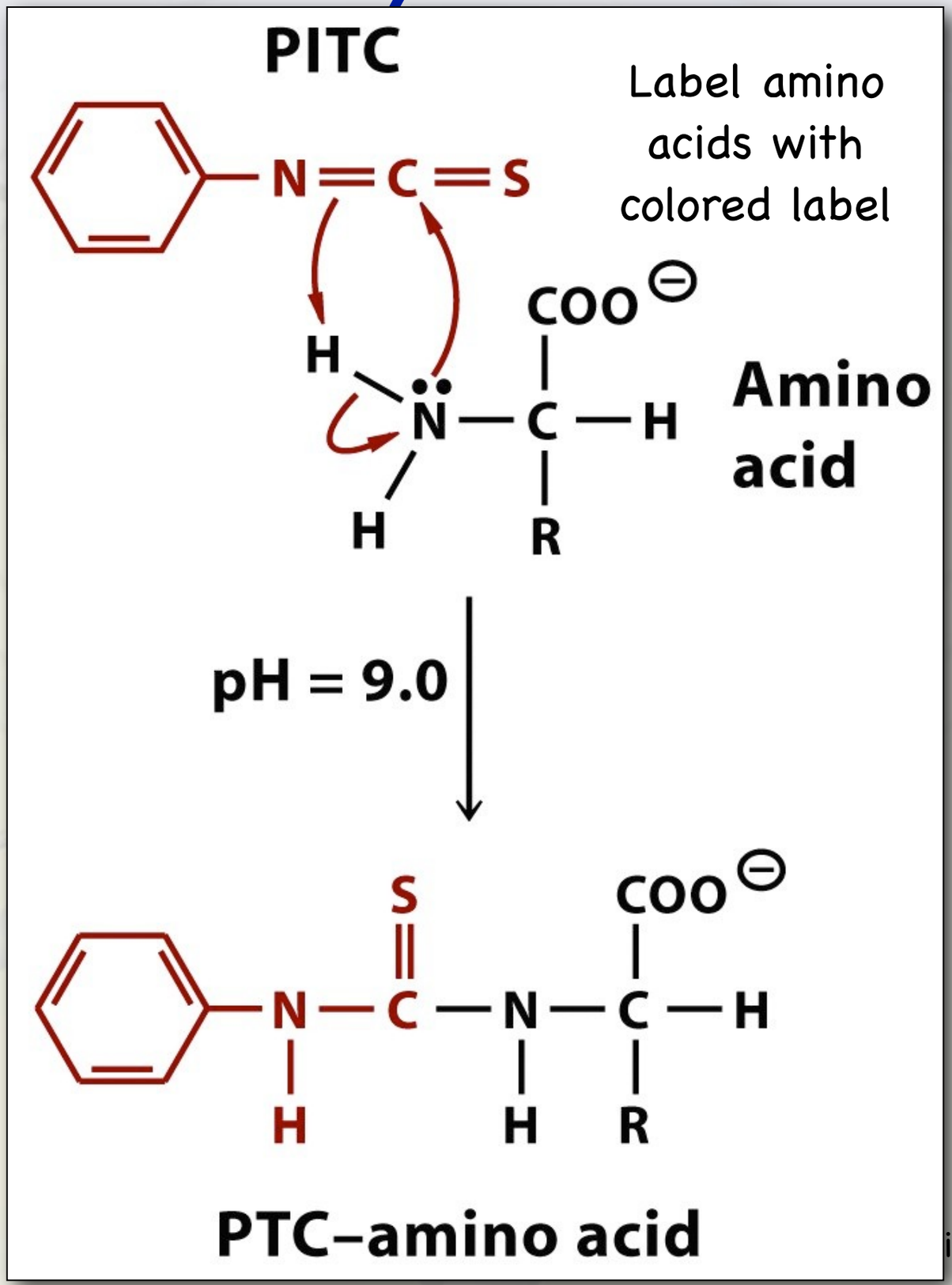
Protein Primary Structure

- Amino acid composition



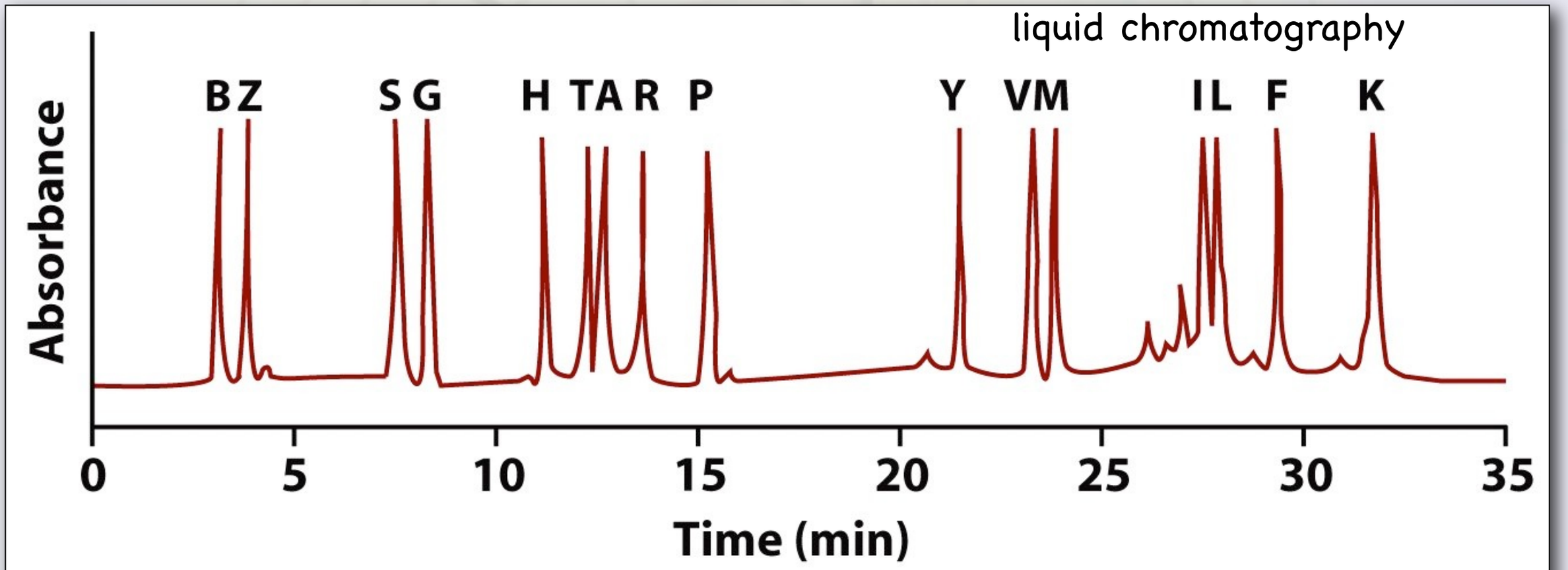
Protein Primary Structure

- Amino



Protein Primary Structure

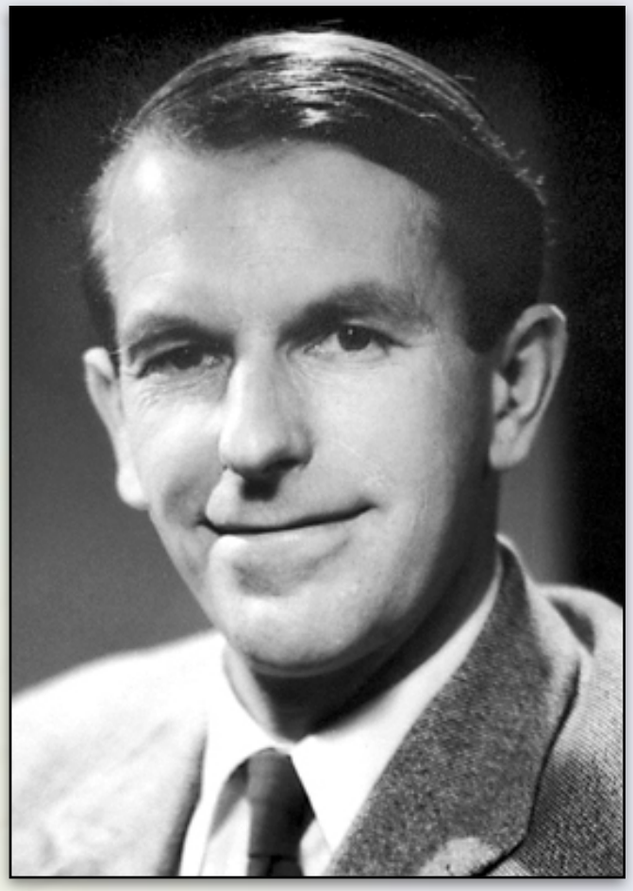
- Amino acid composition



Protein Primary Structure

Fredrick Sanger was the first person to sequence a complete protein

- Insulin (1953)



1956 Nobel Prize in Chemistry

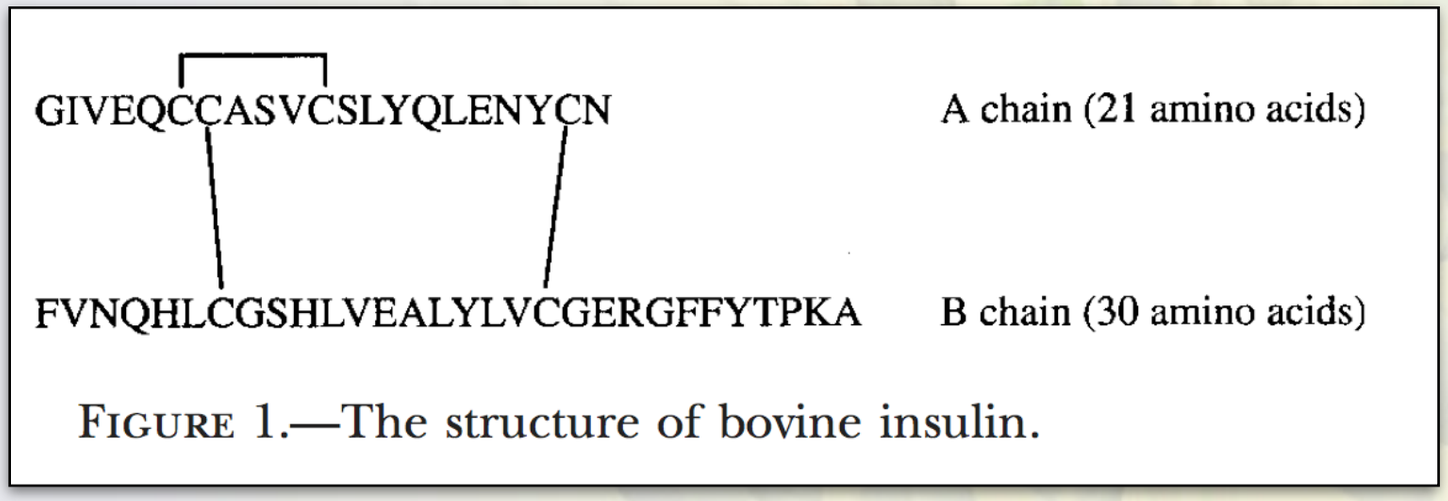
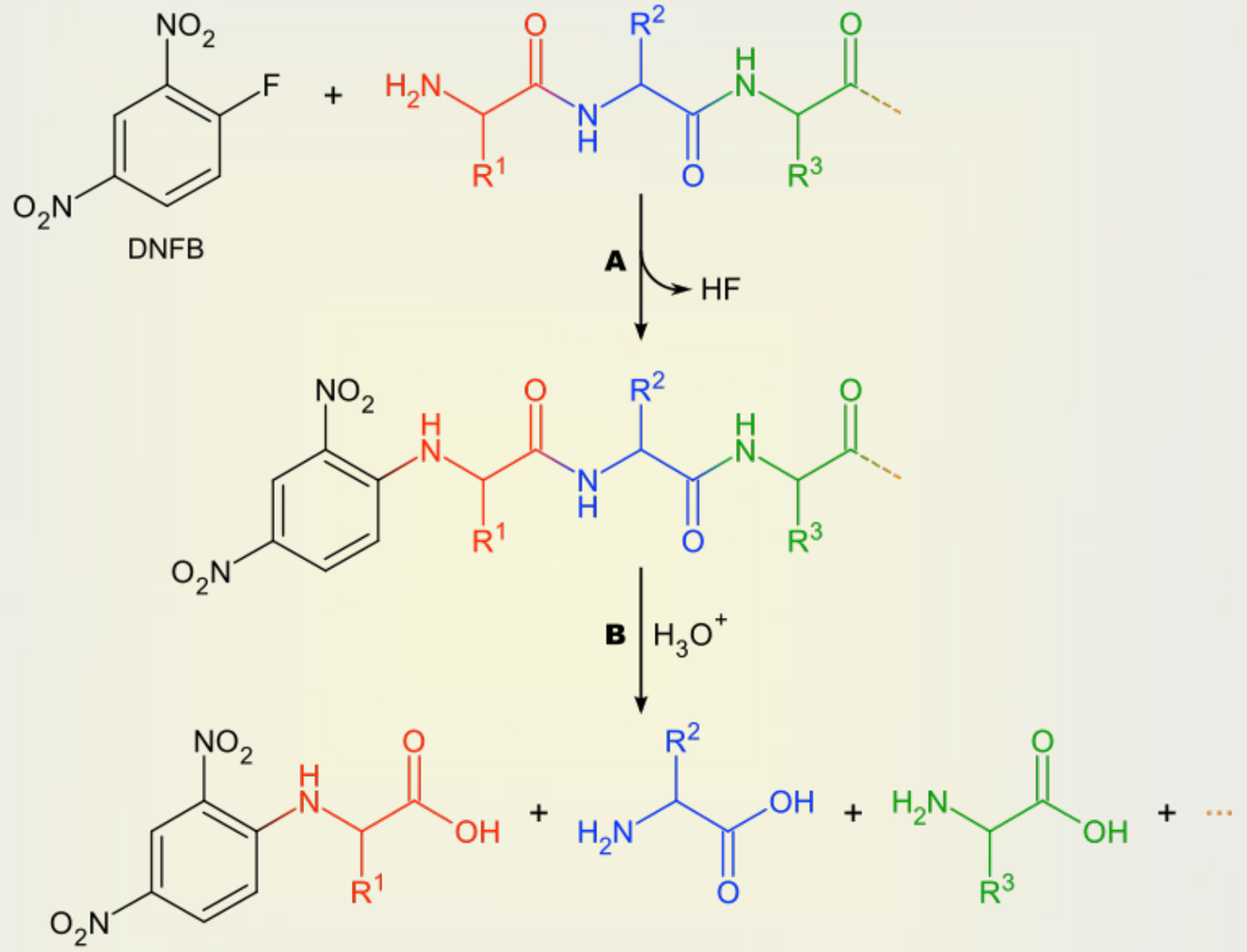


FIGURE 1.—The structure of bovine insulin.

Perspectives on Genetics: Antony Stretton, "The First Sequence: Fred Sanger and Insulin", Genetics 2002, 162, 527-532.

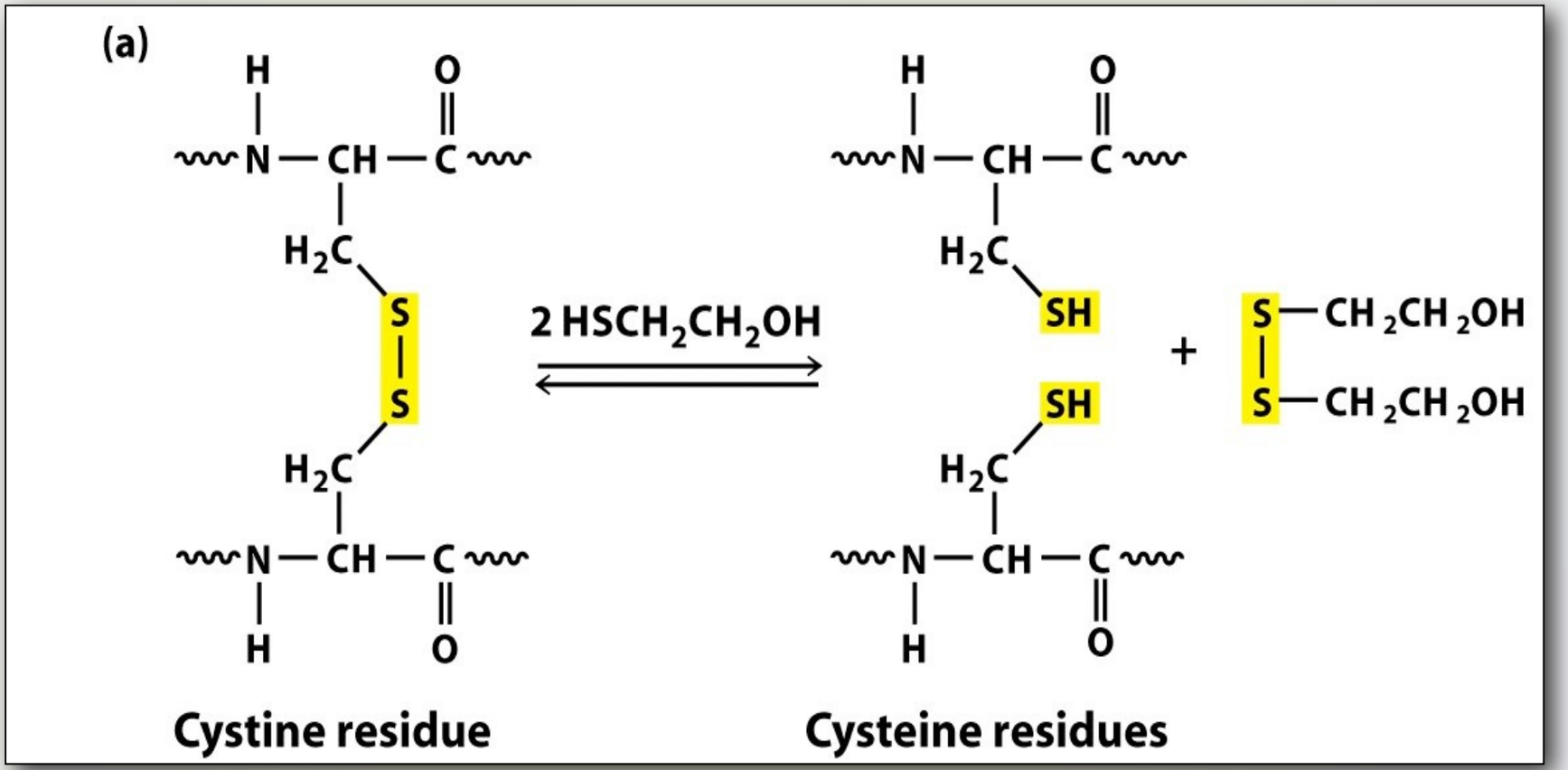
Protein Primary Structure

- Sanger's Reagent
  - (DNFB, 2,4-dinitrofluorobenzene)



Protein Primary Structure

- Cleavage of disulfide bonds with β-mercaptoenthanol.

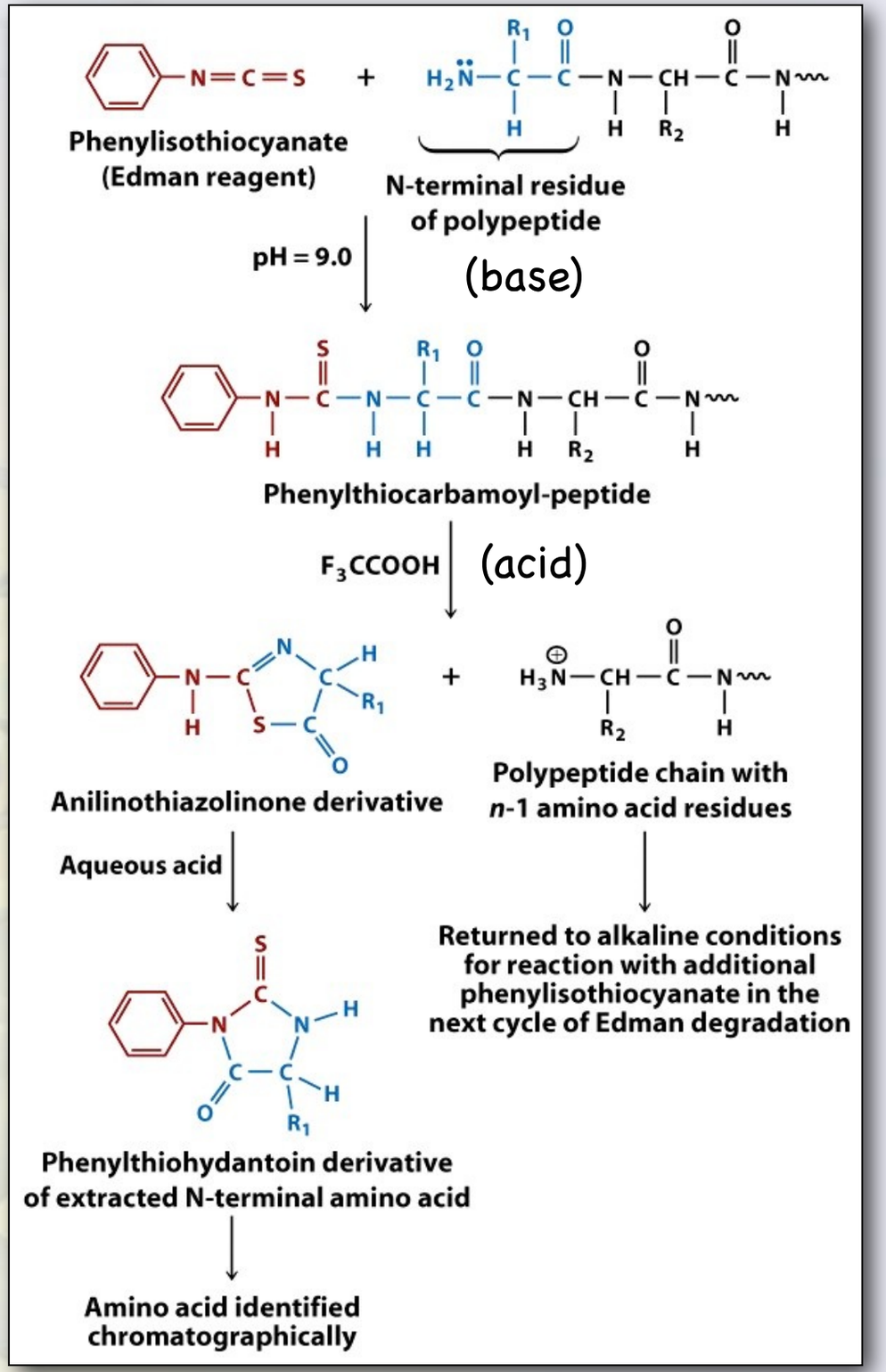


Protein Primary Structure

- Amino acid sequence
  - Edman Degradation

Protein Primary Structure

- Amino acid sequence
  - Edman Degradation



Protein Primary Structure

Amino acid sequence

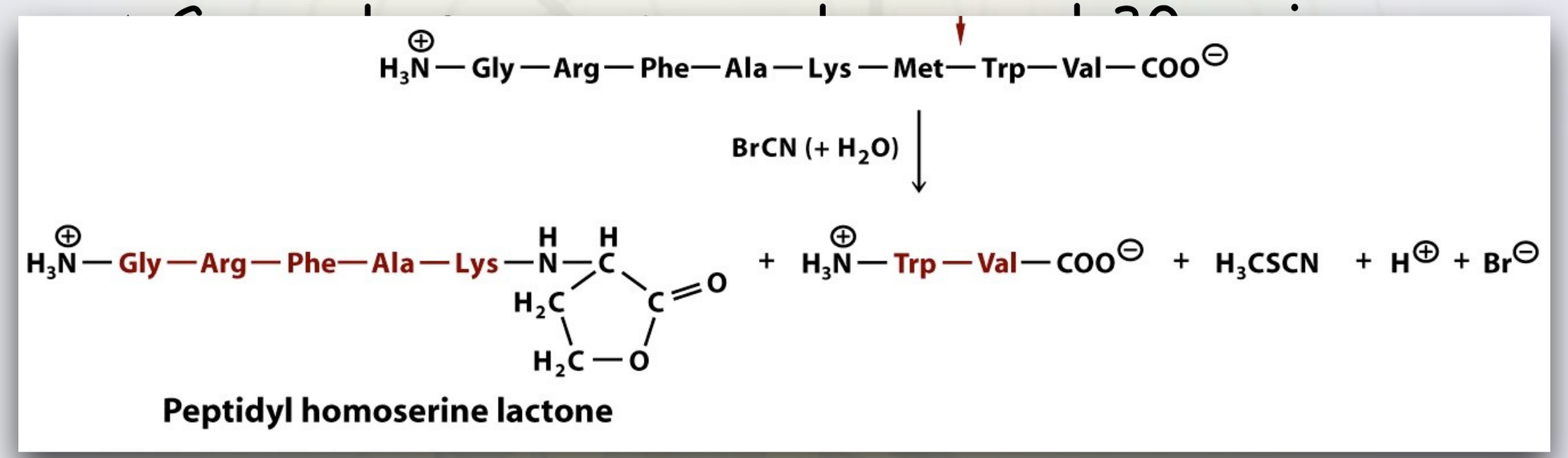
- Edman Degradation
- Can only sequence up to around 30 amino acids at a time, therefore, for larger proteins, the polypeptide is cleaved into smaller segments.

- **CNBr** (cyanogen bromide) cleaves at Met
- **Trypsin** protease cleaves at Lys & Arg (+)
- **Chymotrypsin** cleaves at Phe, Tyr & Trp (aromatic)

Protein Primary Structure

Amino acid sequence

- Edman Degradation



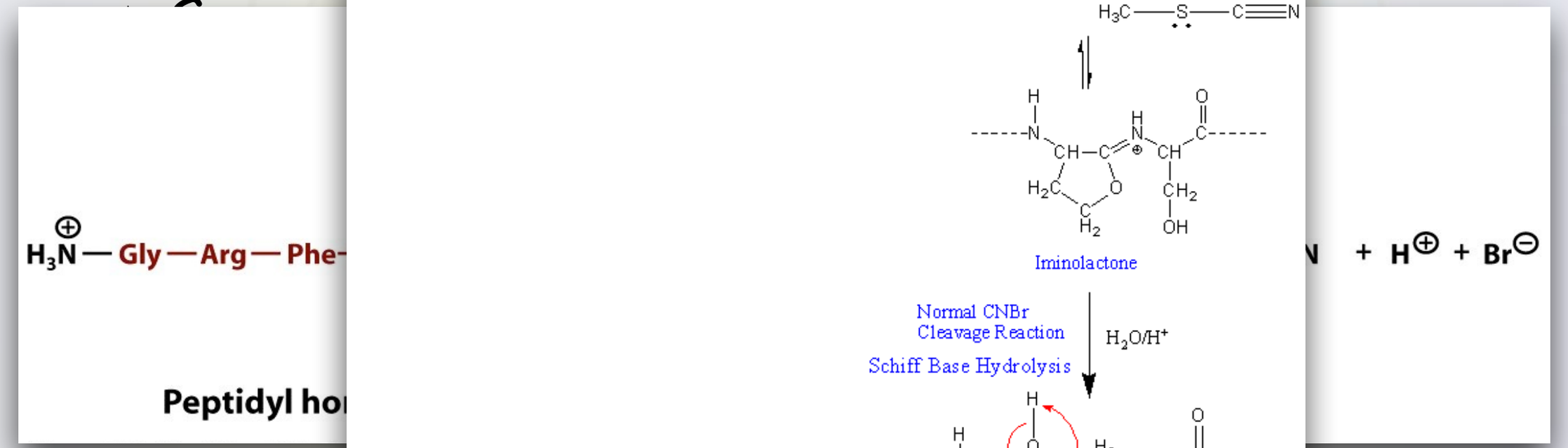
- **Trypsin** protease cleaves at Lys & Arg (+)
- **Chymotrypsin** cleaves at Phe, Tyr & Trp (aromatic)

Protein

Amino a

- Edman

- Tryp
- Chyr (aror



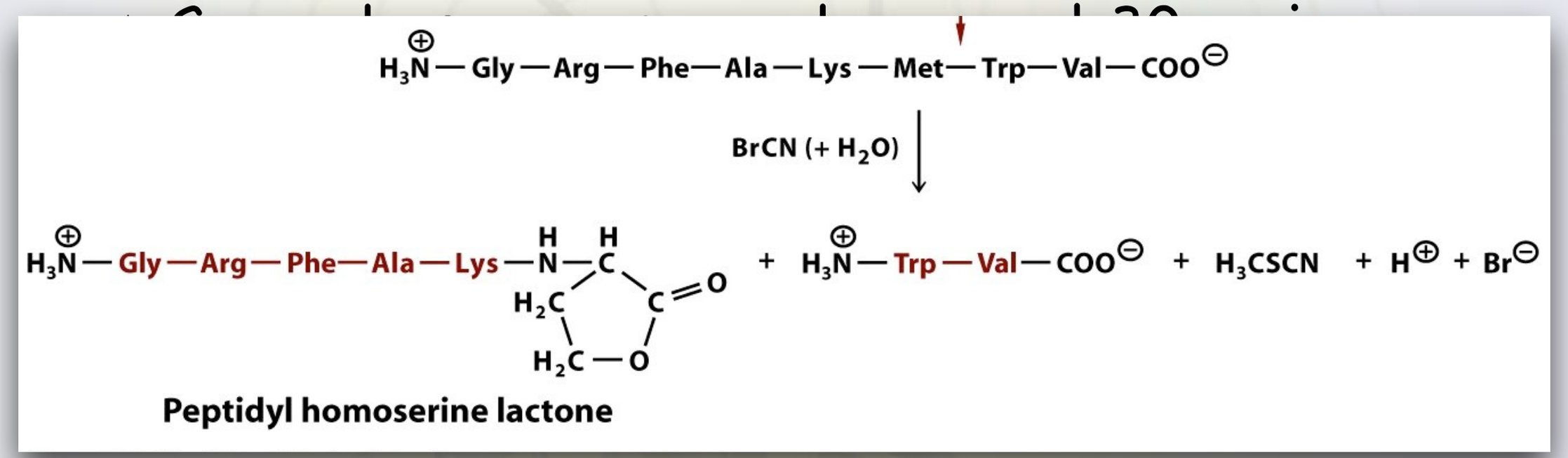
(+)

rp

Protein Primary Structure

Amino acid sequence

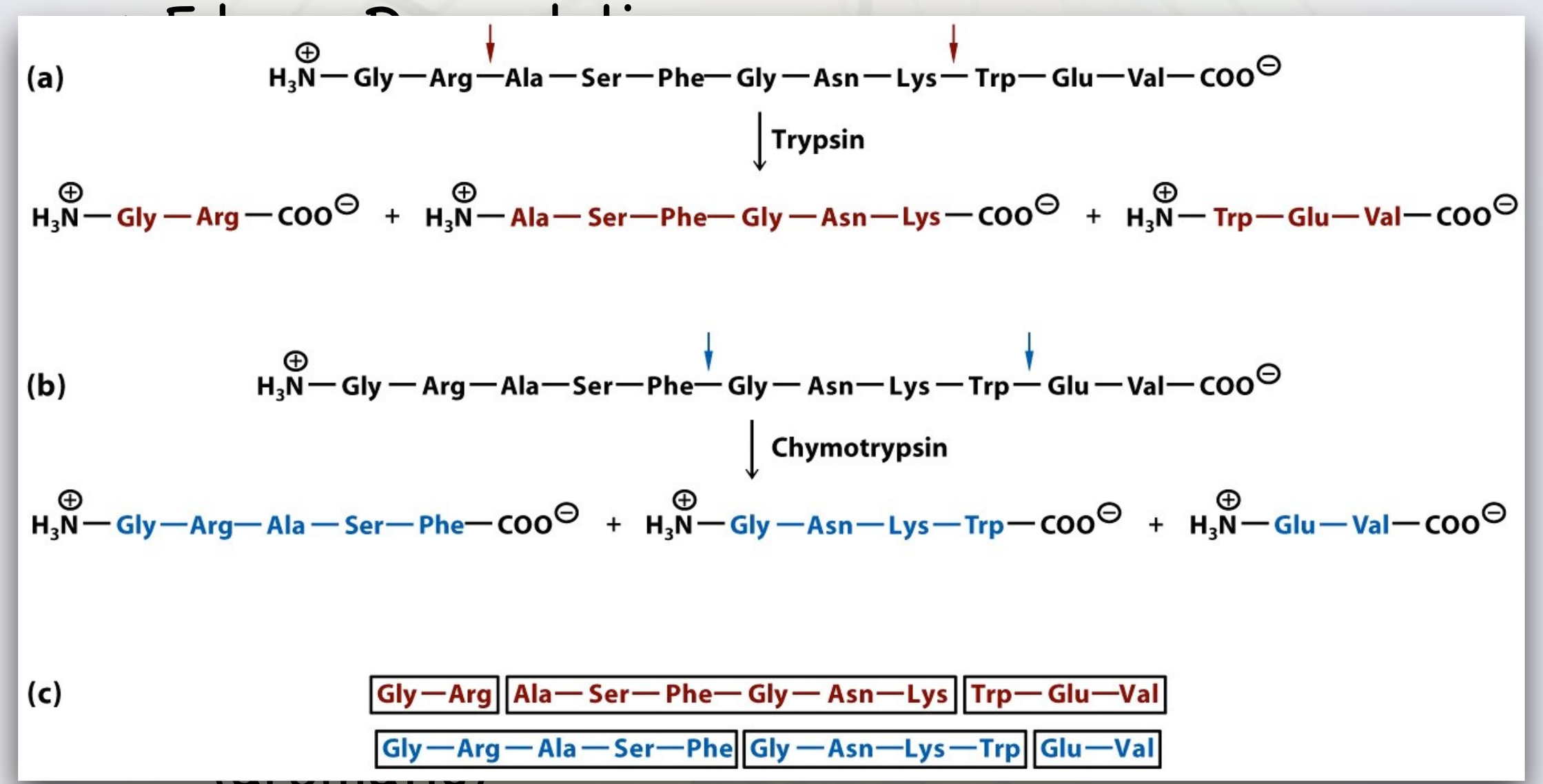
- Edman Degradation



- **Trypsin** protease cleaves at Lys & Arg (+)
- **Chymotrypsin** cleaves at Phe, Tyr & Trp (aromatic)

Protein Primary Structure

Amino acid sequence



Protein Primary Structure

Question:

You have isolated a decapeptide called FP, which has anticancer activity. Determine the sequence of the peptide using the following information:

a. One cycle of Edman degradation of intact FP yields 2 mol of PTH-aspartate per mole of FP.

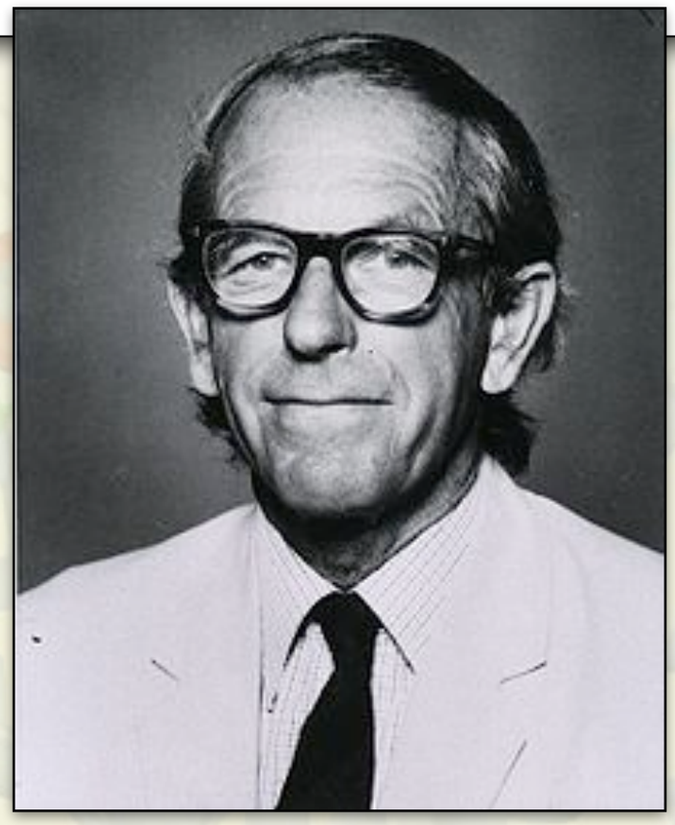
b. Treatment of a solution of FP with 2-mercaptoethanol followed by addition of trypsin yields three peptides with the composition (Ala, Cys,Phe), (Arg,Asp), and (Asp,Cys,Gly,Met,Phe). The intact (Ala,Cys, Phe) peptide yields PTH-cysteine in the first cycle of Edman degradation.

c. Treatment of 1 mol of FP with carboxypeptidase (which cleave the C-terminal residue from peptides) yields 2 mol of phenylalanine.

d. Treatment of the intact pentapeptide (Asp,Cys,Gly,Met,Phe) with CNBr yields two peptides with the composition (homoserine lactone, Asp) and (Cys,Gly,Phe). The (Cys,Gly,Phe) peptide yields PTH-glycine in the first cycle of Edman degradation.

Protein Primary Structure

Protein Primary Structure



1956 Nobel Prize in Chemistry

1980 Nobel Prize in Chemistry

Protein Primary Structure

- Amino acid sequence
  - Reverse translating a DNA sequence

Protein Primary Structure

- Amino acid sequence
  - Reverse translating a DNA sequence

DNA ~~~ **T G T C** ~~~

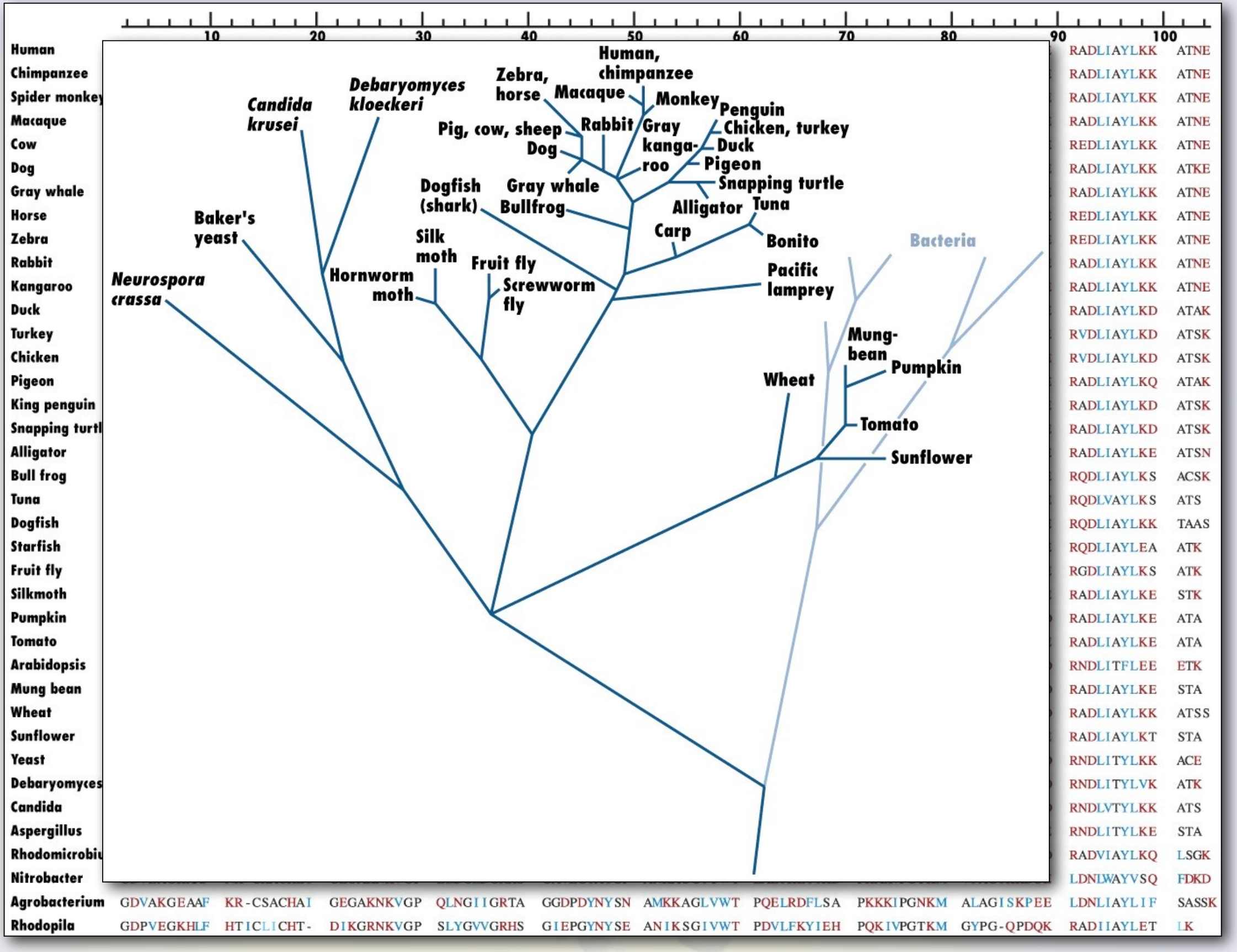
Protein ~~~ — **Val** ~~~

	U	C	A	G	
U	Phe	Ser	Tyr	Cys	**U**
	Phe	Ser	Tyr	Cys	**C**
	Leu	Ser	STOP	STOP	**A**
	Leu	Ser	STOP	Trp	**G**
C	Leu	Pro	His	Arg	**U**
	Leu	Pro	His	Arg	**C**
	Leu	Pro	Gln	Arg	**A**
	Leu	Pro	Gln	Arg	**G**
A	Ile	Thr	Asn	Ser	**U**
	Ile	Thr	Asn	Ser	**C**
	Ile	Thr	Lys	Arg	**A**
	Met	Thr	Lys	Arg	**G**
G	Val	Ala	Asp	Gly	**U**
	Val	Ala	Asp	Gly	**C**
	Val	Ala	Glu	Gly	**A**
	Val	Ala	Glu	Gly	**G**

Protein Primary Structure

- Amino acid sequence
  - Reverse translating a DNA sequence

	10	20	30	40	50	60	70	80	90	100
Human	GDVEKGKKIF	IMKCSQCHTV	EKGGKHKTGP	NLHGLFGRKT	GQAPGYSYTA	ANKNKGIIWG	EDTLMEYLEN	PKKYIPGTKM	IFVGIKKKEE	RADLIAYLKK	ATNE
Chimpanzee	GDVEKGKKIF	IMKCSQCHTV	EKGGKHKTGP	NLHGLFGRKT	GQAPGYSYTA	ANKNKGIIWG	EDTLMEYLEN	PKKYIPGTKM	IFVGIKKKEE	RADLIAYLKK	ATNE
Spider monkey	GDVFKGKRIF	IMKCSQCHTV	EKGGKHKTGP	NLHGLFGRKT	GQASGFTYTE	ANKNKGIIWG	EDTLMEYLEN	PKKYIPGTKM	IFVGIKKKEE	RADLIAYLKK	ATNE
Macaque	GDVEKGKKIF	IMKCSQCHTV	EKGGKHKTGP	NLHGLFGRKT	GQAPGYSYTA	ANKNKGITWG	EDTLMEYLEN	PKKYIPGTKM	IFVGIKKKEE	RADLIAYLKK	ATNE
Cow	GDVEKGKKIF	VQKCAQCHTV	EKGGKHKTGP	NLHGLFGRKT	GQAPGFSYTD	ANKNKGITWG	EETLMEYLEN	PKKYIPGTKM	IFAGIKKKGE	REDLIAYLKK	ATNE
Dog	GDVEKGKKIF	VQKCAQCHTV	EKGGKHKTGP	NLHGLFGRKT	GQAPGFSYTD	ANKNKGITWG	EETLMEYLEN	PKKYIPGTKM	IFAGIKKTGE	RADLIAYLKK	ATKE
Gray whale	GDVEKGKKIF	VQKCAQCHTV	EKGGKHKTGP	NLHGLFGRKT	GQAVGFSYTD	ANKNKGITWG	EETLMEYLEN	PKKYIPGTKM	IFAGIKKKGE	RADLIAYLKK	ATNE
Horse	GDVEKGKKIF	VQKCAQCHTV	EKGGKHKTGP	NLHGLFGRKT	GQAPGFTYTD	ANKNKGITWK	EETLMEYLEN	PKKYIPGTKM	IFAGIKKKTE	REDLIAYLKK	ATNE
Zebra	GDVEKGKKIF	VQKCAQCHTV	EKGGKHKTGP	NLHGLFGRKT	GQAPGFSYTD	ANKNKGITWK	EETLMEYLEN	PKKYIPGTKM	IFAGIKKKTE	REDLIAYLKK	ATNE
Rabbit	GDVEKGKKIF	VQKCAQCHTV	EKGGKHKTGP	NLHGLFGRKT	GQAVGFSYTD	ANKNKGITWG	EDTLMEYLEN	PKKYIPGTKM	IFAGIKKKDE	RADLIAYLKK	ATNE
Kangaroo	GDVEKGKKIF	VQKCAQCHTV	EKGGKHKTGP	NLHGIFGRKT	GQAPGFTYTD	ANKNKGIIWG	EDTLMEYLEN	PKKYIPGTKM	IFAGIKKKGE	RADLIAYLKK	ATNE
Duck	GDVEKGKKIF	VQKCSQCHTV	EKGGKHKTGP	NLHGLFGRKT	GQAEGFSYTD	ANKNKGITWG	EDTLMEYLEN	PKKYIPGTKM	IFAGIKKKSE	RADLIAYLKD	ATAK
Turkey	GDIEKGKKIF	VQKCSQCHTV	EKGGKHKTGP	NLHGLFGRKT	GQAEGFSYTD	ANKNKGITWG	EDTLMEYLEN	PKKYIPGTKM	IFAGIKKKSE	RVDLIAYLKD	ATSK
Chicken	GDIEKGKKIF	VQKCSQCHTV	EKGGKHKTGP	NLHGLFGRKT	GQAEGFSYTD	ANKNKGITWG	EDTLMEYLEN	PKKYIPGTKM	IFAGIKKKSE	RVDLIAYLKD	ATSK
Pigeon	GDIEKGKKIF	VQKCSQCHTV	EKGGKHKTGP	NLHGLFGRKT	GQAEGFSYTD	ANKNKGITWG	EDTLMEYLEN	PKKYIPGTKM	IFAGIKKKAE	RADLIAYLKQ	ATAK
King penguin	GDIEKGKKIF	VQKCSQCHTV	EKGGKHKTGP	NLHGIFGRKT	GQAEGFSYTD	ANKNKGITWG	EDTLMEYLEN	PKKYIPGTKM	IFAGIKKKSE	RADLIAYLKD	ATSK
Snapping turtle	GDVEKGKKIF	VQKCAQCHTV	EKGGKHKTGP	NLNGLIGRKT	GQAEGFSYTE	ANKNKGITWG	EETLMEYLEN	PKKYIPGTKM	IFAGIKKKAE	RADLIAYLKD	ATSK
Alligator	GDVEKGKKIF	VQKCAQCHTV	EKGGKHKTGP	NLHGLIGRKT	GQAPGFSYTE	ANKNKGITWG	EETLMEYLEN	PKKYIPGTKM	IFAGIKKKPE	RADLIAYLKE	ATSN
Bull frog	GDVEKGKKIF	VQKCAQCHTV	EKGGKHKVGP	NLYGLIGRKT	GQAAGFSYTD	ANKNKGITWG	EDTLMEYLEN	PKKYIPGTKM	IFAGIKKKGE	RQDLIAYLKS	ACSK
Tuna	GDVAKGKKTF	VQKCAQCHTV	ENGGKHKVGP	NLWGLFGRKT	GQAEGYSYTD	ANKSKGIVWN	ENTLMEYLEN	PKKYIPGTKM	IFAGIKKKGE	RQDLVAYLKS	ATS
Dogfish	GDVEKGKKVF	VQKCAQCHTV	ENGGKHKTGP	NLSGLFGRKT	GQAQGFSYTD	ANKSKGITWQ	QETLRIYLEN	PKKYIPGTKM	IFAGIKKKSE	RQDLIAYLKK	TAAS
Starfish	GQVEKGKKIF	VQRCAQCHTV	EKAGKHKTGP	NLNGILGRKT	GQAAGFSYTD	ANRNKGITWK	NETLFEYLEN	PKKYIPGTKM	VFAGLKKQKE	RQDLIAYLEA	ATK
Fruit fly	GDVEKGKKLF	VQRCAQCHTV	EAGGKHKVGP	NLHGLIGRKT	GQAAGFAYTD	ANKAKGITWN	EDTLFEYLEN	PKKYIPGTKM	IFAGLKKPNE	RGDLIAYLKS	ATK
Silkmoth	GNAENGKKIF	VQRCAQCHTV	EAGGKHKVGP	NLHGFYGRKT	GQAPGFSYSN	ANKAKGITWG	DDTLFEYLEN	PKKYIPGTKM	VFAGLKKANE	RADLIAYLKE	STK
Pumpkin	GNSKAGEKIF	KTKCAQCHTV	DKGAGHKQGP	NLNGLFGRQS	GTTPGYSYSA	ANKNRAVIWE	EKTLYDYLLN	PKKYIPGTKM	VFPGLKKPQD	RADLIAYLKE	ATA
Tomato	GNPKAGEKIF	KTKCAQCHTV	EKGAGHKEGP	NLNGLFGRQS	GTTAGYSYSA	ANKNMAVNWG	ENTLYDYLLN	PKKYIPGTKM	VFPGLKKPQE	RADLIAYLKE	ATA
Arabidopsis	GDAKKGANLF	KTRCAQCHTL	KAGEGNKIGP	ELHGLFGRKT	GSVAGYSYTD	ANKQKGIEWK	DDTLFEYLEN	PKKYIPGTKM	AFGGLKKPKD	RNDLITFLEE	ETK
Mung bean	GNSKSGEKIF	KTKCAQCHTV	DKGAGHKQGP	NLNGLIGRQS	GTTAGYSYST	ANKNMAVIWE	EKTLYDYLLN	PKKYIPGTKM	VFPGLKKPQD	RADLIAYLKE	STA
Wheat	GNPDAGAKIF	KTKCAQCHTV	DAGAGHKQGP	NLHGLFGRQS	GTTAGYSYSA	ANKNKAVEWE	ENTLYDYLLN	PKKYIPGTKM	VFPGLKKPQD	RADLIAYLKK	ATSS
Sunflower	GNPTTGEKIF	KTKCAQCHTV	EKGAGHKQGP	NLNGLFGRQS	GTTAGYSYSA	GNKNKAVIWE	ENTLYDYLLN	PKKYIPGTKM	VFPGLKKPQE	RADLIAYLKT	STA
Yeast	GSAKKGATLF	KTRCLQCHTV	EKGGPHKVGP	NLHGIFGRHS	GQAEGYSYTD	ANIKKNVLWD	ENNMSEYLTN	PKKYIPGTKM	AFGGLKKEKD	RNDLITYLKK	ACE
Debaryomyces	GSEKKGANLF	KTRCLQCHTV	EKGGPHKVGP	NLHGVVGRTS	GQAQGFSYTD	ANKKKGVEWT	EQDLSDYLEN	PKKYIPGTKM	AFGGLKKAKD	RNDLITYLVK	ATK
Candida	GSEKKGATLF	KTRCLQCHTV	EKGGPHKVGP	NLHGVFGRKS	GLAEGYSYTD	ANKKKGVEWT	EQTMSDYLEN	PKKYIPGTKM	AFGGLKKPKD	RNDLVTYLKK	ATS
Aspergillus	GDAK-GAKLF	QTRCAQCHTV	EAGGPHKVGP	NLHGLFGRKT	GQSEGYAYTD	ANKQAGVTWD	ENTLFSYLEN	PKKFIPGTKM	AFGGLKKGKE	RNDLITYLKE	STA
Rhodomicrobium	GDPVKGEQVF	KQ-CKICHQV	GPTAKNGVGP	EQNDVFGQKA	GARPGFNYSD	AMKNSGLTWD	EATLDKYLEN	PKAVVPGTKM	VFVGLKNPQD	RADVIAYLKQ	LSGK
Nitrobacter	GDVEAGKAAF	NK-CKACHEI	GESAKNKVGP	ELDGLDGRHS	GAVEGYAYSP	ANKASGITWT	EAEFKEYIKD	PKAKVPGTKM	VFAGIKKDSE	LDNLWAYVSQ	FDKD
Agrobacterium	GDVAKGEAAF	KR-CSACHAI	GEGAKNKVGP	QLNGIIGRTA	GGDPDYNYSN	AMKKAGLVWT	PQELRDFLSA	PKKKIPGNKM	ALAGISKPEE	LDNLIAYLIF	SASSK
Rhodopila	GDPVEGKHLF	HTICLICHT-	DIKGRNKVGP	SLYGVVGRHS	GIEPGYNYSE	ANIKSGIVWT	PDVLFKYIEH	PQKIVPGTKM	GYPG-QPDQK	RADIIAYLET	LK



Cytochrome c sequence

Protein Primary Structure

Question:

The functional diversity of proteins results from the large number of possible polypeptides that can be built using the 20 different amino acids

Question: What is the minimum mass it would take to construct one molecule each of all of the possible polypeptides that contain 100 amino acids?

Protein Primary Structure

Question:

The functional diversity of proteins results from the large number of possible polypeptides that can be built using the 20 different amino acids

Question: What is the minimum mass it would take to construct one molecule each of all of the possible polypeptides that contain 100 amino acids?

o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o

Protein Primary Structure

Question:

Number of polypeptides (20^{100})	1.26×10^{130}
Avg. Mass of each polypeptide	1.83×10^{-22} g
Total mass needed	2.32×10^{108} g
Number of Earths	3.9×10^{80}
Number of Suns	1.2×10^{75}
Number of Galaxies	9.7×10^{29}

Protein Primary Structure

Question:

The functional diversity of proteins results from the large number of possible polypeptides that can be built using the 20 different amino acids

Question: What is the minimum mass it would take to construct one molecule each of all of the possible polypeptides that contain 100 amino acids?

o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o

Next up

Lecture 3, Part II - Protein Structure and Function

- Read Chapter 4 of Moran et al.