Amino Acids and Proteins
Amino Acid
> Compound with amino functional group and carboxylic acid functional group
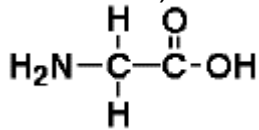
$$H_2N-\overset{|}{\underset{|}{C}}_n-\overset{O}{\overset{||}{C}}-OH$$
Amine   Carboxylic Acid
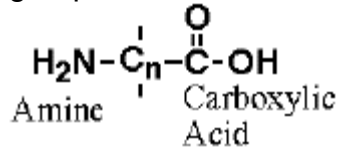
α-amino acids amine and carboxyl separated by 1 carbon atom
β-amino acids amine and carboxyl separated by 2 carbon atoms...
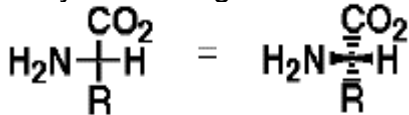

Protein-Derived Amino Acids
> Protein - biological polymer of a-amino acids
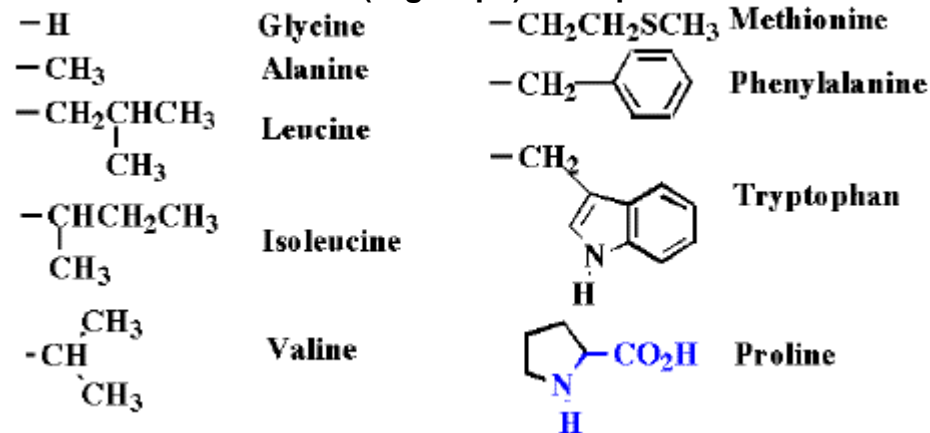> Simplest protein derived amino acid (aa) - glycine (only achiral protein-derived aa)

$$H_2N-\overset{H}{\underset{H}{C}}-\overset{O}{\overset{||}{C}}-OH$$

Chirality of amino acids
> Only one configuration found naturally

$$H_2N\overset{CO_2}{\underset{R}{+}}H \quad = \quad H_2N\overset{CO_2}{\underset{\bar{R}}{=}}H$$

- According to R/S system - S
- According to D/L system - L
  1. put most oxidized carbon at top
  2. continue carbon chain to bottom
  3. non-hydrogen substituent on left = L

Natural AA Sidechains (R groups): Non-polar

$-H$	Glycine	$-CH_2CH_2SCH_3$	Methionine
$-CH_3$	Alanine	$-CH_2-\langle\text{phenyl}\rangle$	Phenylalanine
$-CH_2\overset{|}{C}HCH_3$ (CH_3)	Leucine		
$-\overset{|}{C}HCH_2CH_3$ (CH_3)	Isoleucine	$-CH_2-\langle\text{indole}\rangle$	Tryptophan
$-CH$ (CH_3)(CH_3)	Valine	$\langle\text{pyrrolidine}\rangle-CO_2H$	Proline

Natural AA Sidechains (R groups): Polar, neutral

$-CH_2OH$ Serine

$-\underset{\underset{OH}{|}}{CH}CH_3$ Threonine

$-CH_2\underset{\underset{O}{\|}}{C}NH_2$ Asparagine

$-CH_2CH_2\underset{\underset{O}{\|}}{C}NH_2$ Glutamine

Natural AA Sidechains (R groups): Polar, acidic

$-CH_2\underset{\underset{O}{\|}}{C}OH$ **Aspartic Acid**

$-CH_2CH_2\underset{\underset{O}{\|}}{C}OH$ **Glutamic Acid**

$-CH_2SH$ **Cysteine**

$-CH_2-\bigcirc-OH$ **Tyrosine**

Natural AA Sidechains (R groups): Polar, basic

$-CH_2CH_2CH_2NH\underset{\underset{NH}{\|}}{C}NH_2$ **Arginine**

$-CH_2-\underset{\underset{H}{|}}{N}\overset{N}{\diagup}$ **Histidine**

$-CH_2CH_2CH_2CH_2NH_2$ **Lysine**

Acid-Base Properties of Amino Acids

Amino Acid pKa a-CO₂H pKa b-CO₂H pKa a-NH₃⁺

Amino Acid pKa a-CO_2H pKa b-CO_2H pKa a-NH_3^+

Aspartic Acid 2.10   3.86   9.82

Form: A   B   C   D

$$^+H_3N{-}H \rightleftharpoons {}^+H_3N{-}H \rightleftharpoons {}^+H_3N{-}H \rightleftharpoons H_2N{-}H$$

Very Low pH   Very High pH
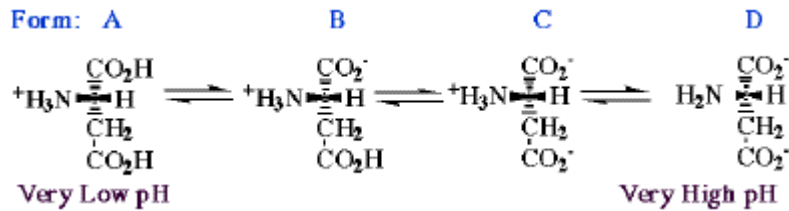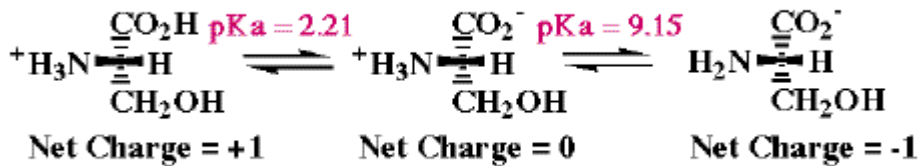

Challenge:

Tyrosine has an a-CO2H pKa of 2.20, and a-NH3+ pKa of 9.11, and a sidechain pKa of 10.07 - sketch a titration curve and give the structure of the major form or forms present at all equivalence and half-equivalence points.
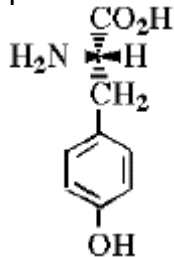
Isoelectric Point

Isoelectric point (pI) - pH at which molecule has no net charge (find by averaging pkaÕs on either side of neutral form)

Serine

$$^+H_3N{-}H \xrightleftharpoons{pKa=2.21} {}^+H_3N{-}H \xrightleftharpoons{pKa=9.15} H_2N{-}H$$

Net Charge = +1   Net Charge = 0   Net Charge = -1
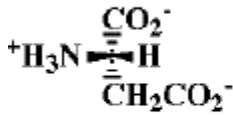
$$pI = (2.21 + 9.15)/2 = 5.68$$

Isoelectric Precipitation

Amino acids with net charges repel each other and do not easily crystallize from solution - solubility is lowest at the isoelectric point

---

Electrophoresis
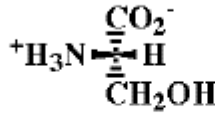A method for separating charged compounds in an electric field

Aspartic Acid	Serine	Lysine
pI=2.98	pI=5.68	pI=9.74

$$^+H_3N \text{—} \overset{CO_2^-}{\underset{CH_2CO_2^-}{\overset{|}{\underset{|}{C}}}} \text{—} H \qquad ^+H_3N \text{—} \overset{CO_2^-}{\underset{CH_2OH}{\overset{|}{\underset{|}{C}}}} \text{—} H \qquad ^+H_3N \text{—} \overset{CO_2^-}{\underset{CH_2CH_2CH_2CH_2NH_3^-}{\overset{|}{\underset{|}{C}}}} \text{—} H$$

Net Charge = ~ -1	Net Charge = ~0	Net Charge = ~ +1
at pH = 7.0	at pH = 7.0	at pH = 7.0
migrates to + electrode	little migration	migrates to – electrode

---

Nutrition
Dietary protein provides aaÕs not made in our systems - these are called essential

Essential Amino Acids Nonessential Amino Acids

Essential Amino Acids	Nonessential Amino Acids
arginine	glycine
histidine	alanine
lysine	cysteine
isoleucine	aspartic acid
leucine	glutamic acid
valine	asparagine
threonine	glutamine
methionine	proline
phenylalaninetyrosine	
tryptophan	serine

Polypeptides and Proteins
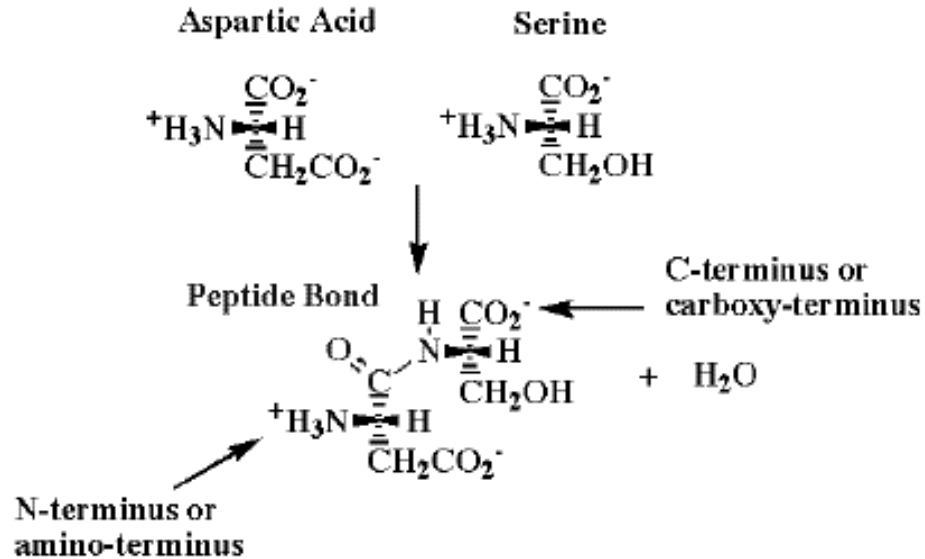**Proteins and polypeptides are chains of aa's**
Dipeptide - 2 amino acids
Tripeptide - 3 amino acids ...
Polypeptides have 10 or more amino acids
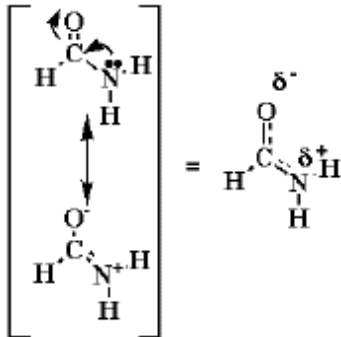Proteins have 100s or more amino acids

---

Amino Acid Connections

Aspartic Acid   Serine

CO_2^-
^+H_3N─┤H
$CH_2CO_2^-$

CO_2^-
^+H_3N─┤H
CH_2OH

C-terminus or
carboxy-terminus

Peptide Bond

H CO_2^-
O═C─N─┤H
CH_2OH + H_2O

^+H_3N─┤H
$CH_2CO_2^-$

N-terminus or
amino-terminus

Peptide Bond Geometry
Peptide bonds are planar due to conjugation of nitrogen lone pair with carbonyl

---

Protein Structural Descriptions
Primary Structure
sequence of amino acids
Secondary Structure
local three-dimensional patterns
Tertiary Structure
three-dimensional shape of an amino acid chain
Quaternary Structure
three-dimensional shape of multiple amino acid chains

Primary Structure Determination

Step 1: Amino Acid Analysis
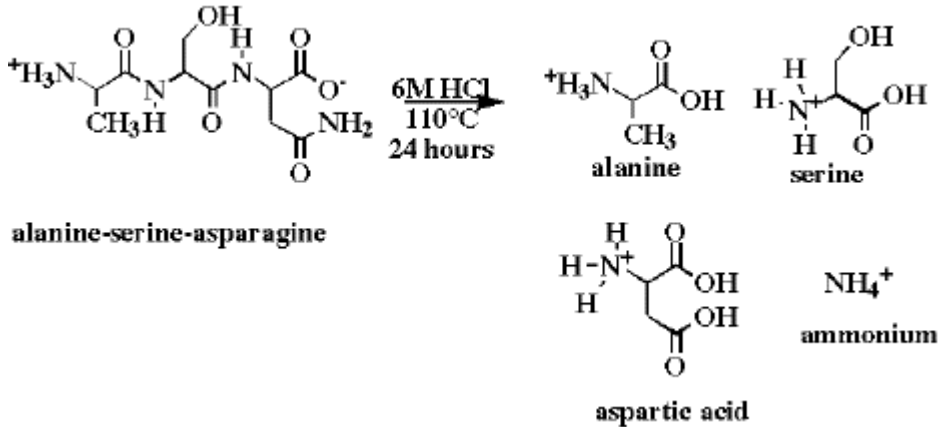Protein is hydrolyzed in strong acid or base at high temperatures
Mixture of amino acids generated by protein hydrolysis is analyzed by chromatography
Step 2: Sequence Analysis
protein cleaved into polypeptides
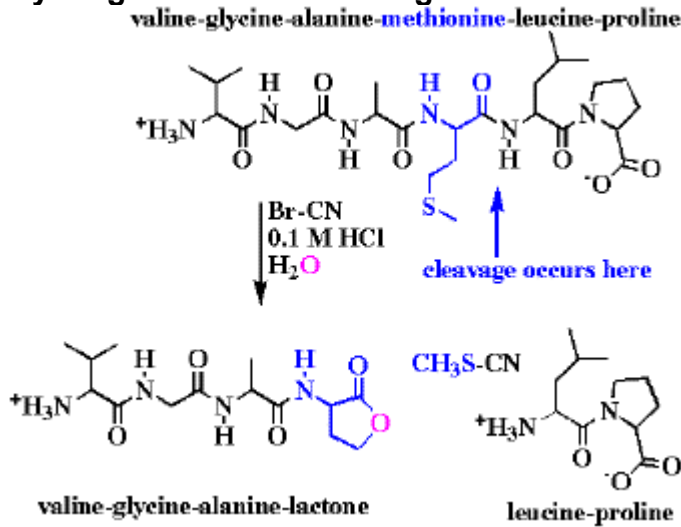polypeptides subjected to amino acid analysis or Edman degradation

Sample Hydrolysis
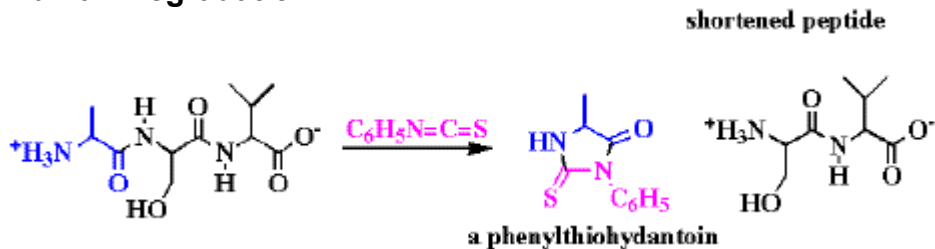

alanine-serine-asparagine

Sequence Analysis
- A Specific Chemical Cleavage:
    - Cyanogen bromide (Br-CN): cleaves peptide bonds formed by carboxyl of methionine
- Specific Enzymatic Cleavages:
    - Trypsin: cleaves peptide bonds formed by carboxyl of lysine and arginine
    - Chymotrypsin: cleaves peptide bonds formed by carboxyl of phenylalanine, tyrosine, and tryptophan
- Sequential Degradation:
    - Edman Degradation: cleaves n-terminal aa

Cyanogen Bromide Cleavage

valine-glycine-alanine-methionine-leucine-proline

Br-CN
0.1 M HCl
H₂O

cleavage occurs here

CH₃S-CN

valine-glycine-alanine-lactone

leucine-proline

Edman Degradation

shortened peptide


$C_6H_5N=C=S$

a phenylthiohydantoin

This process can be repeated ~20-30 times
Long polypeptides require cleavage to small units which can be
individually sequenced

Challenge
Given:

alanine-glycine-methionine-phenylalanine-arginine

Peptide

Trypsin

Br-CN
alanine-glycine-lactone
phenylalanine-arginine-leucine

Chymotrypsin
alanine-glycine-methionine-phenylalanine
arginine-leucine
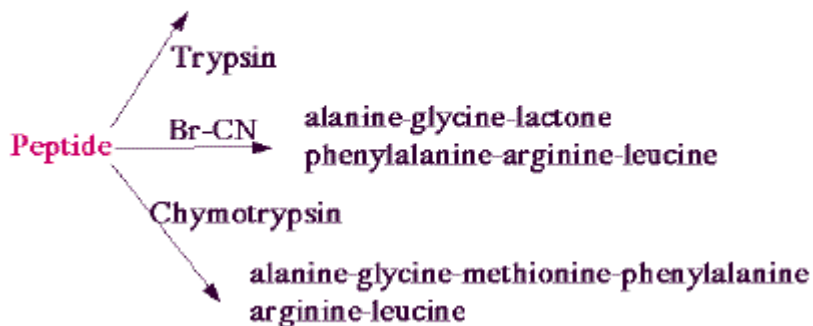
What is the structure of the original peptide?

Secondary Structure
Common local conformations:
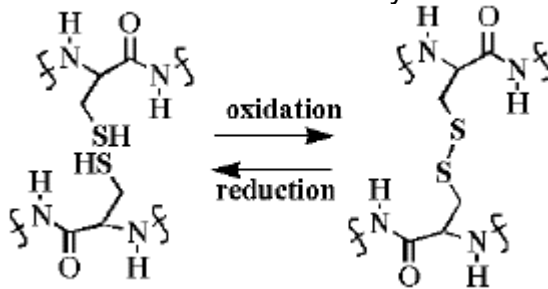- Alpha Helix
- Beta-Pleated Sheet

Stabilized by:
- Hydrogen bonding
- R-groups provided space
- s-trans amide bonds

Tertiary Structure
Stabilized by:
- Clustering of hydrophobic sidechains
- Salt bridges - oppositely charged sidechains clustering
- Disulfide bonds - 2 cysteines can be oxidized

Quaternary Structure
Forces holding protein chains together:
- Clustering of hydrophobic sidechains
- Disulfide Bonds

---

Polypeptide Synthesis
Problems:
- 20 different building blocks
- all building blocks connected by amide bonds
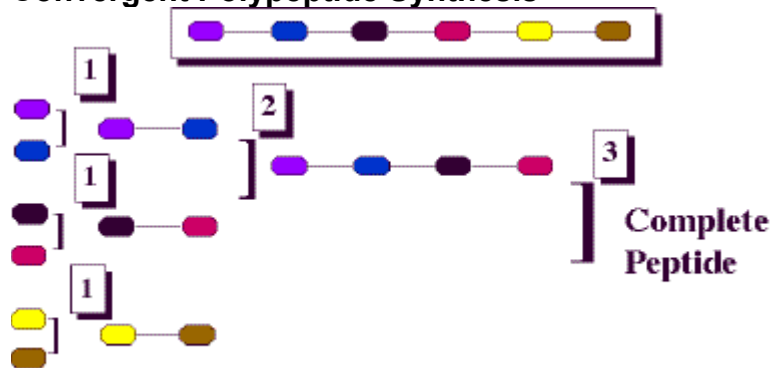- cannot simply mix all needed aa's and allow them to react randomly

---

Polypeptide Synthesis
- Strategies:
    1. Linear Synthesis
    2. Convergent Synthesis
    3. Solid-Phase Synthesis
- Requirements:
    1. Protecting groups to prevent unwanted couplings

Linear Polypeptide Synthesis
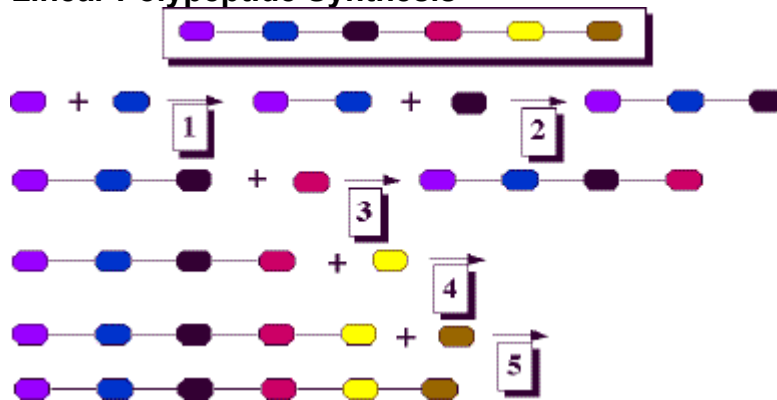
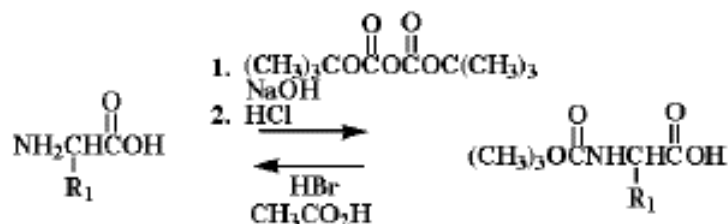**Convergent Polypeptide Synthesis**

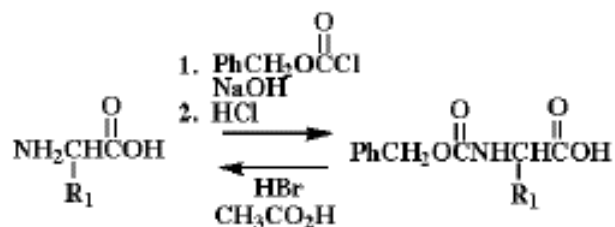Advantage: Fewer steps result in larger yields

Solid-Phase Polypeptide Synthesis: (Merrifield Synthesis)
- Process:
    1. Amino-protected amino acid coupled to resin
    2. Protecting group removed
    3. Another amino-protected amino acid added
       repeat steps 2-3, ending with step 2
    4. Cleave polypeptide from solid resin
- Advantages:
    1. Fully automated
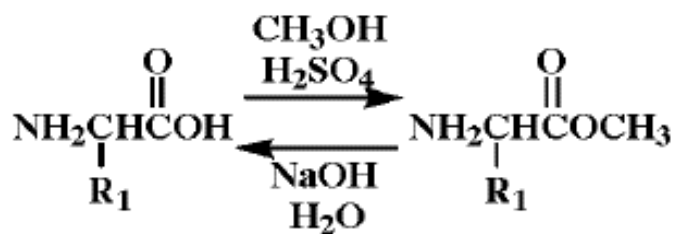   2. By-products easily rinsed away

Amino Protecting Groups
Amino groups commonly protected as carbamates

$$
NH_2CHCOH \xrightleftharpoons[\substack{HBr \\ CH_3CO_2H}]{\substack{1.\ PhCH_2OCCl \\ NaOH \\ 2.\ HCl}} PhCH_2OCNHCHCOH
$$

(structures with R_1 substituents)

$$
NH_2CHCOH \xrightleftharpoons[\substack{HBr \\ CH_3CO_2H}]{\substack{1.\ (CH_3)_3COCOCOC(CH_3)_3 \\ NaOH \\ 2.\ HCl}} (CH_3)_3OCNHCHCOH
$$

(structures with R_1 substituents)

Carboxyl Protecting Groups
Methyl, ethyl, and benzyl esters

$$
NH_2CHCOH \xrightleftharpoons[\substack{NaOH \\ H_2O}]{\substack{CH_3OH \\ H_2SO_4}} NH_2CHCOCH_3
$$

(structures with R_1 substituents)

Peptide Bond Forming Reaction

$$
Z\text{-}NHCHCOH + NH_2CHCOCH_3
$$

(R_1 and R_2 substituents)

dicyclohexylcarbodiimide (DCC)

$$
Z\text{-}NHCHCNHCHCOCH_3 + \text{dicyclohexylurea (DCU)}
$$

(R_1 and R_2 substituents)

Amino Acid Classifications
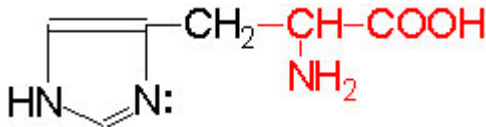
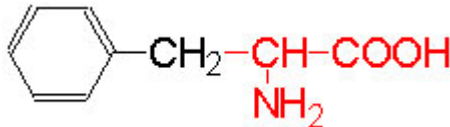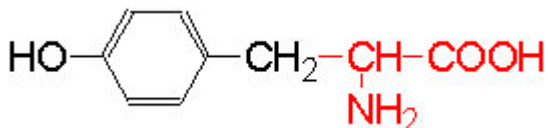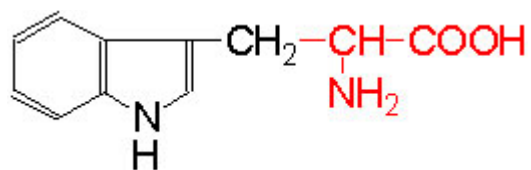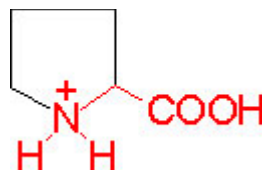There are 20 amino acids, called the standard amino acids, that are found in nearly all proteins. Amino acids are actually zwitterionic in structure, with a α COO- group and a α NH$_3$+ group. Because of their ionic nature, they are soluble in water and have high melting points. Simple proteins contain only amino acids. Conjugated proteins are bonded to a nonprotein prosthetic group such as a sugar, nucleic acid, lipid, or some other group. Each of the 20 a-amino acids found in proteins can be distinguished by the R-group substitution on the a-carbon atom. There are two broad classes of amino acids based upon whether the R-group is hydrophobic or hydrophilic.

The hydrophobic amino acids tend to repel the aqueous environment and, therefore, reside predominantly in the interior of proteins. This class of amino acids does not ionize nor participate in the formation of H-bonds. The hydrophilic amino acids tend to interact with the aqeuous environment, are often involved in the formation of H-bonds and are predominantly found on the exterior surfaces proteins or in the reactive centers of enzymes.

.

Table of α-Amino Acids Found in Proteins

Amino Acid	Symbol	Structure*	pK$_1$ (COOH)	pK$_2$ (NH$_2$)	pK R Group
Amino Acids with Aliphatic R-Groups					
Glycine	Gly - G	H–CH–COOH / NH$_2$	2.4	9.8	
Alanine	Ala - A	CH$_3$–CH–COOH / NH$_2$	2.4	9.9	
Valine	Val - V	H$_3$C / H$_3$C CH–CH–COOH / NH$_2$	2.2	9.7	
Leucine	Leu - L	H$_3$C / H$_3$C CH–CH$_2$–CH–COOH / NH$_2$	2.3	9.7	
Isoleucine	Ile - I	H$_3$C–CH$_2$ / H$_3$C CH–CH–COOH / NH$_2$	2.3	9.8	
Non-Aromatic Amino Acids with Hydroxyl R-Groups					

Serine	Ser - S	HO-CH$_2$-CH-COOH $\qquad$ NH$_2$	2.2	9.2	~13
Threonine	Thr - T	H$_3$C-CH-CH-COOH $\quad$ HO $\quad$ NH$_2$	2.1	9.1	~13
Amino Acids with Sulfur-Containing R-Groups					
Cysteine	Cys - C	HS-CH$_2$-CH-COOH $\qquad$ NH$_2$	1.9	10.8	8.3
Methionine	Met-M	H$_3$C-S-(CH$_2$)$_2$-CH-COOH $\qquad$ NH$_2$	2.1	9.3	
Acidic Amino Acids and their Amides					
Aspartic Acid	Asp - D	HOOC-CH$_2$-CH-COOH $\qquad$ NH$_2$	2.0	9.9	3.9
Asparagine	Asn - N	H$_2$N-C-CH$_2$-CH-COOH $\quad$ O $\quad$ NH$_2$	2.1	8.8	
Glutamic Acid	Glu - E	HOOC-CH$_2$-CH$_2$-CH-COOH $\qquad$ NH$_2$	2.1	9.5	4.1
Glutamine	Gln - Q	H$_2$N-C-CH$_2$-CH$_2$-CH-COOH $\quad$ O $\qquad$ NH$_2$	2.2	9.1	
Basic Amino Acids					
Arginine	Arg - R	HN-CH$_2$-CH$_2$-CH$_2$-CH-COOH $\;$ C=NH $\qquad$ NH$_2$ $\;$ NH$_2$	1.8	9.0	12.5
Lysine	Lys - K	H$_2$N-(CH$_2$)$_4$-CH-COOH $\qquad$ NH$_2$	2.2	9.2	10.8

Histidine	His - H		1.8	9.2	6.0
Amino Acids with Aromatic Rings					
Phenylalanine	Phe - F		2.2	9.2	
Tyrosine	Tyr - Y		2.2	9.1	10.1
Tryptophan	Trp-W		2.4	9.4	
Imino Acids					
Proline	Pro - P		2.0	10.6	

Acid-Base Properties of the Amino Acids

The a-COOH and a-NH_2 groups in amino acids are capable of ionizing (as are the acidic and basic R-groups of the amino acids). As a result of their ionizability the following ionic equilibrium reactions may be written:

$$R\text{-COOH} \longleftrightarrow R\text{-COO}^- + H^+$$

$$R\text{-NH}_3^+ \longleftrightarrow R\text{-NH}_2 + H^+$$
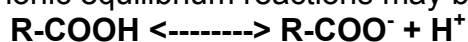
The equilibrium reactions, as written, demonstrate that amino acids contain at least two weakly acidic groups. However, the carboxyl group is a far stronger acid than the amino group. At physiological pH (around 7.4) the carboxyl group will be unprotonated and the amino group will be protonated. An amino acid with no ionizable R-group would be electrically neutral at this pH. This species is termed a **zwitterion**.

Like typical organic acids, the acidic strength of the carboxyl, amino and ionizable R-groups in amino acids can be defined by the association constant, K_a or more commonly the negative logrithm of K_a, the pK_a. The net charge (the algebraic sum of all the charged groups present) of any amino acid, peptide or

protein, will depend upon the pH of the surrounding aqueous environment. As the pH of a solution of an amino acid or protein changes so too does the net charge. This phenomenon can be observed during the titration of any amino acid or protein. When the net charge of an amino acid or protein is zero the pH will be equivalent to the **isoelectric point: pI**

The Peptide Bond

Peptide bond formation is a condensation reaction leading to the polymerization of amino acids into peptides and proteins. Peptides are small consisting of few amino acids. A number of hormones and neurotransmitters are peptides. Additionally, several antibiotics and antitumor agents are peptides. Proteins are polypeptides of greatly divergent length. The simplest peptide, a dipeptide, contains a single peptide bond formed by the condensation of the carboxyl group of one amino acid with the amino group of the second with the concomitant elimination of water. The presence of the carbonyl group in a peptide bond allows electron resonance stabilization to occur such that the peptide bond exhibits rigidity not unlike the typical -C=C- double bond. The peptide bond is, therefore, said to have partial double-bond character.

Protein Primary Structure

The primary structure of peptides and proteins refers to the linear number and order of the amino acids present. The convention for the designation of the order of amino acids is that the N-terminal end (i.e. the end bearing the residue with the free a-amino group) is to the left (and the number 1 amino acid) and the C-terminal end (i.e. the end with the residue containing a free a-carboxyl group) is to the right.

Protein Secondary Structure

The ordered array of amino acids in a protein confer regular conformational forms upon that protein. These conformations constitute the secondary structures of a protein. In general proteins fold into two broad classes of structure termed, globular proteins or fibrous proteins. Globular proteins are compactly folded and coiled, whereas, fibrous proteins are more filamentous or elongated. It is the partial double-bond character of the peptide bond that defines the conformations a polypeptide chain may assume. Within a single protein different regions of the polypeptide chain may assume different conformations determined by the primary sequence of the amino acids.

The Alpha-Helix

The a-helix is a common secondary structure encountered in proteins of the globular class. The formation of the a-helix is spontaneous and is stabilized by H-bonding between amide nitrogens and carbonyl carbons of peptide bonds spaced four residues apart. This orientation of H-bonding produces a helical coiling of the peptide backbone such that the R-groups lie on the exterior of the helix and perpendicular to its axis.

Not all amino acids favor the formation of the a-helix due to steric constraints of the R-groups. Amino acids such as A, D, E, I, L and M favor the formation of a-helices, whereas, G and P favor disruption of the helix. This is particularly true for P since it is a pyrrolidine based imino acid (HN=) whose structure significantly restricts movement about the peptide bond in which it is present, thereby, interfering with extension of the helix. The disruption of the helix is important as it introduces additional folding of the polypeptide backbone to allow the formation of globular proteins.

β -Sheets

Whereas an α-helix is composed of a single linear array of helically disposed amino acids, β-sheets are composed of 2 or more different regions of stretches of at least 5-10 amino acids. The folding and alignment of stretches of the polypeptide backbone aside one another to form b-sheets is stabilized by H-bonding between amide nitrogens and carbonyl carbons. However, the H-bonding residues are present in adjacently opposed stretches of the polypetide backbone as opposed to a linearly contiguous region of the backbone in the α-helix.

β-Sheets are said to be pleated. This is due to positioning of the a-carbons of the peptide bond which alternates above and below the plane of the sheet.

β-Sheets are either parallel or antiparallel. In parallel sheets adjacent peptide chains proceed in the same direction (i.e. the direction of N-terminal to C-terminal ends is the same), whereas, in antiparallel sheets adjacent chains are aligned in opposite directions.

Super-secondary Structure

Some proteins contain an ordered organization of secondary structures that form distinct functional domains or structural motifs. Examples include the helix-turn-helix domain of bacterial proteins that regulate transcription and the leucine zipper, helix-loop-helix and zinc finger domains of eukaryotic transcriptional regulators. These domains are termed super-secondary structures.

Forces Controlling Protein Structure
    Hydrogen Bonding:

Polypeptides contain numerous proton donors and acceptors both in their backbone and in the R-groups of the amino acids. The environment in which proteins are found also contains the ample H-bond donors and acceptors of the water molecule. H-bonding, therefore, occurs not only within and between polypeptide chains but with the surrounding aqueous medium.
    Hydrophobic Forces:

Proteins are composed of amino acids that contain either hydrophilic or hydrophobic R-groups. It is the nature of the interaction of the different R-groups with the aqueous environment that plays the major role in shaping protein structure. The spontaneous folded state of globular proteins is a reflection of a balance between the opposing energetics of H-bonding between hydrophilic R-

groups and the aqueous environment and the repulsion from the aqueous environment by the hydrophobic R-groups. The hydrophobicity of certain amino acid R-groups tends to drive them away from the exterior of proteins and into the interior. This driving force restricts the available conformations into which a protein may fold.

Electrostatic Forces:

Electrostatic forces are mainly of three types; charge-charge, charge-dipole and dipole-dipole. Typical charge-charge interactions that favor protein folding are those between oppositely charged R-groups such as K or R and D or E. A substantial component of the energy involved in protein folding is charge-dipole interactions. This refers to the interaction of ionized R-groups of amino acids with the dipole of the water molecule. The slight dipole moment that exist in the polar R-groups of amino acid also influences their interaction with water. It is, therefore, understandable that the majority of the amino acids found on the exterior surfaces of globular proteins contain charged or polar R-groups.

van der Waals Forces:

There are both attractive and repulsive van der Waals forces that control protein folding. Attractive van der Waals forces involve the interactions among induced dipoles that arise from fluctuations in the charge densities that occur between adjacent uncharged non-bonded atoms. Repulsive van der Waals forces involve the interactions that occur when uncharged non-bonded atoms come very close together but do not induce dipoles. The repulsion is the result of the electron-electron repulsion that occurs as two clouds of electrons begin to overlap.

Although van der Waals forces are extremely weak, relative to other forces governing conformation, it is the huge number of such interactions that occur in large protein molecules that make them significant to the folding of proteins.

Amino-Terminal Sequence Determination

Prior to sequencing peptides it is necessary to eliminate disulfide bonds within peptides and between peptides. Several different chemical reactions can be used in order to permit separation of peptide strands and prevent protein conformations that are dependent upon disulfide bonds. The most common treatments are to use either 2-mercaptoethanol or dithiothreitol. Both of these chemicals reduce disulfide bonds. To prevent reformation of the disulfide bonds the peptides are treated with iodoacetic acid in order to alkylate the free sulfhydryls.

There are three major chemical techniques for sequencing peptides and proteins from the N-terminus. These are the Sanger, Dansyl chloride and Edman techniques.

Sanger's Reagent: This sequencing technique utilizes the compound, 2,4-dinitrofluorobenzene (DNF) which reacts with the N-terminal residue under alkaline conditions. The derivatized amino acid can be hydrolyzed and will be labeled with a dinitrobenzene group that imparts a yellow color to the amino acid. Separation of the modified amino acids (DNP-derivative) by electrophoresis and

comparison with the migration of DNP-derivative standards allows for the identification of the N-terminal amino acid.

Dansyl chloride: Like DNF, dansyl chloride reacts with the N-terminal residue under alkaline conditions. Analysis of the modified amino acids is carried out similarly to the Sanger method except that the dansylated amino acids are detected by fluorescence. This imparts a higher sensitivity into this technique over that of the Sanger method.

Edman degradation: The utility of the Edman degradation technique is that it allows for additional amino acid sequence to be obtained from the N-terminus inward. Using this method it is possible to obtain the entire sequence of peptides. This method utilizes phenylisothiocyanate to react with the N-terminal residue under alkaline conditions. The resultant phenylthiocarbamyl derivatized amino acid is hydrolyzed in anhydrous acid. The hydrolysis reaction results in a rearrangement of the released N-terminal residue to a phenylthiohydantoin derivative. As in the Sanger and Dansyl chloride methods, the N-terminal residue is tagged with an identifiable marker, however, the added advantage of the Edman process is that the remainder of the peptide is intact. The entire sequence of reactions can be repeated over and over to obtain the sequences of the peptide. This process has subsequently been automated to allow rapid and efficient sequencing of even extremely small quantities of peptide.

Synthesis

The key structural element is the peptide bond. This is an amide linkage which joins the ammonium group of one amino acid to the carboxylate group of another by a new covalent bond. The O⁻ of the carboxylate is lost along with two H^+ ions from the ammonium group to form water. This is quite analogous to the formation of an amide by heating a carboxylic acid and an amine. The specific reaction conditions and processes required to do this may be (as we will see) quite sophisticated, but it helps to remember that what is being done is the joining of a carboxylate carbon and an ammonium nitrogen by a new C-N (peptide) bond.
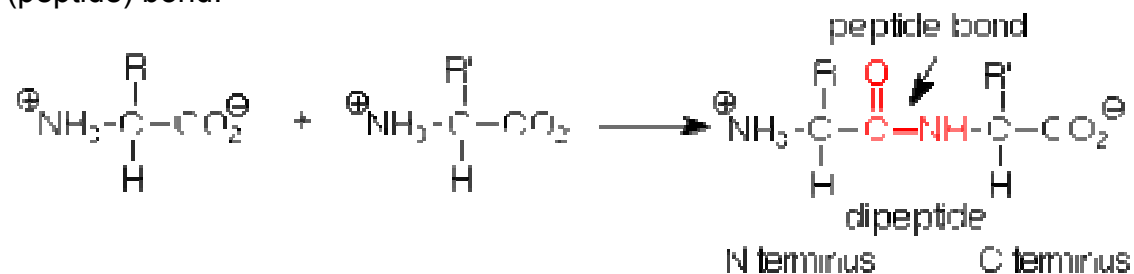

The new compound formed in this way is called a peptide. Our example is a dipeptide, formed from two amino acids. If a third amino acid is connected to the dipeptide by forming a new peptide bond at either the ammonium group or the carboxylate group of the dipeptide, we obtain a tripeptide, and so on. Polypeptides may have many amino acids. Polypeptides with more than 100 amino acids are considered to be proteins.

Since the amino acid whose carboxylic acid group participated in the formation of the peptide bond still has an ammonium group which contains a

nitrogen atom, it is called the N terminus of the peptide. The N terminus is conventionally written to the left. Correspondingly, the amino acid which still has a free carboxylate group is called the C terminus and is written to the right. When the order of amino acids in a peptide is written out, it is conventional to write it left to right from the N terminus to the C terminus. The complete order of amino acids in a protein is called its sequence and is conveniently expressed by using the abbreviated names of the amino acids read from N to C terminus.

a tetrapeptide, phe-val-gly-ala

C terminus

N terminus
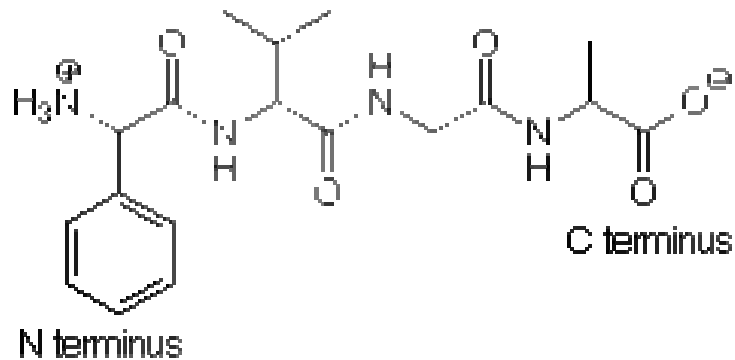
The sequence is held together by peptide bonds. As a part of an amide functional group, these bonds are difficult to break, so the sequence of a protein is quite stable. While there are many possible ways that a protein chain could be folded, the particular folding pattern adopted by the protein is completely determined by the its sequence.