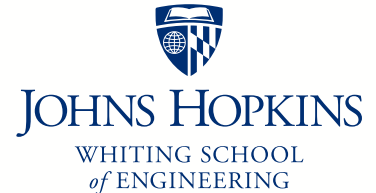


# Johns Hopkins Engineering

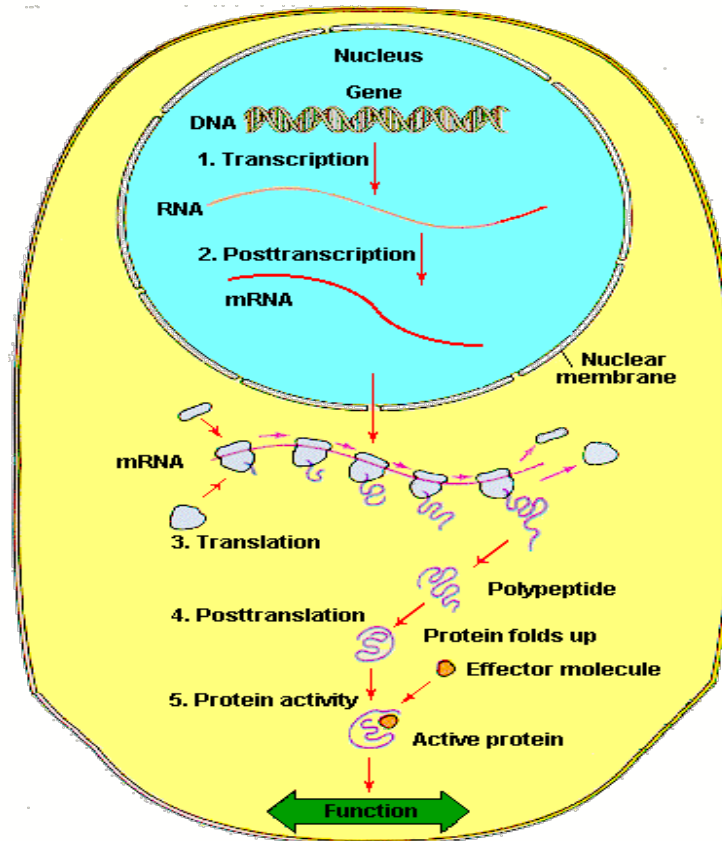
## Molecular Biology

Richard S. Potember, Ph.D.  
Potember@jhu.edu

Module 2 / Lecture 1



# Cell Processes



# The 30 Most Common Small Molecules in Cells

**Table 3-1 The 30 Most Common Small Molecules in Cells**

<i>Kind of Molecule</i>	<i>Number Present</i>	<i>Names of Molecules</i>		<i>Role in Cell</i>	<i>Figure Number for Structures</i>
Amino acid	20	Alanine Arginine Asparagine Aspartate Cysteine Glutamate Glutamine Glycine Histidine Isoleucine	Leucine Lysine Methionine Phenylalanine Proline Serine Threonine Tryptophan Tyrosine Valine	Monomeric units of all proteins	3-2
Aromatic base (purines and pyrimidines)	5	Adenine Cytosine Guanine	Thymine Uracil	Components of nucleic acids (DNA and RNA)	3-15
Sugar (monosaccharides)	2	Ribose		Component of nucleic acids	3-15
		Glucose		Energy metabolism; component of starch and glycogen	3-21
Lipid	3	Choline Glycerol Palmitate		Components of phospholipids	3-27b 3-28a

# Proteins Play Important and Widespread Roles in Cellular Structure and Function

**Table 3-3 Levels of Organization in Protein Structure**

<i>Level of Structure</i>	<i>Basis of Structure</i>	<i>Kinds of Bonds and Interactions Involved</i>
Primary	Amino acid sequence	Covalent peptide bonds
Secondary	Folding into $\alpha$ helix, $\beta$ sheet, or random coil	Hydrogen bonds
Tertiary	Three-dimensional folding of a single polypeptide chain	Hydrogen bonds, disulfide bonds, electrostatic interactions, hydrophobic effect
Quaternary	Association of two or more folded polypeptides to form a multimeric protein	Same as for tertiary structure

©Addison Wesley Longman, Inc.

## Levels of Organization in Protein Structure

# Proteins

**Proteins** are the most important and ubiquitous macromolecules in the cell.

Almost all enzymes are proteins, and proteins form the basis of most cellular structures.

Connective tissue, muscle fibrils, cilia, flagella-all are made of **proteins**.

Carbon dioxide fixation in photosynthesis, oxygen transport in the blood, and the motility of a flagellated bacterium depend on particular proteins with specific properties and functions.

Based on ***function***, proteins fall into **four** major classes.

# Proteins Fall Into 4 Major Classes

Many proteins are **enzymes**, (**catalysts**) that greatly increase rates of chemical reactions in cells.

**Structural proteins** provide support and shape to cells and organelles, giving cells their characteristic appearances.

**Motility proteins** play key roles in the contraction and movement of cells and intracellular structures.

**Regulatory proteins** are responsible for control and coordination of cellular functions, ensuring that cellular activities are regulated to meet cellular needs.

**Mono-functional** proteins have a single function: catalytic, structural, motile, or regulatory.

**Bi-functional** proteins play two different roles.

# The Monomers Are Amino Acids

Proteins are linear polymers of **amino acids**.

More than 60 different kinds of amino acids are typically present in a cell, but only 20 different kinds are used in protein synthesis.

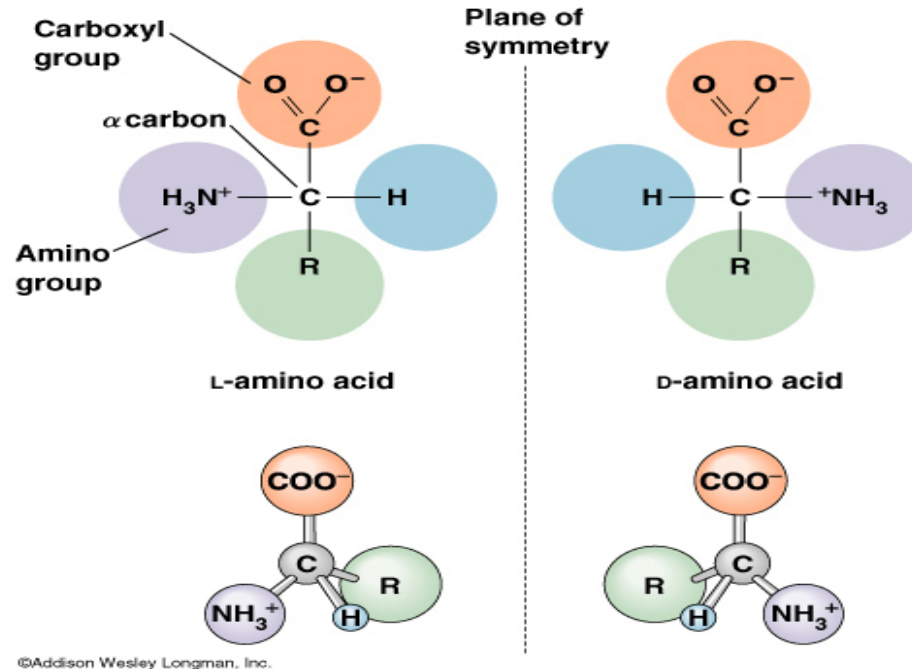
(Some proteins contain more than 20 different kinds of amino acids, but that is because of chemical modifications that occur after the protein has been synthesized.)

Although most proteins contain all or most of the 20 amino acids, the proportions vary greatly among proteins, and no two proteins have the same amino acid sequence.

Every amino acid has the **basic structure** with a **carbonyl group**, an **amino group**, a **hydrogen atom**, and a **so-called R group** all attached to a single carbon atom.

Except for glycine, for which the R group is just a hydrogen atom, **all amino acids have at least one asymmetric carbon atom** and therefore exist in two stereoisomeric forms, called D- and L-amino acids.

# The Structure and Stereochemistry of an Amino Acid



Because the  $\alpha$  carbon atom is asymmetric in all amino acids except glycine, most amino acids can exist in two isomeric forms, designated L and D and shown here as (top) conventional structural formulas and (bottom) ball-and-stick models. L and D forms are stereoisomers, with the vertical dashed line as the plane of symmetry. Of the two forms, only L-amino acids are present in proteins.

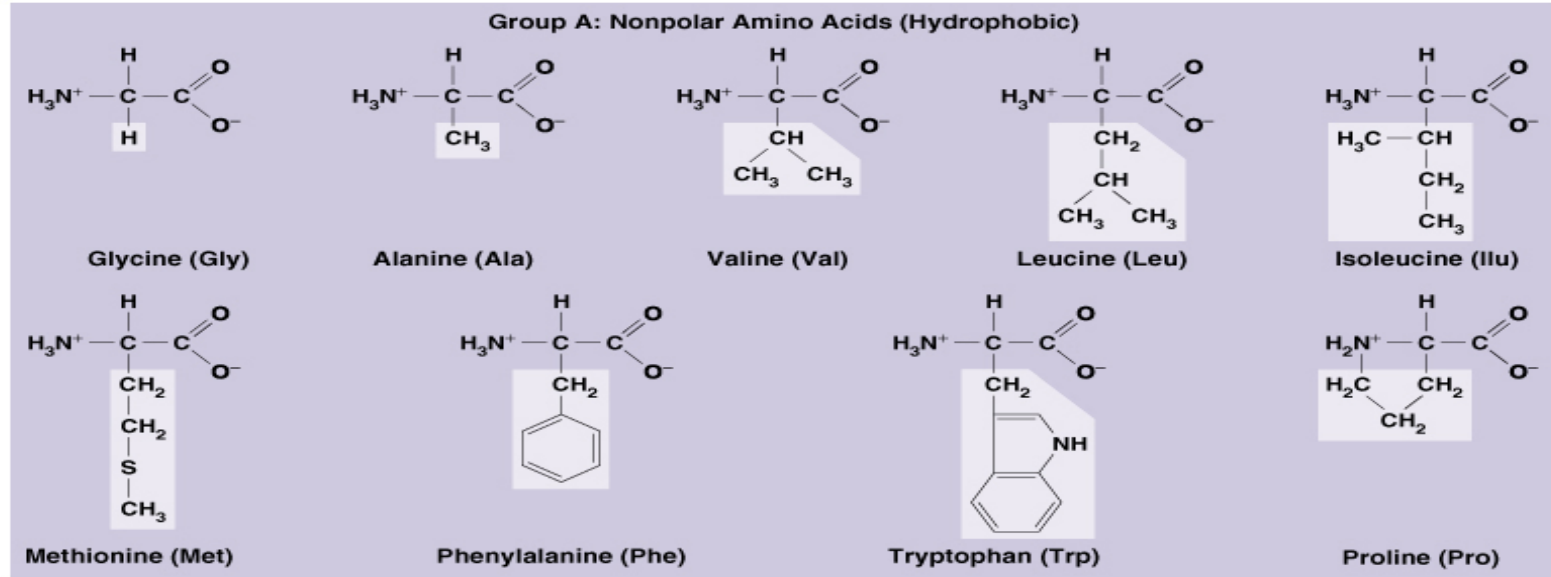


# Abbreviations

**Table 3-2 Abbreviations for Amino Acids**

<i>Amino Acid</i>	<i>Three-Letter Abbreviation</i>	<i>One-Letter Abbreviation</i>
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartate	Asp	D
Cysteine	Cys	C
Glutamate	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

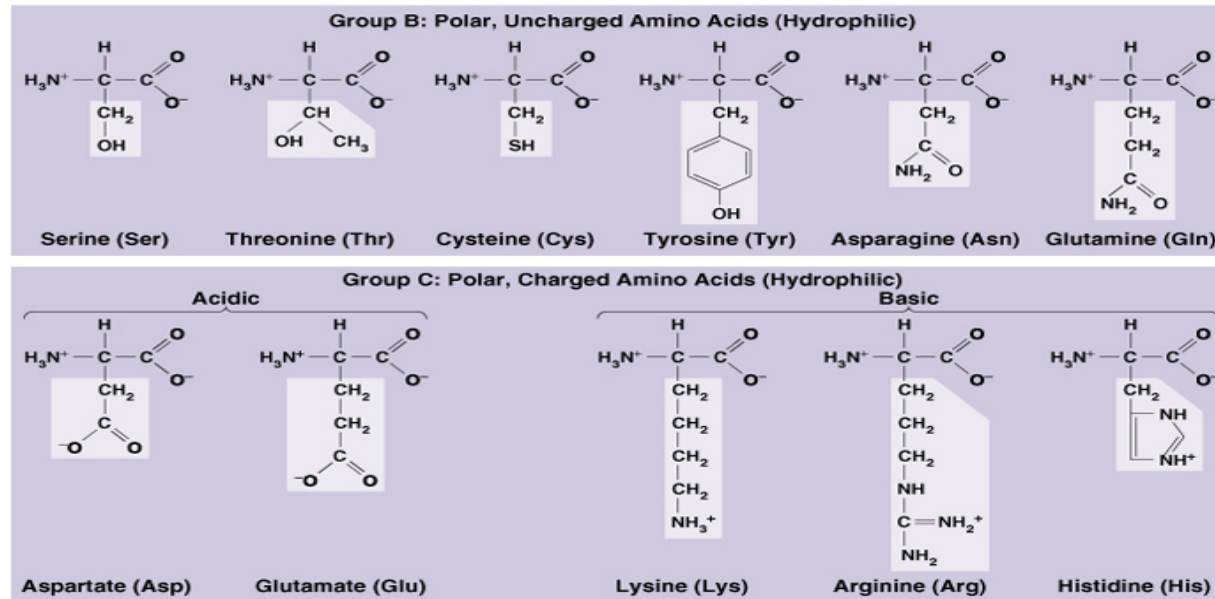
# The Structures of the 20 Amino Acids



©Addison Wesley Longman, Inc.

All amino acids have a carboxyl group and an amino group on the adjacent a carbon (purple), but each has its own distinctive R group (white). Group A have nonpolar R groups and are hydrophobic; notice the hydrocarbon nature of their R groups. The others are hydrophilic, because the R group is polar in nature (Group B) or because the R group is protonated or ionized at cellular pH and carries a formal electrostatic charge (Group C).

# The Structures of the 20 Amino Acids Found in Proteins



©Addison Wesley Longman, Inc.

All amino acids have a carboxyl group and an amino group on the adjacent a carbon (purple), but each has its own distinctive R group (white). Those in Group A have nonpolar R groups and are therefore hydrophobic; notice the hydrocarbon nature of their R groups. The others are hydrophilic, either because the R group is polar in nature (Group B) or because the R group is protonated or ionized at cellular pH and therefore carries a formal electrostatic charge (Group C).

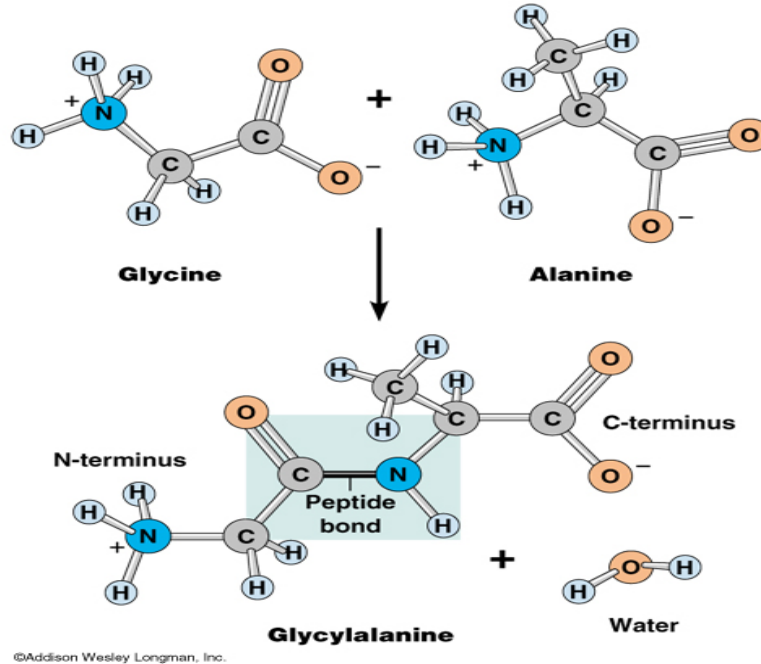
# The Polymers Are Polypeptides and Proteins

Stringing amino acids together into a linear polymer involves **stepwise addition** of **each** new amino acid to the growing chain by a dehydration (condensation) reaction.

The **-H** and **-OR** groups that are removed as **water** come from the carboxyl group of one amino acid and the amino group of the other, respectively.

The covalent bond between a **carboxyl group** and an **amino group** is called an **amide bond**, but in the special case in which both reactants are amino acids, it is referred to as a **peptide bond**.

# Peptide Bond Formation



Successive amino acids in a polypeptide are linked to one another by peptide bonds between the carboxyl group of one amino acid and the amino group of the next.

# Polypeptides and Proteins

The product of amino acid polymerization is a **polypeptide**.

A protein is a polypeptide chain (or several chains) that have attained a unique, stable, 3-D shape and is **biologically active**.

Some proteins consist of a **single** polypeptide and therefore achieve their final shape as a consequence of the folding and coiling that occur spontaneously as the chain is being formed. Such proteins are called **Monomeric proteins**.

On the one hand, a polypeptide is a *polymer*, with amino acids as its repeating units.

# Polypeptides and Proteins

**Ribonuclease** is a **monomeric** protein.

The term **protein synthesis** is appropriate for the reaction because the protein forms spontaneously as the polypeptide is elongated.

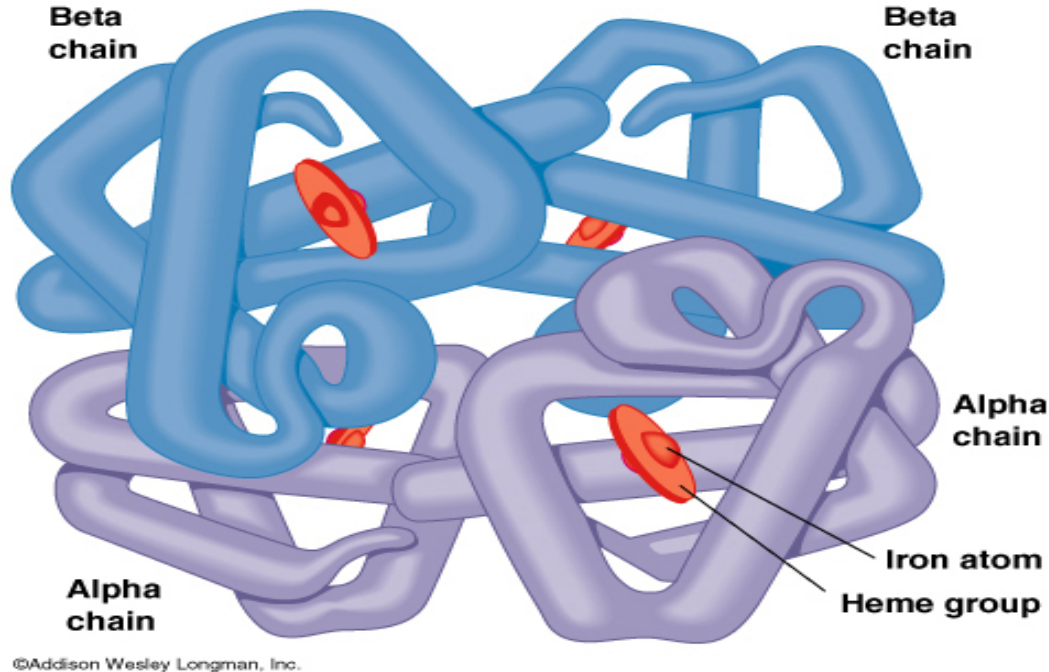
Many proteins are **multimeric proteins**, consisting of two or more polypeptide subunits.

A multimeric protein is said to be *homomeric* if its subunits are identical, heteromeric if it consists of *two* or more different kinds of *subunits*.

The **hemoglobin** carries oxygen in your bloodstream is a **multimeric** protein.

It contains **four** polypeptides, two called “a” chains and two called  $\beta$  chains. In such cases, protein synthesis involves elongation and folding of the polypeptide.

# Structure of Hemoglobin



Hemoglobin is a **multimeric** protein with 4 subunits (two  $\alpha$  polypeptides and two  $\beta$  polypeptides). Each subunit contains a **heme** group with an iron atom. Each **heme** iron can bind a single oxygen molecule.



# Primary Structure

The **primary structure** of a protein is the formal designation for the amino acid sequence of the constituent polypeptide(s).

When we describe the primary structure of a protein, we are simply specifying the order in which its amino acids appear from one end *of* the molecule to the other.

By convention, amino acid sequences are always written from the **N-terminus** to the **C-terminus** of the polypeptide, which is also the direction in which the polypeptide is synthesized.

Once incorporated into a polypeptide chain, individual amino acids are called ***amino acid residues***.

# Primary Structure

First protein to have its complete amino acid sequence determined was the *hormone **insulin***, in 1956 by **Frederick Sanger**, (received a Nobel Prize for this important technical advance).

**Sanger** cleaved ***insulin*** into smaller fragments and analyzed the amino acid order within the individual fragments.

Insulin consists of two polypeptides, called the *A chain* and the *B chain*; these have 21 and 30 amino acid residues, respectively.

Notice also the covalent disulfide (-S--S-) bond between two cysteine residues within the A chain and the two disulfide bonds linking the A and B chains.

Covalent disulfide (-S--S-) bonds play an important role in stabilizing the tertiary structure of many proteins.

# Secondary Structure

**The secondary structure** of a protein involves local interactions between amino acids that are brought into close proximity to each other.

These interactions result in two major structural patterns, the  **$\alpha$  helix and  $\beta$  sheet conformations**.

The helix structure was proposed in **1951** by **Linus Pauling and Robert Corey**.

An “ $\alpha$ ” helix is spiral in shape, consisting of a backbone of peptide bonds with the specific R groups of the individual amino acids jutting out from it.

A helical shape is common to repeating polymers, as we will see in **nucleic acids** and the **polysaccharides**.

# Secondary Structure

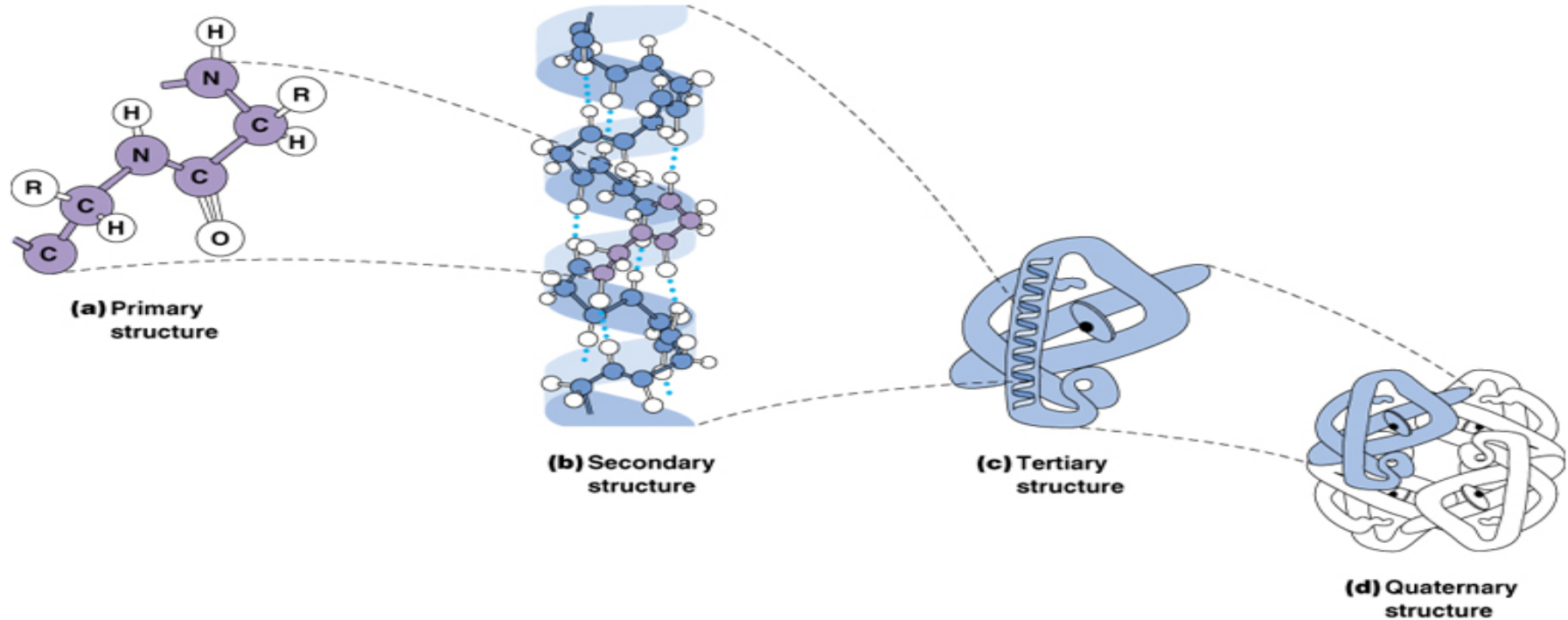
For the  **$\alpha$  helix**, there are **3.6 amino acids** per turn of the helix, bringing the peptide bonds of every fourth amino acid in close proximity. The distance between these juxtaposed peptide bonds is, in fact, just right for the formation of a hydrogen bond between the imido *group* (-NH----) of one peptide bond and the *carbonyl group of the other*.

As a result, every peptide bond in the helix is hydrogenbonded through its carbonyl group to the peptide bond immediately "below" it in the spiral and through its imino group to the peptide bond just "above" it.

In each case, however, the hydrogen-bonded peptide bonds are separated in linear sequence by the three amino acids required to advance the helix far enough to allow the two bonds to be juxtaposed.

These hydrogen bonds are all nearly parallel to the main axis of the helix and therefore tend to stabilize the spiral structure by holding successive turns of the helix together.

# Levels of Organization in Protein Structure



# Categories of Proteins

Proteins can be divided into two categories: ***fibrous*** proteins and ***globular*** proteins.

**Fibrous proteins** have extensive secondary structure (either a helix or beta sheet) throughout the molecule, giving them a highly ordered, repetitive structure.

In general, **secondary structure is much more important** than tertiary interactions in determining the shape of such proteins, which often have an extended, **filamentous structure**.

Prominent **examples of fibrous proteins** include ***silk*** (fibroin) and the ***keratins*** of hair and ***wool***, as well as ***collagen*** (found in tendons and skin) and ***elastin*** (present in ligaments and blood vessels).

As important as fibrous proteins may be, they represent only a small fraction of the kinds of proteins present in most cells.

# Globular Proteins

**Most** of the proteins involved in cellular structure are **globular proteins**, so named because their polypeptide chains are folded into compact structures rather than extended filaments.

The polypeptide chain of a **globular protein** is often folded locally into regions with alpha-helical or Beta-sheet structures, and these regions are themselves folded on one another to give the protein its compact, globular shape.

This folding is possible because regions of a helix or beta sheet are interspersed with **random coils**, irregularly structured regions that allow the polypeptide chain to loop and fold.

# Globular Proteins

**Globular proteins** have unique tertiary structures, made up of secondary structural elements (helices and sheets) folded in a specific way especially suited to the protein's particular functional role.

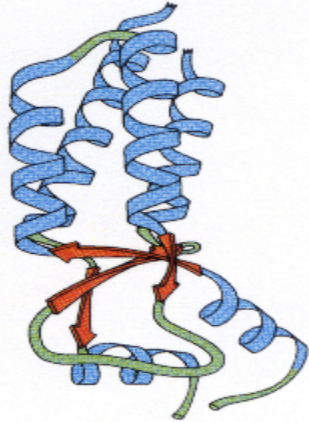
Most globular proteins consist of a number of segments called **domains**. A **domain** is a discrete, locally folded unit of tertiary structure, often containing regions of a helices and beta sheets packed together compactly.

A **domain** typically includes about 50-350 amino acids and usually has a specific function.

Small globular proteins tend to be folded into a single domain, as shown for **ribonuclease**, a relatively small protein.



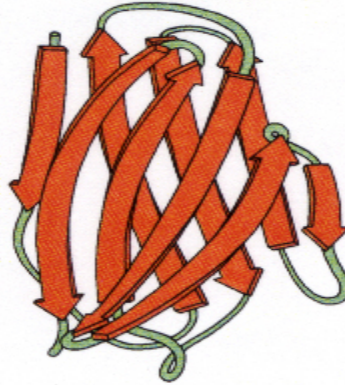
# Structures of Several Globular Proteins



Tobacco mosaic coat protein

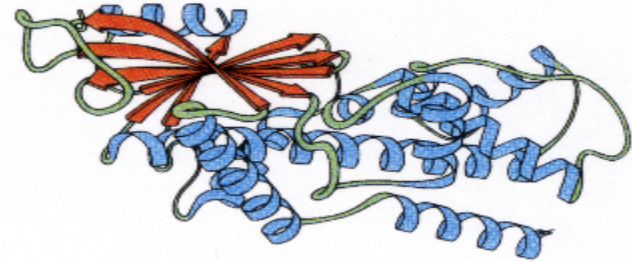
**(a) Predominantly  $\alpha$  helix**

©Addison Wesley Longman, Inc.



Immunoglobulin, V<sub>2</sub> domain

**(b) Predominantly  $\beta$  sheet**



Hexokinase, domain 2

**(c) Mixed  $\alpha$  helix and  $\beta$  sheet**

Shown above are proteins with different tertiary structures: (a) a predominantly alpha-helical structure (blue spirals), the coat protein of TMV; (b) a mainly beta-sheet structure (orange ribbons with arrows), the V(sub)2 domain of immunoglobulin; and (c) a structure that mixes alpha helix and beta sheet, domain 2 of hexokinase.

The immunoglobulin V(sub)2 domain is an example of an antiparallel beta-barrel structure, whereas the hexokinase domain 2 illustrates a twisted beta sheet. (Green segments are random coils.)

# Quaternary Structure

The quaternary structure of a protein is the level of organization concerned with subunit interactions and assembly.

Quaternary structure applies only to multimeric proteins.

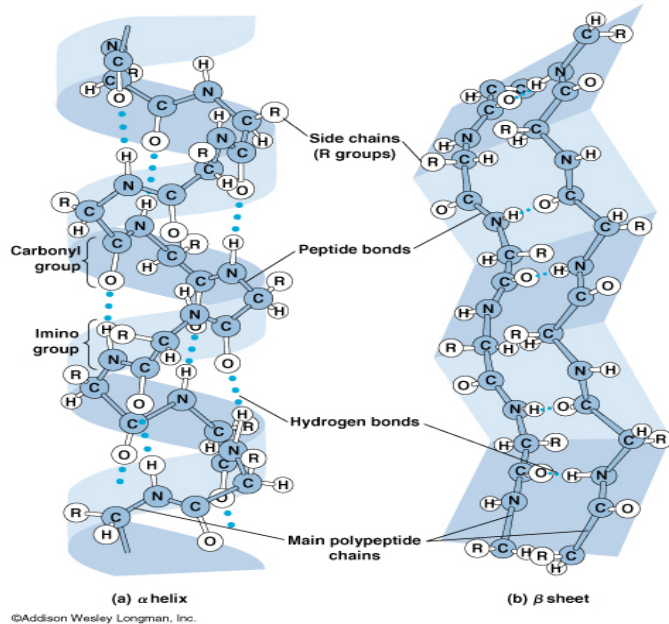
Many proteins are included in this category, particularly those with MW above 50,000.

**Hemoglobin** is a multimeric protein.

Some **multimeric proteins** contain identical polypeptide subunits; others, such as **hemoglobin**, contain several different kinds of polypeptides.

The bonds and forces that maintain quaternary structure are the same as those responsible for tertiary structure: [hydrogen bonds](#), [electrostatic interactions](#), [hydrophobic interactions](#), and [covalent disulfide bonds](#).

# The alpha helix and beta sheet



The alpha helix and beta sheet are two important elements in secondary structure of proteins.

(a) The alpha helix resembles a coiled telephone cord, but with each turn of the coil stabilized by hydrogen bonds (blue dots) between the carbonyl and imino groups of one peptide bond and those of the peptide bonds just "below" and "above" it in the helix.

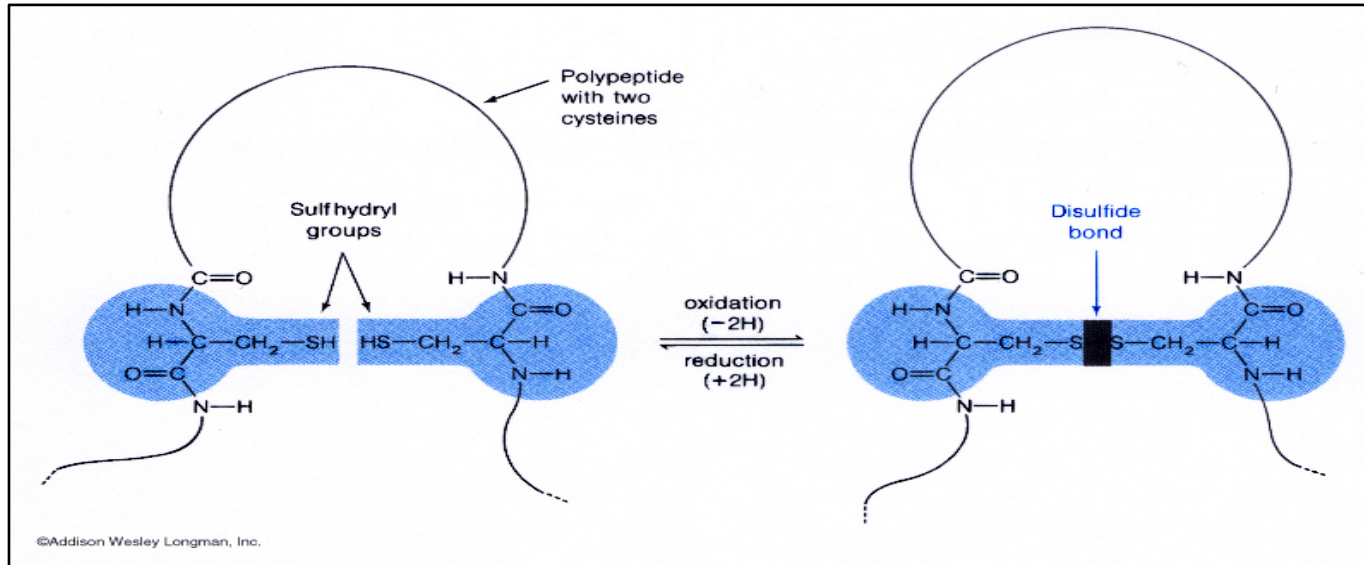
The hydrogen bonds of an alpha helix are within a single polypeptide chain and parallel to the polypeptide axis.

(b) The beta sheet resembles a pleated skirt, with successive atoms of each polypeptide chain located at folds of the pleats and with the R groups of the amino acids jutting out on alternating sides of the sheet.

The structure is stabilized by hydrogen bonds (blue dots) between the carbonyl and imino groups of peptide bonds in the adjacent polypeptides (or adjacent segments of the same polypeptide).

Hydrogen bonds are between adjacent polypeptide chains, but beta sheets can also form within the same polypeptide.

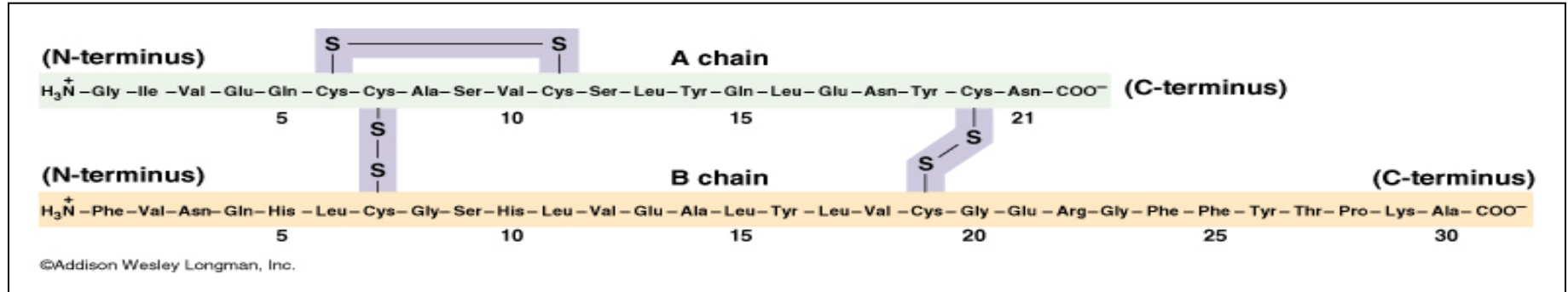
# Disulfide Bond Formation in Insulin



Disulfide Bond Formation in **Insulin**. A disulfide bond is formed by the oxidation of the sulfhydryl (JSH) groups of two cysteines (blue) located at different positions in a polypeptide chain (represented by the continuous line). A disulfide bond can be broken by reducing it to the two sulfhydryl groups from which it was originally formed.

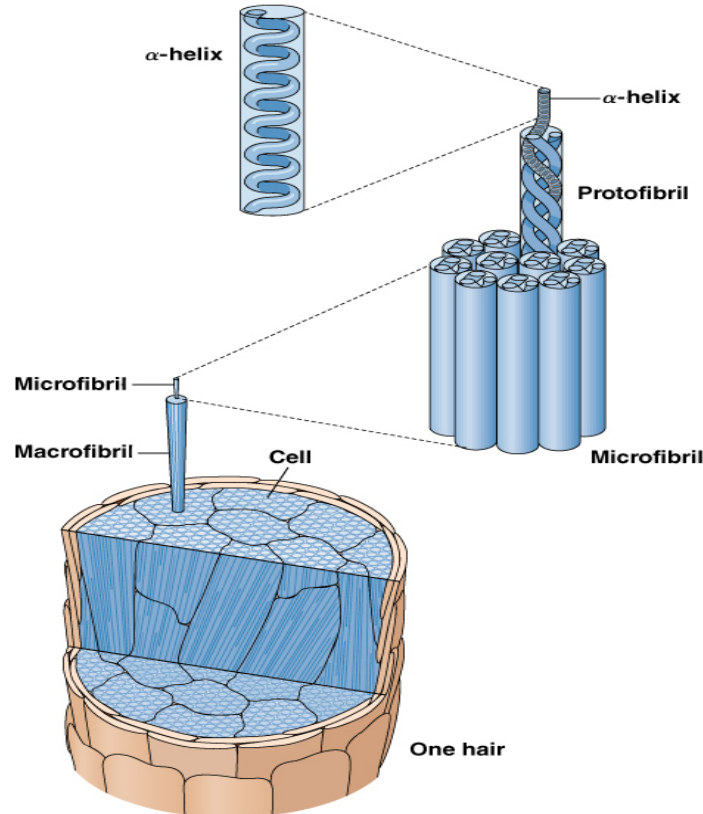
Disulfide bonds can also be formed between sulfhydryl groups in two separate polypeptide chains, thereby linking the two polypeptides covalently.

# The Primary Structure of Insulin



Insulin consists of two polypeptides, called the A and B chains, each shown here from its N-terminus to its C-terminus. The two chains are covalently linked by two inter-chain disulfide bonds.

# Structure of Hair



The structural protein of hair is alpha keratin, a fibrous protein with an alpha-helical shape.

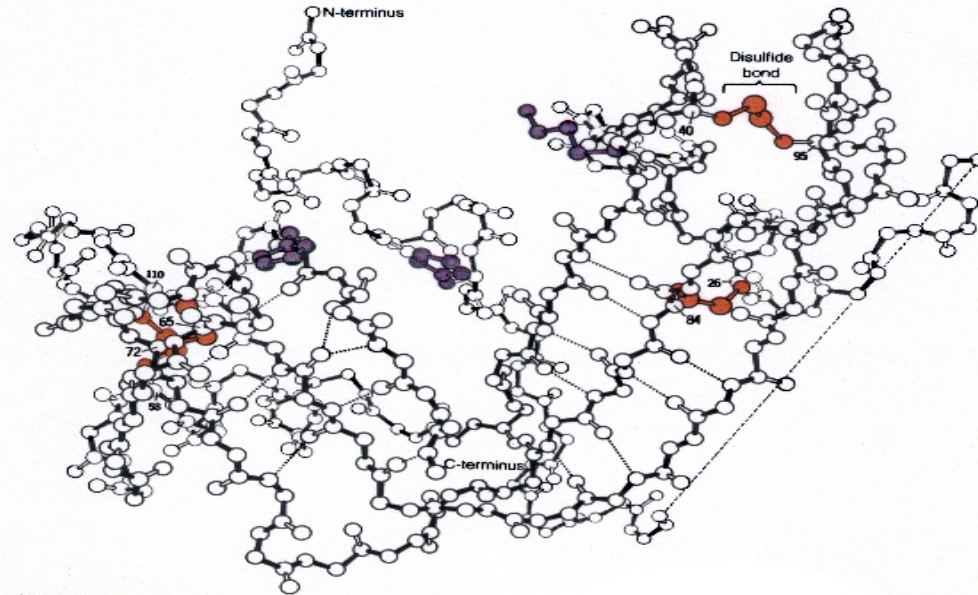
Three helices of alpha-keratin wrap into **protofibrils**, which then bond together to form **microfibrils**.

**Microfibrils** have a characteristic "9 + 2" structure in which two central protofibrils are surrounded by nine others.

Microfibrils, in turn, aggregate laterally to form **macrofibrils**. Bundles of macrofibrils pass through and around the cells in a hair.



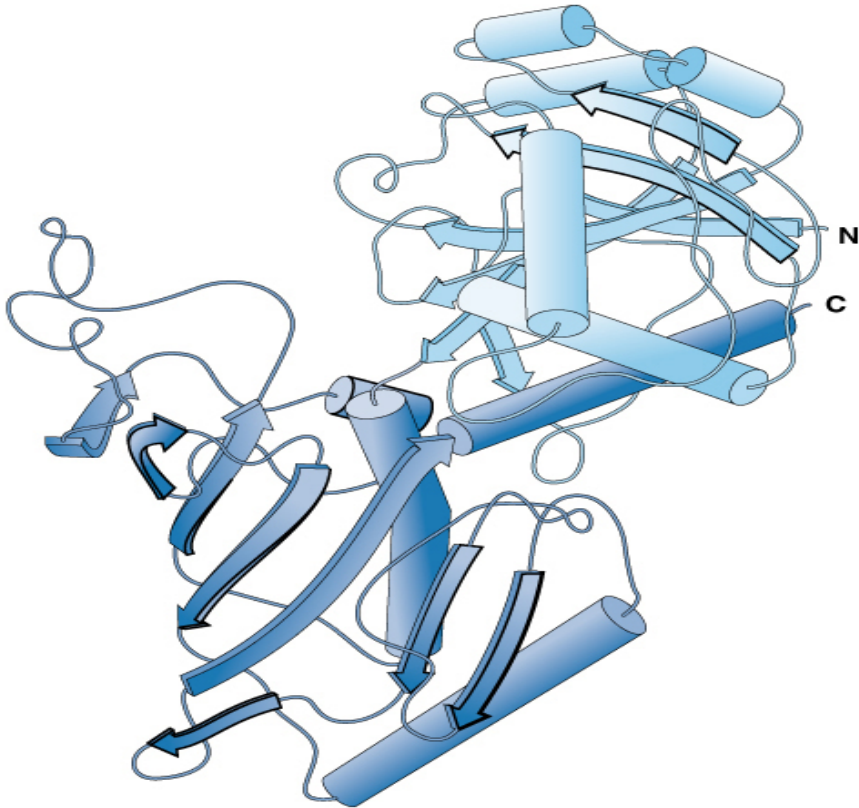
# Three-Dimensional Structure: Ribonuclease



(a) Ball-and-stick model  
©Addison Wesley Longman, Inc.

Ribonuclease is a typical globular protein with significant alpha-helical and beta-sheet segments connected by random coils. (a) A ball-and-stick model with disulfide bonds shown in orange and each of the relevant cysteines identified by its number in the amino acid sequence. The only side (R) groups shown are two histidines and a lysine that are important in the catalysis by this enzyme (purple).

# A Protein Containing Two Functional Domains



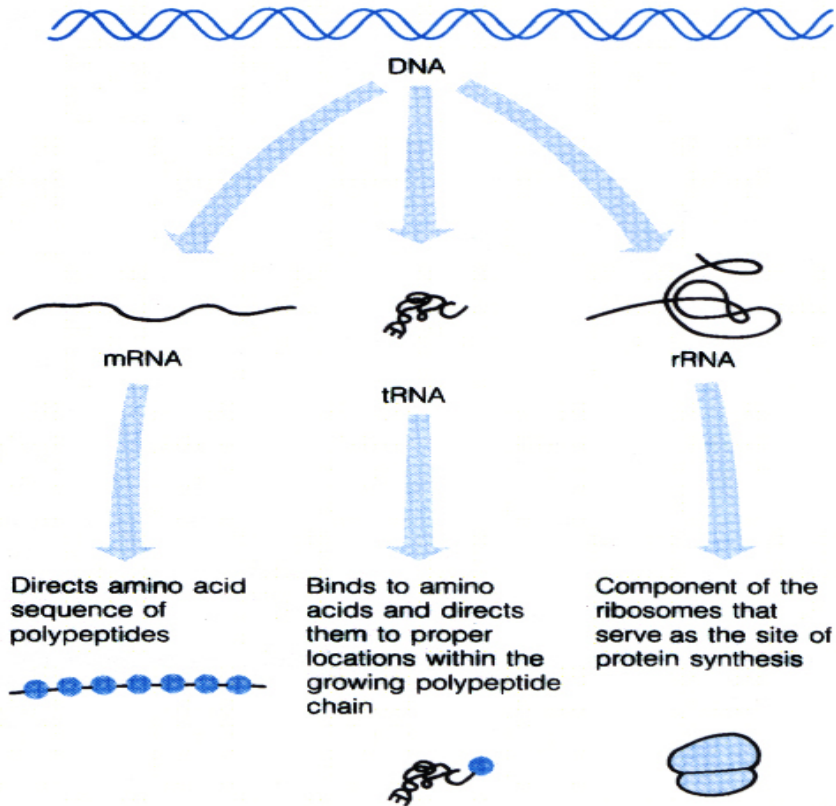
Model of the enzyme **glyceraldehyde phosphate dehydrogenase** shows a single polypeptide chain folded into two domains:

- one domain binds to the substance being metabolized
- the other domain binds to a chemical factor required for the reaction to occur

The two domains are indicated by different colored shading. The letters N and C refer to the N-terminus and C-terminus, respectively.

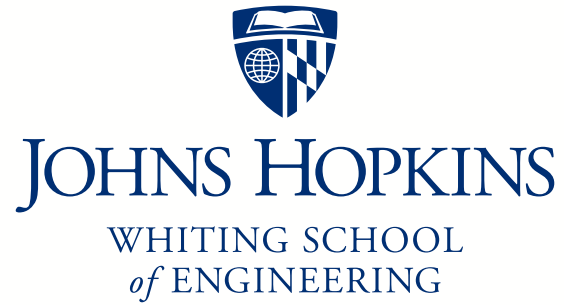


# The Roles of DNA and RNA in Protein Synthesis



The genetic information in DNA is transcribed into molecules of messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA).

Each type of RNA plays a specific role in protein synthesis.



© The Johns Hopkins University 2019, All Rights Reserved.