CHAPTER 2

THE CHEMICAL BASIS OF LIFE

OBJECTIVES

- Define and explain the chemical principles that form the basis of the chemistry of life.
- Clarify the principle of chemical bonding (covalent and noncovalent bonds).
- Explain ionization.
- Describe the chemistry of water and its relationship to biological chemistry and cell biology.
- Explain the chemistry of hydrophobic and hydrophilic molecules.
- Define and explain acids, bases, pH, and buffers for your students.
- Familiarize students with the structure and function of the four major groups of biological macromolecules.
- Get students to appreciate the similarities of and differences between the macromolecules.
- Explain the importance of polymerization to the production of macromolecules.
- Emphasize the importance of shape in biological chemistry.

LECTURE OUTLINE

Covalent Bonds

I. Molecular atoms are joined together by covalent bonds in which electron pairs are shared between atoms
   A. Formation of a covalent bond is governed by the basic principle that atoms are most stable with a full outer electron shell
      1. Number of bonds an atom forms determined by how many electrons are needed to fill outer shell
      2. Outer & only shell of hydrogen & helium atoms is filled when it contains 2 electrons; outer shells of other atoms are filled when they contain 8 electrons
      3. Example: oxygen with 6 outer-shell electrons can fill its outer shell by combining with 2 H atoms, forming a molecule water; oxygen atom linked to each H by a single covalent bond
   B. Bond formation is accompanied by energy release
      1. Later reabsorption of energy by bond breaks it; C—C, C—H or C—O covalent bonds require 80 - 100 kcal/mole to break
      2. This energy is quite large so these bonds are stable under most conditions
         a. 1 calorie = the amount of thermal energy required to raise the temperature of 1 gram of water 1°C; 1 kilocalorie (kcal; or 1 large Calorie) = 1000 calories
         b. Energy also expressed in Joules (measure of energy in terms of work); 1 kcal = 4186 Joules
         c. 1 mole = Avogadro's number (6.023×10^{23}) of molecules; a mole of a substance is its molecular weight expressed in grams
   C. Atoms can be joined by bonds in which >1 pair of electrons are shared: if 2 pairs are shared -> double bond (O_2); if 3 pairs shared -> triple bond (N_2); no quadruple bonds are known
   D. Type of bond can determine molecular shape - atoms joined by single bond can rotate relative to one another; atoms of double & triple bonds cannot

II. Electronegativity and unequal or equal sharing of electrons
   A. When atoms sharing electrons are the same, electrons are shared equally between the 2 atoms

B. If 2 unlike atoms share electrons, positively charged nucleus of 1 atom (the more **electronegative** atom) exerts a greater attractive force on the outer electrons than the other
   1. Thus, the outer electrons are located closer to the more electronegative atom
   2. Of atoms most often seen in biological molecules, nitrogen, oxygen - highly electronegative

III. Polar and non-polar molecules
   A. Water - O-H bonds in H_2O polarized; O atom is partially negative; the other [H] partially positive
      1. It is a polar molecule – such molecules have an asymmetric charge distribution or *dipole*
      1. O atom attracts electrons much more forcefully than does either of its H atoms
   B. Biologically important polar molecules have one or more electronegative atoms - usually O, N and/or S)
   C. Molecules without electronegative atoms & polar bonds (those made of C & H) are nonpolar
   D. Presence of strongly polarized bonds is of utmost import in determining molecular reactivity
      1. Large nonpolar molecules without electronegative atoms (waxes & fats) are relatively inert
      2. Molecules with electronegative atoms tend to be more reactive
      3. Many interesting biological molecules (proteins, phospholipids) have both polar & nonpolar regions & behave very differently

IV. Ionization - some atoms are so strongly electronegative that they can capture electrons from other atoms during a chemical reaction
   A. Sodium (Na; silver-colored metal) & chlorine (Cl; toxic gas); mix them; together form table salt
      1. Single electron in Na outer shell migrates to electron-deficient chlorine atom
      2. Each atom thus becomes charged (**ion**): Cl^- (anion) and Na^+ (cation); together form crystal
   B. Ions like Na^+ and Cl^- are relatively stable because they have a filled outer shell
   C. A different electron arrangement in atom can produce a highly reactive species (**free radical**)

Noncovalent Bonds

I. A variety of noncovalent bonds govern interactions between molecules or different parts of a large biological molecule; such bonds are typically weaker linkages, while covalent bonds are stronger
   A. Depend on attractive forces between atoms having an opposite charge
      1. Involve interaction between positively & negatively charged regions within same molecule or on 2 adjacent molecules; usually weaker than covalent bonds, which are strong
      2. Individual noncovalent bonds are often weak (~1 - 5 kcal/mole); they readily break & reform
      3. When many of them act in concert (DNA, protein, etc.), attractive forces add up & provide structure with considerable stability
   B. Noncovalent bonds mediate the dynamic interactions among molecules within the cell

II. Types of noncovalent bonds: Ionic bonds (or salt bridges)
   A. Ionic bonds - result from transfer of electron(s) from 1 atom to another leading to atoms with positive & negative charges that attract each other; can hold molecules together (DNA-protein)
      1. In crystal, strong; in water, ions surrounded by water, prevents attraction between them
      2. Water surrounds individual ions & inhibits oppositely charged ions from approaching each other closely enough to form ionic bonds
   B. Bonds between free ions not important in cells because cells are mostly water; weak ionic bonds between oppositely charged groups of large molecule are much more important
      1. Ionic bonds in cell are generally weak (~3 kcal/mole) due to presence of water
      2. Deep in protein core where water is excluded, they can be influential

III. Types of noncovalent bonds: hydrogen (H) bonds - hydrophilic (water-loving); enhance solubility in & interactions with water
   A. If H is bonded to electronegative atom (O or N), the shared electron pair is displaced toward electronegative atom so H is partially positive; H shared between two electronegative atoms
      1. Bare positively charged nucleus of H can approach unshared pair of outer electrons of second electronegative atom —> an attractive (weak electrostatic) interaction (an H bond)
      2. Occur between most polar molecules; important in determining structure & properties of water, also form between polar groups present in large biological molecules (like DNA)
   B. Strong collectively because their strength is additive; weak individually (2 - 5 kcal/mole in aqueous solutions); a result of polar covalent bonding; makes DNA double helix very stable

IV. Types of noncovalent bonds - hydrophobic (water-fearing) interactions
   A. Polar molecules like amino acids & sugars are said to be hydrophilic (water-loving); nonpolar molecules (fat molecules or steroids; water-fearing) are essentially insoluble in water
      1. Molecules with nonpolar covalent bonds lack charged region that can interact with poles of water molecules & are thus insoluble in water
      2. Hydrophobic molecules form into aggregates, minimizing exposure to polar surroundings (fat on chicken or beef soup); this type of interaction is called hydrophobic interaction
      3. Hydrophobic, nonpolar R groups congregate in soluble protein interior away from H_2O
   B. Most believe that they are not true bonds since not usually thought of as attraction between hydrophobic molecules
      1. Some believe they are driven by increased entropy, since nonpolar molecules in H_2O form H_2O into ordered cage; when hydrophobic groups cluster, H_2O becomes more disordered
      2. Others believe that hydrophobic interactions are driven by formation of weak bonds

V. Types of noncovalent bonds - van der Waals interactions (forces)
   A. Hydrophobic groups can form weak bonds with one another based on electrostatic interactions; due to slight perturbations of electron distributions
      1. Polar molecules associate because they contain permanent asymmetric charge distributions within their structure
      2. Electron distributions of nonpolar covalent bonds (like those in CH_4 or H_2) are not always symmetric & vary moment to moment
      3. Electron density may be larger on one side of atom or other even though electrons are shared equally; transient charge asymmetries result in momentary charge separations (dipoles)
   B. If 2 such molecules are very close together & appropriately oriented, 2 electrically neutral molecules will experience weak attractive force bonding them together (**van der Waals forces**)
      1. Formation of temporary charge separation in one molecule can induce similar separation in adjacent molecule & lead to additional attractive forces among nonpolar molecules
      2. Single van der Waals very weak (0.1 - 0.3 kcal/mole) & very sensitive to distance separating 2 atoms
      3. Molecules must be close together & interacting portions have complementary shapes that allow close approach; many atoms of both interactants can approach each other closely
      4. Important biologically as with interactions between antibodies and viral antigens

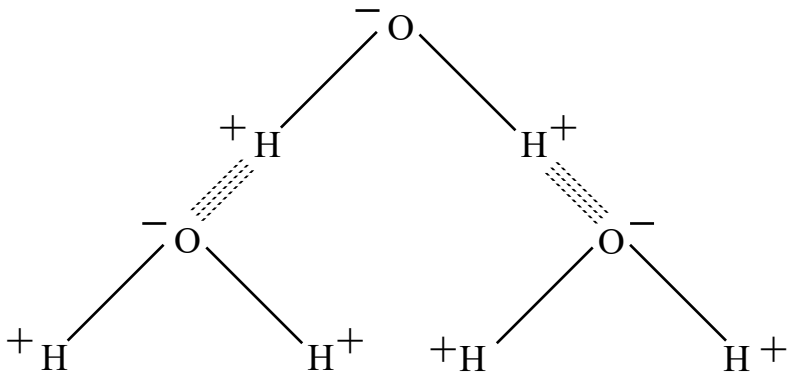
The Life-Supporting Properties of Water

I. Life on Earth totally dependent on water (maybe life anywhere in the Universe as well)

II. Unique water structure responsible for properties: highly asymmetric (O at one end, 2 H's at other end), its 2 highly polarized covalent bonds, very adept at forming H bonds
   A. Life-supporting attributes of water stem from above properties
   B. Each H_2O molecule H bonds with up to 4 others; forms highly interconnected molecular network
      1. Partially negative O at one end of molecule aligns with partially positive H of another one
      2. H_2O molecules have an unusually strong tendency to adhere to each other due to H bonds
   C. Comparison of water structure with that of H_2S (hydrogen sulfide)
      1. Like oxygen, sulfur has 6 outer-shell electrons & forms single bonds with 2 hydrogens
      2. Because sulfur is larger atom, it is less electronegative than oxygen & its ability to form H bonds is greatly reduced
      3. At room temperature, H_2S is a gas, not a liquid; temperature must drop to –86°C before it freezes to a solid
   D. The plentiful H bonds of water lead to its properties that relate to its importance to life

III. Tendency of water molecules to adhere to each other is evident in water's thermal properties
   A. H_2O has high heat capacity - heat energy disrupts H bonds instead of causing molecular motion that is measured as increased temperature so temperature does not rise too fast
   B. H_2O has a high heat of vaporization - H bonds must be broken to allow evaporation; explains the high energy needed to evaporate H_2O & convert it to steam
      1. When mammals sweat, heat absorbed from body used; explains sweat's cooling effect on body

IV. Also a good solvent - dissolves many things (**solutes**; more than any other solvent) but is inert itself
   A. Solubilizes ions & organic molecules - forms shell around ions separating them; H bonds with organic molecules containing polar groups (e. g. amino acids, sugars) & larger macromolecules
      1. Since they can form weak H bonds with water polar molecules are soluble within cell
   B. Determines structure of biological molecules & types of interactions in which they engage
   C. Water is the fluid matrix around which the insoluble fabric of the cell is constructed
      1. It is also the medium through which materials move from compartment to compartment
      2. It is a reactant or product in many cellular reactions
      3. It also protects cell from excessive heat, cold & damaging radiation
   D. High surface tension due to H bonding and capillary action
   E. Ice is less dense than liquid water, so ice floats; very important to aquatic ecosystems



Acids, Bases and Buffers

I. Acids & bases exist in pairs (couples)

A. Acid - a molecule able to release (or donate) a hydrogen ion to medium (**dissociation**); proton dissociates & is released into medium whenever a hydrogen atom loses an electron
   1. Once dissociated, proton can combine with other molecules forming H_3O^+, H_2O, NH_3^+, etc.
   2. When acid loses a proton, it becomes a base (termed the conjugate base of that acid)

B. Base - any molecule capable of accepting a hydrogen ion (proton)
   1. When base picks up a proton, it becomes an acid (the conjugate acid of that base)
   2. Acid always contains one more positive charge than its conjugate base

C. Amphoteric molecule - a molecule that can serve as both an acid & a base (usually both a positive & negative charge); water and amino acids are examples

II. Acids vary greatly in the ease with which they give up proton

A. If proton readily lost, attraction of conjugate base is lower & acid is stronger (ex.: HCl); it readily transfers its proton to water

B. Strong acid's conjugate base (ex.: Cl) is weak base; H^+ dissociates since H_2O is a stronger base

C. Weak acid (ex.: acetic acid) is mostly undissociated in H_2O; acetate ion is stronger base than H_2O

III. pH (measure of H^+ concentration) = - $\log_{10}$ $[H^+]$ where $[H^+]$ is the molar concentration of protons

A. Logarithmic scale - increase of 1 pH unit means 10X increase in OH^- or 10X decrease in H^+

B. Formula for dissociation of water into a hydroxyl ion & a proton: H_2O <—> $H^+ + OH^-$ or more accurately 2 H_2O <—> $H_3O^+ + OH^-$
   1. In aqueous solutions, protons do not exist in the free state, but rather as H_3O^+ or $H_5O_2^+$ ions but for simplicity one can refer to them as hydrogen ions or protons
   2. Equilibrium constant of water dissociation reaction is $K_{eq} = [H^+][OH^-]/[H_2O]$
   3. Since the concentration of pure H_2O is always 55.51 M, a new constant, K_w, the ion-product constant for water can be generated: $K_w = [H^+][OH^-] = 10^{-14}$ at 25°C; thus pH + pOH = 14

C. In pure water, $[H^+] = [OH^-] = \sim 10^{-7}$ M; low dissociation indicates water is very weak acid
   1. In presence of acid, $[H^+]$ rises & $[OH^-]$ drops (combine with protons to form water)
   2. Ion produce remains at 10^{-14}

IV. Most biological processes sensitive to pH changes since pH affects biological molecule ionic states

A. Amino acid R groups can acquire charge (—COOH —> $—COO^-$; $—NH_2$ —> $—NH_3^+$)

B. Even slight pH changes altering these groups can disrupt shape & activity of entire protein & impede biological reactions

V. Buffer - minimizes pH fluctuations & resists changes in pH; binds or releases (reacts with) H^+ & OH^- ions depending on conditions; they thus protect organisms & their cells

A. Usually contain weak acid with its conjugate base

B. Blood – H_2CO_3 & HCO_3^- ions; neutralizes H^+ rise during exercise, OH^- rise during hyperventilation
   1. Excess H^+ ions bind HCO_3^-; excess OH^- ions neutralized by protons derived from H_2CO_3
   2. Blood stays at pH 7.4

C. pH of fluid within cell regulated by phosphate buffer system ($H_2PO_4^-$ & HPO_4^{-2})

The Nature of Biological Molecules: Background

I. Organic molecules - contained in cell dry weight; once thought to only be found in living organisms; their name distinguishes them from inorganic molecules found in inanimate world

A. Chemists learned to synthesize them so some of the mystique was dispelled

B. Called them **biochemicals** (compounds made by living organisms)

II. Organic chemistry centers around carbon – both its size & electronic structure allow carbon to generate many molecules (several 100,000 known)
   A. Binds to up to 4 other atoms, since it has only 4 outer-shell electrons (8 needed to fill shell)
   B. Form carbon-containing backbones with long chains, which may be linear, branched or cyclic
   C. Carbons can be connected by single, double (with O and N) or triple bonds (with N)
   D. Compounds very stable since strength of covalent bond inversely proportional to atomic weight of elements involved; example: silicon (just below carbon in periodic table)
      1. Silicon (4 outer-shell electrons) is too large for its +-charged nucleus to attract neighboring atom valence (outer-shell) electrons enough to hold such large molecules together

III. Hydrocarbons - contain only hydrogen & carbon atoms (simplest group of organic molecules)
   A. As more carbons added, skeletons increase in length & structure becomes more complex
      1. As get bigger, can have same formula but different structures (structural isomers) & properties
   B. Fully reduced or saturated when each carbon bound to maximum number of hydrogen atoms
   C. Unsaturated compounds have double or triple bonds; lack maximum number of H atoms
   D. Rotation of carbons around single bonds, but not around double & triple bonds

IV. Functional groups - particular atom groupings that often behave as unit; responsible for physical properties, chemical reactivity & solubility in aqueous solutions; replace H's in hydrocarbons
   A. Hydrocarbons do not occur often in living cells although they form the bulk of fossil fuels formed from the remains of ancient plants & animals
      1. Many organic molecules important in biology contain chains of carbons like those in hydrocarbons but some of the hydrogens are replaced by various functional groups
   B. Some major functional groups
      1. Hydroxyl group - —OH
      2. Carboxyl group - —COOH; acquires charge —COO^-; carboxylic acids react with alcohols to form **ester bond**
      3. Sulfhydryl group - —SH; react to form disulfide bonds in polypeptides
      4. Amino group - —NH_2; acquires charge —NH_3^+; react with carboxylic acids & form **amide bonds**
   C. How do functional groups affect or change the properties of biochemicals?
      1. Usually contain one or more electronegative atoms (N, P, O and/or S) & thus make organic molecules more polar, more water soluble & more reactive
      2. Many are capable of ionization & may become positively or negatively charged
   D. Example of functional group importance (ethane -> ethanol -> acetic acid -> ethyl mercaptan)
      1. Ethane (CH_3CH_3) - toxic, flammable gas; if replace one H with hydroxyl (-OH) get....
      2. Ethyl alcohol (CH_3CH_2OH) which is palatable; if replace -CH_2OH with -COOH get .....
      3. Acetic acid (CH_3COOH), strong-tasting vinegar ingredient; if replace -COOH with -CH_2SH get..
      4. Ethyl mercaptan (CH_3CH_2SH) - strong, foul-smelling agent used to study enzyme reactions

The Nature of Biological Molecules: Functional Classification of Biological Molecules

I. **Macromolecules** - form structure & carry out activities of cells; usually huge & highly organized molecules; contain from dozens to millions of carbon atoms
   A. Because of their size & the intricate shapes they can assume, some can perform complex tasks with great precision & efficiency
   B. Endow organisms with properties of life & set them apart chemically from inanimate world
   C. Divided into 4 major categories: proteins, nucleic acids, polysaccharides, lipids - first 3 are **polymers**; made of large number of low MW building blocks (**monomers**)
   D. Basic structure & function of each type of macromolecule are similar in all organisms

1. If look at special sequences of monomers making up these various macromolecules, the diversity among organisms becomes apparent

II. Macromolecule building blocks – most macromolecules in cell have short lifetime compared with cell (except DNA); steadily broken down & replaced by new macromolecules
   A. Most cells contain supply (pool) of low MW precursors to build macromolecules
   B. Monomers - building blocks of macromolecules (sugars/polysaccharides, amino acids/proteins, nucleotides/nucleic acids, fatty acids & glycerol/lipids)
      1. Monomers joined together & form polymers by process like coupling railroad cars onto train

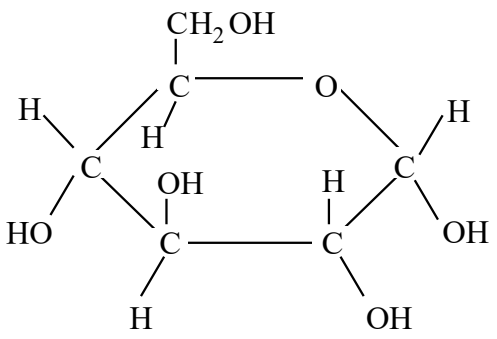
III. Metabolic intermediates (metabolites) – molecules in cell have complex chemical structures & must be synthesized in step-by-step sequence beginning with specific starting materials
   A. In cell, each series of chemical reactions is called a **metabolic pathway**
      1. Pathway starts with a compound & converts it to other ones sequentially until an end product that can be used in other reactions (like an amino acid building block of protein) is made
   B. Compounds formed along pathways leading to end products might have no function per se except as a stop on the way to the end product & are called **metabolic intermediates**

IV. Molecules of miscellaneous function – vast bulk of cell dry weight is made up of macromolecules & their direct precursors
   A. Vitamins – function primarily as adjuncts to proteins
   B. Certain steroid or amino acid hormones
   C. Molecules involved in energy storage (ATP, creatine phosphate)
   D. Regulatory molecules - cyclic AMP
   E. Metabolic waste products - urea

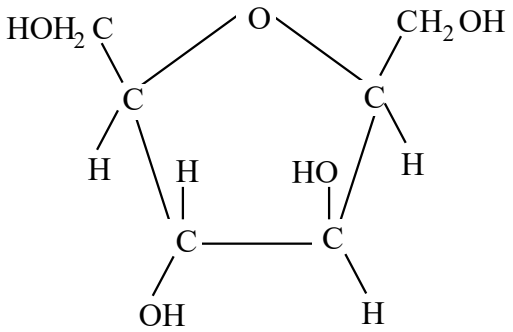
The Types of Biological Molecules: Carbohydrates

I. Carbohydrates comprise a group of substances, including simple sugars (**monosaccharides**) & larger molecules made from them
   A. Serve primarily as chemical energy storehouse & durable building material for biological construction
   B. Most have general formula $(CH_2O)_n$
      1. Important ones in cell metabolism have from 3 to 7 carbons (n = 3 - 7)
      2. Trioses, tetroses, pentoses, hexoses, & heptoses - 3, 4, 5, 6, & 7 carbons, respectively

II. The structure of simple sugars – each sugar molecule consists of carbon atom backbone linked together in linear array by single bonds
   A. Each carbon of backbone is linked to single OH group except for one bearing carbonyl (C=O) group
      1. Ketose - carbonyl group found at internal chain position; forms ketone group (e. g., fructose)
      2. Aldose - carbonyl group at one end of sugar forms aldehyde group (e. g., glucose)
   B. Sugars with 5 or more carbons convert by self-reaction into closed, ring-containing molecule with H's & OH's above or below ring; ring not planar but in 3D-conformation resembling chair



GLUCOSE



FRUCTOSE

III. Stereoisomerism - arrangement of groups around a carbon atom is depicted with carbon in center of tetrahedron with bonded groups projecting into its 4 corners

A. With 4 different groups attached to carbon, it can exist in 2 configurations that cannot be superimposed on one another

B. The 2 configurations are mirror images (**stereoisomers** or **enantiomers**) of each other; D-compound if OH group of C projects to right, L-compound if OH of C projects to left

1. Stereoisomers have essentially the same chemical reactivities
2. Carbon acting as site of stereoisomerism is **asymmetric** carbon; molecules can have more than one such carbon, which increases number of isomers
3. As backbone of sugar molecule increases in length so does number of asymmetric carbons & consequently the number of stereoisomers
4. D or L designation of molecule is based on arrangement of groups on carbon farthest from aldehyde (carbon associated with aldehyde is designated as C1)
5. Enzymes distinguish between D- & L-sugars; usually organism uses only one stereoisomer

C. Straight-chain glucose converts by self-reaction into 6-membered pyranose ring with carbon 1 being asymmetric

1. Unlike open chain precursor, C1 of ring form bears 4 different groups & thus becomes a new center of asymmetry within sugar molecule
2. If hydroxyl group of carbon 1 is below plane of ring called α-pyranose or α-glucose; if above plane of ring called β−pyranose or β−glucose
3. Difference in two forms important; results in compact shape of glycogen/starch (α) & extended conformation of cellulose (β)

IV. Linking sugars together to make larger molecules – bond joining sugars together called **glycosidic** linkage or bond (–C—O—C–); forms by reaction between C1 of one sugar & OH of another

A. Sugars can be joined by a variety of different glycosidic linkages

B. 2 monosaccharides covalently bond together to form **disaccharide**; serve primarily as readily available energy stores

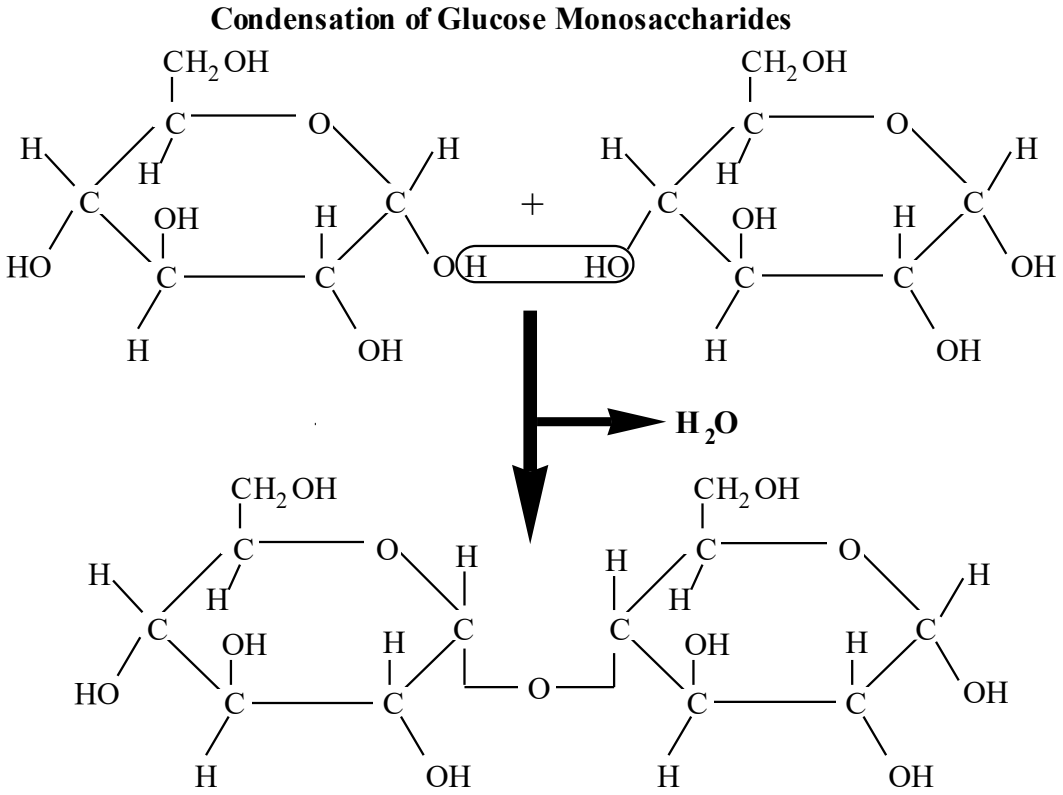
1. Sucrose (table sugar) - major component of plant sap; carries chemical energy from one part of plant to another
2. Lactose (milk sugar) - fuel for early growth & development of newborns
   a. Enzyme lactase that hydrolyzes it is found in membranes of cells lining intestines
   b. If lose this enzyme after childhood, eating dairy products causes digestive discomfort

C. Oligosaccharides - small chains of sugars (*oligo* - few), usually attached to lipids & proteins converting them to glycolipids & glycoproteins, respectively

1. Particularly important on plasma membrane from which they project

2. They may be composed of many different combinations of sugar units & can thus play an informational role
3. They can distinguish one cell type from another & help mediate specific interactions of a cell with its surroundings

D. Polysaccharides – many, many sugars hooked together; very large molecules



V. Claude Bernard & Diabetes – by mid-19th century, it was known that the blood of diabetics was sweet due to elevated glucose; tried to find blood sugar source (at first, thought it came from diet)

A. Found dogs on carbohydrate-free diet still had normal blood glucose levels -> body makes it
B. Found liver releases glucose to blood by hydrolyzing glycogen (insoluble glucose polymer)
C. Concluded food converted to glucose, which is stored as glycogen, released from liver if needed
D. Balance between glycogen formation & breakdown in liver is the prime determinant in maintaining the relatively constant (homeostatic) level of glucose in the blood

VI. Polysaccharide types - sugars, starches, cellulose, chitin, peptidoglycan, glycosaminoglycans

A. Glycogen – branched glucose polymer mostly joined by α(1—>4) bonds
   1. Branches every 10 or so units; sugar at branch joined to 3 neighboring units instead of 2; branch bond is α(1—>6) glycosidic linkage
   2. Surplus chemical energy storehouse in most animals; typical MW from 1 - 4 million daltons
   3. Human skeletal muscles have enough glycogen to fuel about 30 min of moderate activity
   4. Stored in cells as highly concentrated, dark-staining, irregular granules
B. Starch - glucose polymer; mixture of 2 different polymers (amylose & amylopectin); plants bank their surplus chemical energy in form of starch (potatoes & cereals are primarily starch)
   1. Amylose - unbranched, helical molecule; sugars joined by α(1—>4) linkages

2. Amylopectin - branched (less than glycogen with irregular branching pattern); α(1—>6) bonds at branch
3. Starch stored as densely packed granules (starch grains) found in membrane-bound plastids within plant cells
4. Animals possess enzyme (amylase) to hydrolyze starch even though they don't synthesize it

C. Cellulose – tough, durable structural material (cotton, linen); major plant cell wall component
1. Long, unbranched polymer; ordered into side-by-side aggregates to form molecular cables that resist pulling (tensile) forces; accounts for durability of cotton textiles
2. Solely glucose units joined by β(1—>4); its properties differ dramatically from the above polysaccharides because of the difference in bonds joining the glucose units
3. Most multicellular organisms lack enzyme to degrade it even though it is the most abundant organic material on Earth
4. Organisms that digest it & make a living from its (termites, sheep) harbor bacteria & protozoa that make needed cellulase

D. Chitin - unbranched polymer of N-acetylglucosamine (acetyl amino group instead of OH on glucose C_2)
1. Occurs widely as structural material among invertebrates (outer covering of insects, spiders, crustaceans)
2. Tough, resilient yet flexible; similar to certain plastics; insects owe much of their success to this highly adaptive polysaccharide covering

E. Glycosaminoglycans (GAGs) - repeating disaccharides (2 different sugars; —A—B—A—B—); occur in spaces surrounding cells

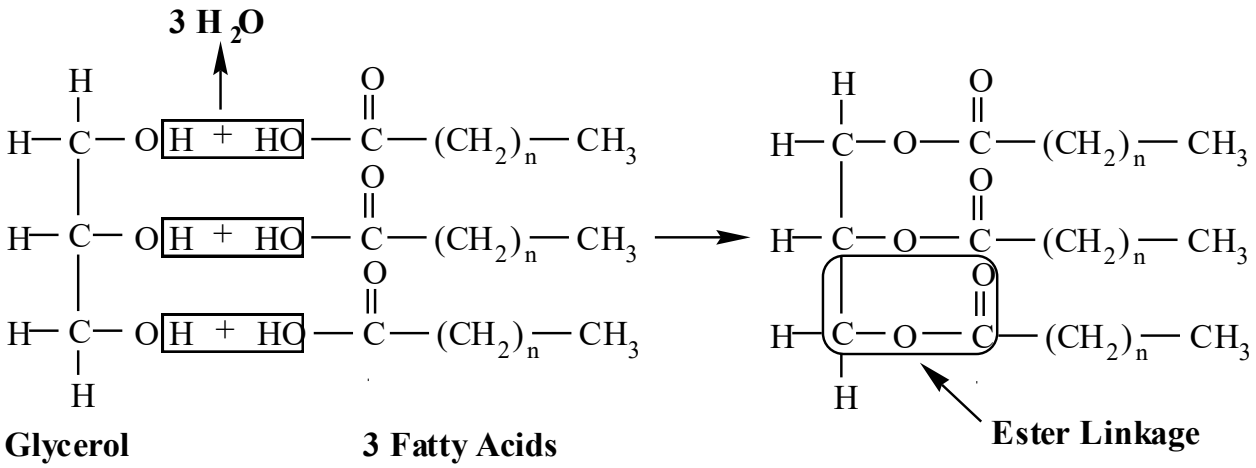
Polysaccharide	Location	Component Sugar	Type of Bonds	Function
Glycogen	Animal tissues	D-glucose	α(1—>4); highly branched every 8-10 residues via α(1—>6) linkages	Principal animal energy storage product (especially in liver & muscle)
Starch				
α-Amylose	Plants	D - glucose	α(1 —> 4); unbranched, forms helical coil	Principal higher plant energy storage product with amylopectin
Amylopectin	Plants	D - glucose	α(1—>4); branched every ~12-25 residues along backbone via α(1—>6) bonds. Branches ~12 residues long	Principal higher plant energy storage product with α–amylose
Cellulose	Some lower invertebrates & plants; usually extracellular (ex. pure cotton & linen)	D-glucose (disaccharide - cellobiose)	β (1—>4); no branching	Mainly structural; nutrient if can break it down (mammals who use as food lack enzyme to digest cellulose but get it from bacteria & protozoa in rumen); highly insoluble

Peptidoglycans	Bacterial cell wall	NAG-NAM (N-acetyl glucosamine & N-acetyl-muramic acid)	β(1—>4); no branching; cross-linked by peptide bonds between attached aminos	Structural, possibly some minor barrier functions
Chitin	Exoskeletons of insects & crustaceans	NAG (N-acetyl-D-glucosamine)	β(1—>4); no branching; close relative of cellulose	Structural component of exoskeleton (tough, resilient, flexible)
Glycosaminoglycans (GAGs)	Extracellular material (in spaces around cells) & connective tissue	Two alternating sugars (usually one is an amino sugar)	β(1—>4) & β(1—>4) or β(1—>3); no branching	Structural; extremely important in development

Lipids

I. Composed principally of C, H & O - not macromolecules, but aggregate to form large complexes
   A. Includes diverse, heterogeneous group of nonpolar biological molecules (fats, oils, phospholipids, sterols)
   B. Lumped together due to solubility in organic substances (benzene, chloroform) & their insolubility in H_2O, which explains many of their varied biological functions
   C. Serve as fuel molecules, very rich in chemical energy (contain more [>2X] energy than carbohydrates); structural components
      1. Carbohydrates - short-term, rapidly available energy source
      2. Fat stores - store energy more efficiently on a long-term basis
   D. Person of average size contains ~0.5 kg of carbohydrate primarily in form of glycogen (~2000 kcal of total energy) & ~16 kg of fat (144,000 kcal of energy)
      1. During strenuous day's exercise, a person can virtually deplete the body's entire carbohydrate store; fat stores take a long time to deplete
   E. Since they lack polar groups & are extremely water-insoluble, stored as dry lipid droplets in cells (extremely concentrated storage fuel)
      1. In many animals, fats are stored in special cells, adipocytes, whose cytoplasm is filled with one or a few large lipid droplets
      2. Adipocytes show remarkable ability to change volume to accommodate varying fat amounts

II. Triglyceride (neutral lipid, fats, triacylglycerol) - serves as lipid storage form for fuel (stored in adipocytes)
   A. Formed by 3 condensation reactions, which form ester linkages (—C—O—C—) between glycerol (a polar molecule) & 3 fatty acids



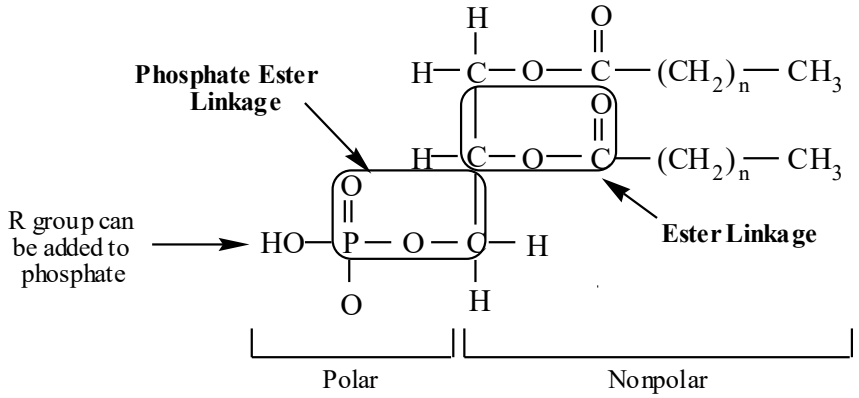
B. Fatty acid chains can vary in length & degree of saturation (see below)
C. 3 fatty acids of triglyceride need not be identical but may be; if they contain more than one fatty acid species called mixed fats
   1. Most natural fats (olive oil, butter fat) - mixtures of molecules with different fatty acid species

III. Fatty acids - long, unbranched hydrocarbon chains with single carboxyl group at one end
   A. Both hydrophobic (long C chain) & hydrophilic (carboxyl; "-" charge at physiological pH) in character (**amphipathic**); they have unusual & biologically important properties
   B. Fatty acids can differ in size (usually even & 14 to 20 carbons) & degree of saturation (presence or absence of double bonds)
      1. Saturated chains have no double bonds; every C attached to maximum number of H's; chains straight, pack tightly together (solid above room temperature [stearic acid, animal fats])
         a. Tristearate's fatty acids lack double bonds, is a common component of animal fats & remains in a solid state well above room temperature
      2. Unsaturated chains - 1 or a few double bonds (each causes bend/kink & if in *cis* configuration prevents tight packing); if prevalent, liquid at room temperature (ex.: plant fats called oils)
         a. In oils, double bonds lower the temperature at which a fatty acid containing lipid melts
         b. Multiple double bonds leads to oils being called polyunsaturated
         c. Example: linseed oil – highly volatile lipid extracted from flax seeds, remains liquid at much lower temperatures than tristearate
   C. Soap - in past, soap was made by heating animal fat in strong alkali (NaOH, KOH) to break fatty acid - glycerol bonds; most are now made synthetically
      1. Hydrophobic end of fatty acids embed in grease; hydrophilic end interacts with water
      2. Greasy materials form complexes that can be dispersed by water (micelles)
   D. Margarine formed from unsaturated vegetable oils by chemically reducing double bonds with H atoms (hydrogenation)

IV. Sterols and steroids – complex & characteristic 4 ringed hydrocarbon structures (4 joined rings differ in numbers & positions of double bonds & functional groups)
   A. Most common & one of most important - cholesterol; a component of animal cell membrane, but not in internal membranes or in plants
      1. Precursor for synthesis of many steroid hormones (testosterone, progesterone, estrogen)
      2. While largely absent from plant cells (vegetable oils considered to be cholesterol-free), plants may contain lots of related compounds
   B. Adrenocortical hormones
   C. Sex hormones - estrogen, progesterone, testosterone, etc.

D. Vitamin D_3 and the bile acids - involved in lipid digestion in the intestine

V. Phospholipids (phosphoglyceride, diacylglycerol) - glycerol + 2 fatty acids + phosphate group on third hydroxyl (often an amino group as well); highly charged at physiological pH; amphipathic
   A. Major cellular function - presence in membranes (properties of which depend on phospholipids)
   B. Usually has hydrophilic alcohol attached to phosphate [choline, ethanolamine, inositol (only 1 lacking charged amino group)]; with just phosphate group called phosphatidic acid
      1. The 2 ends have different properties: phosphate group end is hydrophilic & fatty acid tails are hydrophobic
   C. Usually have 16 & 18-C fatty acids, 1 saturated and 1 unsaturated fatty acid chain
   D. Length & degree of unsaturation affect membrane fluidity and are regulated



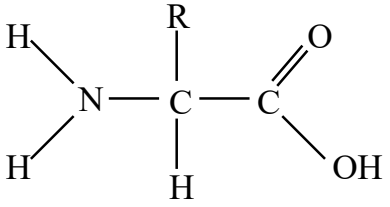
Proteins: General Information

I. Composed of H, C, O, N & usually S or P; very large macromolecules; polymers of amino acids
   A. Traits & functions - more varied role than other organism molecules (enzymes, structural or both); execute almost all cell activities; typical cell may have ~10,000 different ones; high specificity
      1. Enzymes catalyze and vastly accelerate rate of metabolic reactions
      2. Cytoskeletal elements serve as structural cables, provide mechanical support in & out of cells
      3. Hormones, growth factors, gene activators – wide variety of regulatory functions
      4. Membrane receptors & transporters - determine what cell reacts to, what can leave, enter cell
      5. Contractile elements & molecular motors – machinery for biological movements
      6. Antibodies and toxins
      7. Form blood clots
      8. Absorb or refract light
      9. Transport substances from one part of body to another

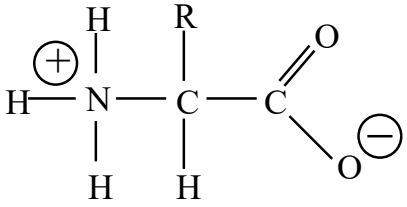
   B. The wide variety of protein functions comes from the virtually unlimited shapes they can assume
      1. They can exhibit a great variety of structures and thus a great variety of activities
      2. Each protein has a unique & structure enabling it to carry out a particular function
      3. Their shapes allow them to interact selectively with other molecules (high degree of specificity)
   C. Protein polymer sequences give them their unique properties
      1. Many protein capabilities can be understood by examining the chemical properties of its constituent amino acids
      2. 2 aspects of amino acid structure: that which is common to all & that which is unique to each
   D. Basic monomer building block is amino acid (20 types) that has central C with 4 attached groups
      1. Amino acid backbone is α-carbon between NH_2 & COOH groups; all but one (glycine) have symmetric centers so there are 2 stereoisomers D- & L-amino acids; only L-form in proteins
      2. In aqueous environment, —COOH ionizes to —COO^- & —NH_2 ionizes to —NH_3^+

3. Microorganisms use D-amino acids to synthesize certain small peptides, including those of cell wall & several antibiotics (e.g., gramicidin A)
   a. D-aminos cross-link peptidoglycans in bacterial cell walls (penicillin inhibits enzyme, transpeptidase, that crosslinks peptidoglycans by making peptide bonds)
   b. Found in certain peptides made by skin of South American tree frogs (one, **dermorphin**, acts as highly potent pain killer in rats through interaction with brain opiate receptors)
   c. Frog peptide D-aminos created by enzymes acting on L-aminos initially placed in peptide

Generalized Amino Acid



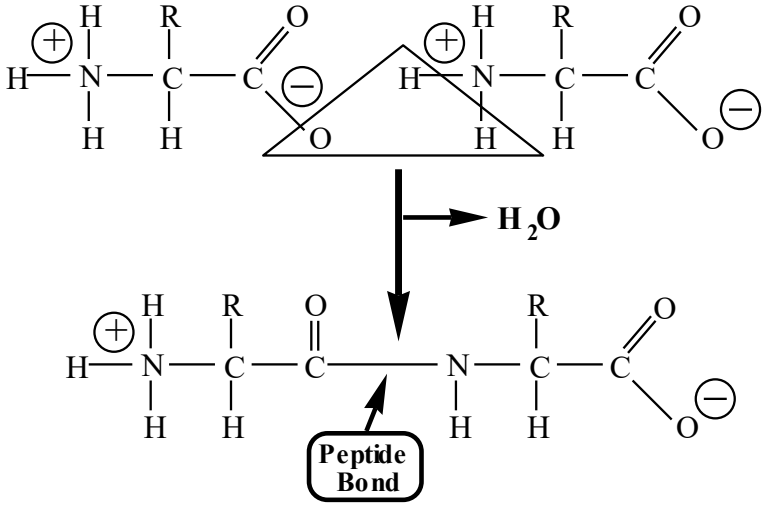
Generalized Amino Acid at Physiological pH



R = Radical or Side Group

E. During protein synthesis, each amino is joined to 2 other amino acids forming a long, continuous, unbranched polymer (**polypeptide chain**); have N-terminus & C-terminus
   1. Peptide bonds form by condensation reactions (elimination of water molecule) attaching α-carboxyl of 1 amino acid to α-amino group of another; form polypeptides by repeating it
   2. Average chain has ~450 amino acid residues; longest protein is muscle titin (27,000 aminos)
   3. Once incorporated into chain amino acids are termed **residues**; residue on one end (N-terminus) contains free α-amino group; other end (C-terminus) has free α-carboxyl group

A Dipeptide Formed By Condensation-Dehydration



II. Properties of R groups - determine inter- & intramolecular interactions that determine molecular structure & protein activities, respectively; give amino acids their variability
   A. Polypeptide backbone is made of that part of each amino acid that is common to all of them
      1. Common amino acid backbones allow polymerization to form by same reaction no matter which two amino acids are being joined

B. Side chain (R group) bonded to α-carbon is highly variable; this gives proteins their diverse structures & activities
   1. The various side chains considered together exhibit a large variety of structural features, fully charged to hydrophobic
   2. Participate in wide variety of organic reactions; can form many covalent & noncovalent bonds

C. Side chains of enzyme active sites can facilitate (catalyze) many different organic reactions
   1. Assorted protein side chains are important in intramolecular (determine molecule structure & activity) intermolecular (determine relationship of protein to other molecules) bonds

III. Four amino acid & R group categories – classified by R group character; not all of the amino acids are found in all proteins; nor are various amino acids distributed in an equivalent manner

A. Polar charged - contain R groups that act as stronger organic acids, bases; can form ionic bonds
   1. Almost always fully charged (lysine, arginine, aspartic acid, glutamic acid) at pH 7; side chains are relatively strong organic acids & bases
   2. Can form ionic bonds due to charges; histones with arginine (+-charge) bind to negatively charged phosphate groups of DNA
   3. Histidine - usually only partially charged at pH 7; often important in enzyme active sites due to its ability gain or lose a proton in physiologic pH ranges

B. Polar uncharged - R groups weakly acidic or basic; not fully charged at pH 7; can form H bonds with other molecules like water since they have atoms with a partial negative or positive charge
   1. Asparagine & glutamine [amides of aspartic & glutamic acid], threonine, serine, tyrosine
   2. Often quite reactive

C. Nonpolar - R groups hydrophobic; generally lack O & N; cannot interact with water or form electrostatic bonds; vary primarily in size & shape; allows them to pack tightly into protein core
   1. Alanine, valine, leucine, isoleucine, tryptophan, phenylalanine, methionine
   2. Associate with one another via hydrophobic & van der Waals interactions in protein core

D. The other three – glycine, proline, cysteine
   1. Glycine (R = H) - small R group makes backbone flexible & able to move so it is useful in protein hinges; small R group allows 2 backbones (of same or different protein) to approach closely
   2. Proline – R group forms ring with amino group (making it an imino acid); hydrophobic amino acid that does not readily fit into orderly secondary structure (α-helix)
   3. Cysteine – R group has reactive —SH; forms disulfide (—S—S—) bridge with other cysteines often at some distance away in polypeptide backbone or in another chain
      a. Stabilize proteins especially outside cells where they get added chemical & physical stress

IV. Not all of amino acids are found in all proteins, nor are the various amino acids distributed in an equivalent manner

V. A number of other amino acids also seen in chains (thyroxine, hydroxyproline, hydroxylysine)

A. Result from R group alteration of 20 basic aminos *after* their incorporation into polypeptide; they are called posttranslational modifications (PTMs)

B. Dozens of different types of PTMs have been documented, the most widespread of which is the covalent addition of a phosphate group to a serine, threonine or tyrosine residue
   1. PTMs can generate dramatic changes in the properties & function of a protein, most notably by modifying its interactions with other molecules or shortening its lifespan in cell
   2. The presence or absence of a single phosphate group on a key regulatory protein has the potential to determine whether or not a cell will behave as a cancer cell or normal cell
   3. Because of PTMs, a single polypeptide can exist as a number of distinct biological molecules

VI. Character of amino acid R groups (ionic, polar, nonpolar) is very important to protein structure & function; side chains also affect solubility (amino acids can be separated on basis of solubility)
   A. Most soluble (i.e., nonmembrane) proteins set up so polar residues are on molecule surface
      1. They associate with surrounding H_2O & contribute to protein solubility in aqueous solution
   B. Nonpolar residues situated predominantly in core of protein
      1. These hydrophobic residues are usually tightly packed together in 3-D puzzle excluding H_2O
      2. Contribute substantially to overall stability of protein; driving force during protein folding
      3. Reactive polar groups can project into nonpolar interior giving protein its catalytic activity
      4. Nonpolar environment can enhance ionic interactions between charged groups that would be lessened by competition with water in aqueous environment
      5. Reactions that might proceed very slowly in H_2O occur in thousandths of second within protein

VII. Conjugated proteins - involve another type of molecule attached covalently or noncovalently to protein; they include:
   A. Nucleoproteins - protein + nucleic acids
   B. Lipoproteins - protein + lipids
   C. Glycoproteins - protein + carbohydrate
   D. Various low-molecular-weight materials, like metals & metal-containing groups, often attached

VIII. Proteins are good illustration of intimate relationship between form & function
   A. The structure of most proteins is completely defined & predictable
   B. Each amino acid in one of these giant macromolecules is located at a specific site within the structure, giving the protein the precise shape & reactivity required for the job at hand
   C. Protein structure described at several levels of organization – each emphasizes a different aspect & each is dependent on different types of interactions
      1. 4 such levels are described: primary, secondary, tertiary & quaternary
      2. The first, primary structure, concerns amino acid sequence of a protein; the latter 3 levels concern the organization of the molecules in space

Proteins: Levels of Protein Structure – Primary & Secondary Structure

I. Primary (1°) structure - specific linear sequence of amino acids in chain; all levels of structure are ultimately determined by the primary level
   A. Number of chains that can be made = 20^n, where n = number of amino acids in chain; most polypeptides have >100 aminos (some several 1000); variety of possible sequences is unlimited
   B. Genome contains instructions for building them (precisely specifies amino acid sequence)
   C. Amino acid sequence contains most, if not all, information needed to specify protein 3D shape & thus its function; changes in sequence resulting from mutation may not be readily tolerated
      1. Example: sickle cell anemia - single change in amino acid sequence in hemoglobin molecule; valine replaces glutamic acid (nonpolar amino acid replaces charged, polar amino acid)
      2. Changes in amino acid sequence caused by changes (mutations) in DNA; problems with red blood cell shape & decreased O_2-carrying capacity result from this change
      3. Not all changes as big as above; related organisms show variations in sequence of same protein
   D. Degree to which changes in primary sequence are tolerated depends on degree to which protein shape or critical functional residues are disturbed
   E. Now know sequences of tens of thousands of proteins – first was protein hormone insulin determined by Sanger & coworkers, Cambridge, early 1950s
      1. Beef insulin chosen for its availability & small size (2 polypeptide chains of 21 & 30 aminos)

2. Insulin sequencing showed that proteins have a definable substructure that is neither regular nor repeating; each particular protein has precise amino acid sequence that does not vary
3. With DNA sequencing, primary structure can be deduced from the nucleotide sequence of encoding gene
4. Complete sequences of genomes of hundreds of organisms have been revealed; this will eventually allow researchers to learn about every protein an organism can make

II. Secondary (2°) structure - describes polypeptide conformation (spatial organization) chain portions; preferred ones provide maximum possible number of H bonds between neighboring amino acids
   A. α-helix - backbone assumes form of cylindrical, twisting spiral; backbone inside helix, R groups project outwards (Linus Pauling & Robert Corey, Cal Tech proposed both α- & β-structures)
      1. Stabilized by H bonds between atoms of one peptide bond & those above & below it in spiral; H bonds parallel to molecular axis
      2. Seen in X-ray diffraction patterns of actual proteins in 1950s; found in keratin from hair & various oxygen-binding proteins like myoglobin & hemoglobin; proof
      3. Opposing surfaces of α-helix may have contrasting properties – in water-soluble proteins, often polar amino acids are on outside of helix & nonpolar R groups facing inward
      4. Since it is coiled & held together by weak, noncovalent bonds, an α-helix can be extended in length if subjected to pulling forces
         a. Example is wool (mostly α-helix) - H bonds break if pulled, stretching fibers; when tension is relieved, H bonds reformed & fiber shortens to original length
         b. Human hair is less extensible because it is also stabilized by disulfide bridges
   B. β-pleated sheet- consists of several polypeptide segments lying side-by-side; the backbone of each segment of polypeptide adopts a folded or pleated conformation
      1. Characterized by a large number of H bonds perpendicular to polypeptide chain long axis; project across from one part of chain to another
      2. β strands highly extended; sheet resists pulling (tensile) forces; very strong (ex.: silk fibroin)
      3. A single fiber of spider silk (one-tenth the thickness of a human hair, is roughly 5 times stronger than a steel fiber of comparable weight
      4. Spider silk being produced from cultured epithelial cells; hoping to use them in making strong, lightweight, resilient products like bulletproof vests
   C. Chain portions not organized into α-helix or β-pleated sheet may consist of hinges, turns, loops, fingerlike extensions; often most flexible portions of chain & sites of greatest biological activity
      1. Antibody interactions with antigens are very specific & mediated by a series of loops at one end of antibody

Proteins: Levels of Protein Structure – Tertiary Structure

I. Tertiary (3°) structure is the conformation of entire protein; results from (**intramolecular**) noncovalent interactions between R groups in same chain; virtually unlimited number of structures unlike limited 2°
   A. X-ray crystallography can be used to determine tertiary structure; ~20,000 3D structures already reported, increased pace in structure discovery per year
      1. In X-ray crystallography, protein crystal is bombarded by thin X ray beam
      2. Radiation scattered (diffracted) by atoms of protein is allowed to strike radiation-sensitive plate or detector
      3. Forms an image of spots, pattern of which is subjected to complex mathematical analysis
      4. Investigator can work backward to derive structure responsible for producing pattern; if more information in patterns is analyzed, a more detailed picture of the molecule is obtained

5. Most newly discovered proteins evolutionarily related to those already determined so some idea of 3-D organization immediately available from amino acid sequence

B. NMR spectroscopy (not described here) – 3D structure of small proteins (<30 kDa) can also be determined by NMR spectroscopy

C. It has become evident recently, as more & more protein structures have been solved, that a surprising number of proteins contain sizable segments that lack a defined conformation
   1. These unstructured or disordered segments of proteins can be present in many different positions & thus cannot be studied by X-ray crystallography
   2. Disordered regions have key roles in vital cell processes (often bind DNA or other proteins); may seem strange that regions lacking fully defined structure can engage in useful function
   3. Remarkably, these segments often undergo a physical transformation once they have bound an appropriate partner & are then seen to possess a defined, folded structure

D. Fibrous proteins (highly elongated shape) - long strands or flattened sheets that resist pulling or shearing forces to which they are exposed; structural materials outside cell are usually these
   1. Collagens & elastins of connective tissue, keratins (hair, skin, fingernails), silk

E. Globular proteins – most proteins in cell; compact shape; chains folded & twisted into complex shapes; distant points brought next to each other, linked by various types of bonds; ex.: myoglobin
   1. Myoglobin - storage site for O_2 in muscle tissue; reported by John Kendrew, Cambridge (1957)
   2. Oxygen binds to Fe atom at heme group center (gives muscle red color), a prosthetic group (portion of protein not made of amino acids that is added to polypeptide after translation)
   3. First report low resolution – compact (globular), complex, folds back on itself; no regularity or symmetry indicated, 8 rodlike α-helix sections (7-24 aa long; ~75% of chain); no β-pleated sheet
   4. Later X-ray analysis – heme group in pocket of hydrophobic R groups; promotes O_2 binding without Fe oxidation; no disulfide bonds; 3° structure totally result of noncovalent interactions
   5. All of the noncovalent bonds thought to occur between R groups within proteins (H bonds, ionic bonds, hydrophobic interactions) have been found
   6. Unlike myoglobin, most proteins have both α-helix & β-pleated sheet; triosephosphate isomerase is mostly β-sheet
   7. Most importantly, these early landmark studies revealed that each protein has a unique tertiary structure that can be correlated with its amino acid sequence & its biological function

II. Unlike myoglobin, most eukaryotic proteins made of ≥2 spatially distinct modules (**domains**) that fold independent of one another; often represent parts that function in semi-independent manner

A. May bind various things (coenzyme & substrate; DNA strand & another protein) or move relatively independent of one another

B. Proteins with >1 domain may have arisen during evolution by fusion of genes coding for different ancestral proteins; each domain representing a part that once was separate molecule
   1. Mammalian phospholipase C - each domain identified as homologous unit in other protein

C. Some domains have been found only in one or a few proteins; other domains have been shuffled widely about during evolution
   1. These widely shuffled domains appear in variety of proteins whose other regions show little or no evidence of an evolutionary relationship
   2. Domain shuffling creates proteins with unique combinations of activities
   3. On average, mammalian proteins tend to be larger & contain more domains than proteins of less complex organisms (fruit flies, yeast)

III. In the last few years, there has been an explosion of interest in protein structure

A. ~25,000 3-D protein structures have already been reported – recent advances in protein crystallization & X-ray diffraction technology has led to increased pace in structures reported

1. Vast majority of newly described structures are similar to those of proteins whose structure has already been determined

B. Proteins (or the domains of which they are composed) can be grouped into families whose members have overall structural similarity
   1. Domains grouped in same family have backbone folded into roughly same configuration; said to share common fold; structural biologists want to describe all the various folds in nature

C. There is considerable disagreement as to how many different families comprise protein universe; estimates range from <2,000 to >10,000
   1. Protein Structure Initiative mounted to solve structures of as many different folds as possible over next few years
   2. Once representative structure of given fold determined, other family members' structures can usually be predicted from amino acid sequence with improved computer modeling techniques

IV. Dynamic changes within proteins - proteins not rigid, inflexible; capable of internal movements; with NMR

A. Since so tiny, greatly influenced by energy of their environment; random small-scale fluctuations in protein bond arrangement create incessant thermal motion in molecule seen with NMR

B. Spectroscopic techniques like NMR can monitor dynamic movements within proteins; reveal H bond shifts, external side chain waving, full aromatic ring rotation about single bonds (phe, tyr)

C. Ex.: acetylcholinesterase – degrades acetylcholine left behind after neurotransmission
   1. When structure first revealed by X-ray diffraction, there was no obvious path for acetylcholine to enter catalytic site at bottom of deep gorge in molecule
   2. Narrow entrance to active site blocked by a number of bulky amino acid side chains
   3. Computer analysis showed that dynamic movements of side chains in protein would lead to rapid opening & closing of gate allowing acetylcholine to diffuse into active site

V. Conformational changes - predictable (nonrandom) movements triggered by specific stimulus & associated with function; accompany virtually every activity in which a protein takes part

A. Ex.: muscle myosin binds actin & small 20° head rotation moves adjacent actin filament 50 - 100Å

B. Ex.: Bacterial protein (GroEL) as it interacts with another protein GroES

Proteins: Levels of Protein Structure – Quaternary Structure & Multiprotein Complexes

I. Quaternary (4°) structure is the linking of polypeptide chains to form multisubunit functional protein via **intermolecular** R group interactions

A. May be linked by disulfide bonds, but more often noncovalent bonds (hydrophobic, H bonds, etc.) like hydrophobic patches on complementary surfaces of neighboring polypeptides

B. Chains may be identical or nonidentical
   1. Protein composed of 2 identical subunits - homodimer
   2. Protein composed of 2 nonidentical subunits - heterodimer

C. Ex.: hemoglobin (best studied); Max Perutz, Cambridge (1959) - 2 α-globin and 2 β-globin subunits; each binds one molecule of O_2
   1. Subunits like myoglobin; probably evolved from common ancestral O_2-binding polypeptide
   2. O_2-binding was accompanied by movement of bound iron atom closer to plane of heme group
   3. Shift in iron atom position pulled on a-helix to which iron is connected; this, in turn, leads to series of increasingly larger movements within & between subunits
   4. Finding revealed that complex functions of proteins may be carried out by means of small changes in conformation

II. Multiprotein complexes – different proteins, each with a specific function, become physically associated to form a much larger complex
   A. Example - *E. coli* pyruvate dehydrogenase complex - 60 polypeptide chains constituting 3 different enzymes; stable
      1. Its enzymes catalyze reaction series connecting 2 metabolic pathways: glycolysis & TCA cycle
   B. Because they are so closely associated, the product of one enzyme can be channeled directly to next enzyme in sequence; prevents dilution in cell's aqueous medium
   C. Some associations stable; some not (transient; dynamic; associate, dissociate based on conditions); have complementary surfaces (part of one fits in pocket on other); stabilized by noncovalent bonds
      1. General rule – most proteins interact with other proteins in highly dynamic patterns;
      2. Ex.: SH3 domain – part of many different proteins involved in molecular signaling; act like knobs that allow binding to proteins with complementary handle
         a. In this case, handle is rich in proline (a polyproline motif)
   D. A number of different structural domains like SH3 identified that function as adaptors to mediate interactions between proteins
      1. Often interactions between proteins are regulated by changes like phosphate addition to key amino acid; may greatly increase or decrease its ability to bind a protein partner
      2. Transient protein interaction important in DNA synthesis, ATP formation, RNA processing, etc.; done by molecular machines made of many interacting proteins (transient or stable)
   E. How to determine whether two proteins interact physically – the yeast two-hybrid (Y2H) system
      1. Genes encoding the 2 proteins being studied are introduced into same yeast cell
      2. If yeast cell tests positive for particular reporter protein, which is indicated by obvious color change in yeast cell, then the 2 proteins in question had to have interacted in yeast cell nucleus

III. In recent years, a number of research teams have attempted to study protein-protein interactions on a global scale, rather than one at a time as generally has been done
   A. One might want to know all of the interactions occurring among the 14,000 or so proteins encoded by fruit fly *Drosophila melanogaster* genome, which has now been fully sequenced
      1. Virtually every gene within genome is available as individual DNA segment that can be cloned & used as desired
      2. Thus, it should be possible to test the proteins encoded by the fly genome, two at a time, for possible interactions in a Y2H assay
      3. The first such study reported that, of the millions of possible combinations, >20,000 interactions were detected among 7048 proteins tested
   B. Although the Y2H has been the mainstay in the study of protein-protein interactions for >15 years, it is an indirect assay & it is fraught with uncertainties
      1. A large percentage of interactions known to occur between specific proteins fail to be detected in these experiments; the assay gives a significant number of false negatives
      2. Assay also generates large number of false positives; it indicates that 2 proteins are capable of interacting when other studies show they do not do so under normal cellular conditions
   C. Fruit fly protein study – authors used computer analyses to narrow findings from the original 20,000 interactions to ~5000 interactions in which they have high confidence
      1. Overall, it is estimated that, on average, each protein encoded in a eukaryotic organism's genome interacts with ~5 different protein partners
      2. According to this estimate, human proteins would engage in roughly 150,000 different interactions

IV. What do we learn about cellular activities from large-scale protein-protein interaction studies other than obtaining a long list of potential interactions?

A. Most importantly, they provide a guideline for further investigation
B. Genome-sequencing studies have provided amino acid sequences of many previously unknown proteins – can determine a protein's function by identifying the proteins with which it associates
    1. If known protein has been shown to be involved with DNA replication & unknown protein interacts with known protein, then unknown one probably is part of replication machinery
    2. Thus, regardless of their limitations, large-scale Y2H studies & others provide a starting point to explore a large number of previously unknown protein-protein interactions
    3. Each of these interactions has the potential to lead investigators to a previously unknown biological process

IV. Chart below summarizes features and definitions of the four levels of protein structure:

Level of Structure	Definition	Bonds Involved	Comments
Primary (1°)	Absolute sequence of amino acids from amino end to carboxyl end	Peptide bonds	All 3 higher levels are direct consequences of 1° structure (contains information about their final shape). Changes can lead to disease (ex. sickle cell) or little or no effect.
Secondary (2°)	Results from interactions between backbone portions of adjacent or nearly adjacent amino acids	H bonds	α-helix - spiral shaped; H bonding maximal and parallel to main molecular axis of helix; allows extensibility (ex.: wool & human hair). β-pleated sheet - highly flattened, extended sheetlike shape; H bonding maximal and perpendicular to main molecular axis; strong and flexible (ex.: silk fibroin). Without α or β structure adopts hinges, turns, loops or fingerlike extensions with most biological activity.
Tertiary (3°)	Results from interactions within a single chain between R groups or between R groups at a distance and backbone	H bonds, disulfide bonds, van der Waals forces, ionic bonds, hydrophobic interactions	Proteins fibrous (highly elongated like collagen) or globular (myoglobin). Protein domains - compact regions functioning semi-independently (linked by flexible part of chain serving as hinge). Protein motifs - recurring protein substructures with certain functions. Proteins flexible & can change shape.
Quaternary (4°)	Results from R group interactions between multiple protein chains (subunits) which form a functional protein unit	H bonds, disulfide bonds, van der Waals forces, ionic bonds, hydrophobic interactions	Assembly spontaneous & usually bound together by noncovalent bonds (electrostatic or hydrophobic). Homodimers - 2 identical subunits; heterodimers - at least 2 nonidentical subunits (Ex.: hemoglobin)

Proteins: Protein Folding

I. Christian Anfinsen, NIH (1956) – studied properties of ribonuclease A (RNase; small enzyme consisting of 1 chain; 124 aa; 4 disulfide [diS] bonds)
   A. To denature (unfold) RNase —> break diS bonds with mercaptoethanol (MerETOH [CH_3CH_2SH]; a reducing agent) which converts disulfide bridge to a pair of sulfhydryl (—SH) groups
      1. To make all disulfide links accessible to MerETOH had to first unfold protein (**denaturation**)
      2. Can use detergents, organic solvents, radiation, heat, urea, guanidine chloride to break up various interactions stabilizing 3° structure & denature protein
      3. Treat RNase with concentrated urea & MerETOH -> enzyme activity disappears (it's unfolded)
   B. When he removed urea & MerETOH -> normal enzyme activity returned
      1. The reformed molecules were indistinguishable structurally & functionally from native ones at beginning of experiment
      2. Concluded that amino acid sequence contained all information needed for self-assembly of 3D-conformation

3. Events tend to progress toward states of lower energy so 3° structure assumed after refolding is the accessible structure with lowest energy
4. It is the most thermodynamically stable structure possible that can be formed by that chain

II. Numerous controversies in study of protein folding, one concerns the types of events that occur at various stages during the folding process; describe simple proteins (RNase) with single domain
   A. Scheme 1
      1. Protein folding is initiated by interactions among neighboring residues that lead to the formation of much of the secondary structure of the molecule
      2. Once α helices & β sheets are formed, subsequent folding is driven by hydrophobic interactions that force nonpolar residues together in the central core of the protein
   B. Alternate scheme
      1. First major event in protein folding is collapse of polypeptide to form a compact structure
      2. Only then, does significant secondary structure develop
   C. Recent studies suggest that the two pathways above lie at opposite extremes with most proteins probably folding by a middle-of-the-road scheme
      1. Secondary structure & formation & compaction occur simultaneously in this scheme
      2. These early folding events lead to formation of partially folded, transient structure that resembles native protein
      3. The transient structure formed lacks many of the specific interactions between amino acid side chains that are present in the fully folded molecule

III. If information that governs folding is embedded in protein's amino acid sequence, then alterations in sequence have potential to change the way a protein folds
   A. This change leads to a molecule with an abnormal tertiary structure
   B. Many mutations responsible for inherited disorders have been found to alter a protein's 3-D structure; in some cases, consequences of protein misfolding can be fatal

IV. Molecular chaperones - help proteins to fold into final 3D conformation; nonspecific (one can work on many proteins); prevent proteins
   A. Not all proteins can assume final 3° structure by simple process of self-assembly
   1. 1° structure does not lack required information for proper folding
      2. Proteins undergoing folding must be prevented from interacting nonselectively with other cell molecules
   B. Several protein families have evolved whose function is to help unfolded or misfolded proteins achieve proper 3-D conformation; called **molecular chaperones**
   C. Selectively recognize & bind short stretches of exposed hydrophobic amino acids that tend to be exposed in non-native proteins but buried in proteins having native conformation
   D. Polypeptides are made on ribosomes by adding amino acids beginning at chain's N-terminus
      1. Chaperones of Hsp70 family bind to elongating polypeptide chains as they exit ribosome
      2. Hsp70 chaperones thought to prevent these partially formed (nascent) polypeptides from binding to other proteins in cytosol that would cause them either to aggregate or misfold
      3. Once synthesis is done, many are simply released from chaperones into cytosol where they spontaneously fold into native state
   E. Many of larger polypeptides are transferred from Hsp70 proteins to different type of chaperone called chaperonin
      1. Chaperonins are cylindrical protein complexes with central cavity in which newly synthesized polypeptides can fold without interference from other cell macromolecules

2. Best studied chaperonin from eukaryotic cytosol is TriC; thought to assist in folding of up to 15% of polypeptides made in mammalian cells

F. Most abundant & poorly understood chaperones are members of Hsp90 family
1. Until recently, the only demonstrated role of Hsp90 chaperones was to interact with a small number of specific proteins
2. Then fruit flies were found carrying a mutant Hsp90 chaperone; they developed into adults that had a wide variety of abnormal traits
3. Plants with Hsp90 mutations also had bizarre abnormalities
4. It was concluded that animals & plants carry mutations that would produce abnormalities if not for the presence of Hsp90 chaperones
5. Hypothesized that these hidden mutations encode abnormal proteins that can be stabilized in native conformation by Hsp90 chaperones
6. As long as Hsp90 chaperones are fully operational, these mutant proteins function normally & organism develops without disturbance
7. If Hsp90 chaperones are incapacitated, then flies exhibit abnormal phenotype whose nature depends on specific genes that are mutated
8. If this is correct, chaperones could play important role in evolution by allowing hidden genetic variation to accumulate in a population
9. Under stressful conditions (elevated environmental temperature), demand for Hsp90 would increase as normal proteins become destabilized
10. As more Hsp90 proteins get tied up stabilizing heat-stressed normal proteins, then fewer of the chaperones are available to stabilize mutant proteins
11. Hidden reservoir of genetic variation might be released leading to the sudden appearance of individuals with novel phenotypes
12. When fruit flies with normal Hsp90 protein are raised at higher temperature, a significant number do show abnormal phenotypes as would be expected by this hypothesis
13. A small proportion of such individuals might be better adapted to the new stressful conditions & thus favored by natural selection

The Emerging Field of Proteomics

I. Human genome contains roughly 25,000 genes, each of which can potentially gives rise to a variety of different proteins, only a fraction of which have been characterized

A. One gene can give rise to more than one polypeptide in a number of ways – 2 prominent mechanisms
1. Alternative splicing
2. Posttranslational modification

B. It should also be noted that many proteins have more than one distinct function
1. Myoglobin, long studied as an oxygen-storage protein, has recently been shown to be involved in the conversion of nitric oxide (NO) to nitrate (NO_3^-)

B. If all cells in body, at all stages in life, are taken together as a group, humans may synthesize >100,000 different protein molecules (only a very small percentage have been characterized)

II. **Proteome** – entire collection of proteins produced by an organism (human or otherwise)

A. Also applied to complete set of proteins present in a particular tissue, cell or cell organelle
1. Once an entire genome has been sequenced, genes present in that genome become available to produce the encoded proteins
2. Thus, scientists have at their disposal a huge number of proteins whose properties & functions are largely unknown

3. Techniques have been developed that allow the determination of the properties & activities of a large number of proteins at the same time

B. The term **proteomics** was coined to describe the expanding field of protein biochemistry
   1. Advanced technologies & high speed computers are being used to perform systematic, large-scale studies on diverse protein mixtures
   2. Trying to separate & identify the large numbers of proteins made in cells & tissues of complex tissues & answer a host of difficult questions about each protein

C. Proteomics is inherently more difficult than genomics because proteins are more difficult to work with than DNA
   1. One gene is the same as all others; proteins have unique chemical properties & handling requirements

D. In addition, small quantities of a particular DNA segment can be expanded greatly using readily available enzymes, whereas protein quantities cannot be increased
   1. This is particularly troublesome when one considers that many of the proteins regulating important cellular processes are present in only a handful of copies per cell

II. Two-dimensional polyacrylamide gel electrophoresis (invented in mid-1970s) – one of the best ways to separate large numbers of proteins in a mixture; difficult to identify each one

A. In the past few years, mass spectrometry & high-speed computation have together made it possible to identify any or all of the proteins present on a gel

B. Mass spectrometry – technique to determine the precise mass of a molecule or fragment of a molecule, which can be used to identify the nature of that molecule
   1. Remove spot from gel & digest it with trypsin, which cleaves polypeptides into peptides that have either a lysine or arginine residue at their carboxyl terminus
   2. Introduce peptides into mass spectrometer —> they are separated according to mass & displayed as series of peaks of known molecular mass
   3. The pattern of peaks constitutes a highly characteristic peptide mass fingerprint of protein

C. Advances in computer technology
   1. Once a genome has been sequenced & a large percentage of the embedded genes identified, then amino acid sequences of encoded proteins can be predicted
   2. These sequences can be subjected to a theoretical trypsin digestion & the masses of the resulting peptides can be calculated & entered into a database
   3. Actual peptide masses of purified protein obtained by mass spectrometer compared with high-speed supercomputers to masses predicted by theoretical digests of all genome polypeptides
   4. Protein isolated & subjected to mass spectrometry can be directly identified based on database search
   5. Mass spectrometers are not restricted to handling one purified protein at a time, but are also capable of analyzing proteins present in complex mixtures
   6. Mass spectrometry - very useful in revealing how protein complement of cell/tissue changes over time as might happen after hormone secretion in body, taking drug or during disease

III. Pharmaceutical industry is rushing to identify & patent proteins that play a role in human disease using proteomics

A. Compare proteins in blood or tissues of healthy individuals with those suffering from particular disease – example is study by a team at Nat'l Cancer Institute (2002)
   1. Used mass spectrometry to compare proteins in blood of patients with ovarian cancer to those in blood of normal, healthy individuals —> could distinguish the 2 groups
   2. The serum proteome of women with ovarian cancer showed a distinctive pattern of peaks representing numerous proteins of unknown identity

B. Based on these findings, a simple noninvasive test has been developed (OvaCheck) that may be able to detect this deadly cancer at an early stage
   1. Such a test would have far-reaching consequences since ovarian cancer is seldom caught at early stage when it is curable
C. While test was being developed, a number of proteomics researchers took an independent look at the original findings presented in publications & online postings
   1. They concluded that the data had been misinterpreted
   2. Instead, they found no clinically significant differences in the mass spectral profiles reported between serum of healthy people & those suffering from ovarian cancer (or other cancers)
   3. Issue will only be decided as validity of OvaCheck & other proteomics-based diagnostic tools is analyzed in large-scale clinical settings
D. Despite controversy, many clinical researchers have gut feeling that most human diseases leave telltale patterns (biomarkers) among the 1000s of proteins found in human serum
   1. International effort (Plasma Proteome Project) has been organized to study the protein patterns in serum samples obtained from patients suffering from a host of disorders
   2. One day, it might be possible that a single blood test could reveal the existence of early-stage heart, liver or kidney disease that could be treated before it became life-threatening condition
E. Once such proteins are found, their role in disease development can be investigated
   1. Once a protein is determined to play role in causing disease, like activated ABL protein that causes certain leukemias, it may be possible to develop drugs that target that specific protein
F. Proteins are also potential pharmaceutical products – examples below & more will be developed
   1. Growth hormone is given to exceptionally small children to spur their growth
   2. Erythropoietin is given to patients undergoing chemotherapy to stimulate red blood cell production & prevent anemia

IV. Researchers are working to develop techniques that allow protein function to be determined on a large scale, rather than one protein at a time – example: protein microarrays (or protein chips)
A. Protein microarray uses solid surface (glass microscope slide) covered by microscopic-sized spots, each containing an individual protein sample
   1. Constructed by application of tiny volumes of individual proteins to specific sites on slide generating an array of proteins
   2. To obtain proteins to be used, individual genes can be isolated from genome & introduced into suitable host cells that will manufacture the encoded protein
   3. The 6000 or so yeast proteins encoded by yeast genome can fit comfortably as individual spots on single glass slide
B. Once protein microarray created, proteins on it can be screened for various types of activities
   1. Phosphoinositides (certain group of lipids) – play important role in transmitting signals from one part of cell to another
   2. Useful to know identity of all proteins in cell that can bind to a particular phosphoinositide like PI $(3,4)P_2$
   3. Incubate microarray with PI $(3,4)P_2$ & proteins that have bound the lipid are indicated by fluorescent spots
   4. Incubate same microarray with calmodulin, which binds calcium ions & is key mediator in many activities triggered by rise in calcium ion concentration
   5. Proteins that bind calmodulin are likely to be proteins involved in calcium signaling activity; turned up 33 new calmodulin-binding proteins
C. Thus microarrays can be used to identify proteins involved in one or another type of critical cellular activity or they may be valuable in screening new drugs
   1. Inhibitors of an enzyme might bind proteins other than target enzymes; could be determined by protein microarray

2. If inhibitor binds to more proteins, it would be more likely to produce undesirable side effects in patients

D. Protein chips may be used by clinical labs to screen for proteins characteristic of particular disorders; proteins would most likely be identified by mass spectrometry as described with ovarian cancer
   1. Use antibodies to identify presence of protein associated with disease in blood or urine sample, like PSA test for prostate cancer
   2. PSA is protein found in blood of normal men, but it is elevated in those with prostate cancer
   3. Soon microarrays will be available with antibodies capable of detecting presence & amount of blood proteins that indicate a person may be suffering from one of a variety of diseases

Protein Engineering

I. Scientists can now design & mass produce novel proteins that are different from those made by living organisms
   A. Can make artificial gene to be used in making protein with any desired amino acid sequence; can also make polypeptides from scratch in lab using chemical techniques
      1. This latter strategy allows researchers to incorporate building blocks other than the 20 amino acids that normally occur in nature
      2. The problem with these engineering efforts is knowing which of the virtually infinite variety of possible proteins one could make might have some useful function
      3. Example: manufacture protein that would bind to AIDS virus surface & remove it from aqueous solution, like bloodstream
      4. Assume computer simulations could predict shape such a protein should have to bind virus surface; requires detailed insight into rules governing relationship between 1° & 3° structure
   B. Can now synthesize polypeptides that fold into small domains with relatively simple 2° structures (like bundles of α-helices and/or 2 – 3 stranded β-sheets)
      1. Creation of more complex polypeptide structures from scratch has proven more difficult
      2. It is hard to put particular sequence into context of entire protein
      3. Same sequence or similar stretch can form different 2° structures in different proteins
      4. Depends on external influences from other regions of molecule

II. Another approach is to modify proteins already made by cells
   A. Can isolate human gene, alter its sequence in precise way & synthesize modified protein with altered amino acid sequence (**site-directed mutagenesis**); used to study protein structure
      1. If want to know role of particular residue on protein folding or function, gene can be modified in specific way
      2. Substitute amino acid with different charge, hydrophobic character or H-bonding properties —> assess effect of substitution on modified protein structure & function
      3. Site-directed mutagenesis is proving invaluable in analysis of specific function of minute parts of many proteins of interest to biologists
   B. Site-directed mutagenesis is also used to modify structure of clinically useful proteins to bring about various physiological effects; example:
      1. The drug Somavert (approved by FDA in 2003) is modified version of human growth hormone (GH) containing several alterations
      2. GH normally acts by binding to a receptor on the surface of target cells, which triggers a physiological response
      3. Somavert competes with GH in binding to the GH receptor, but interaction between drug & receptor fails to trigger the cellular response
      4. Somavert is prescribed for the treatment of acromegaly, a disorder that results from excess production of growth hormone

III. Structure-based drug design - develop new drugs that act by binding known proteins & thus inhibiting their activity
   A. Expose protein being targeted to combinations of organic compounds made over years by drug companies or isolated from plants or microorganisms (there are millions of them)
      1. Determine which ones, if any, happen to bind to the protein with reasonable affinity
      2. Alternative if protein 3° structure is known – use computers to design virtual drug molecules whose size & shape might allow them to fit into protein cracks & crevices, making it inactive
   B. Illustrate both technologies by looking at development of drug Gleevec; it has revolutionized treatment of some relatively rare cancers, most notably chronic myelogenous leukemia (CML)
      1. Tyrosine kinases (TKs) are often involved in transformation of normal cells into cancer cells
      2. TKs catalyze addition of phosphate groups to specific tyrosine residues within target protein; this may activate or inhibit the target protein
      3. CML development is driven almost single-handedly by the presence of an overactive TK called ABL
   C. During 1980s, researchers identified compound called 2-phenylaminopyrimidine that was able to inhibit TKs
      1. Compound was discovered by randomly screening a large chemical library for compounds that exhibited this particular activity
      2. As is usually the case in blind screening experiments, 2-phenylaminopyrimidine would not have made a very effective drug
      3. First, it was only a weak enzyme inhibitor, which meant it would have had to be used in very large quantities
      4. 2-phenylaminopyrimidine is described as a lead compound, a starting point, from which usable drugs might be developed
      5. Starting with this lead molecule, compounds of greater potency & specificity were synthesized using structure-based drug design
      6. One compound to emerge from this process was Gleevec; it was found to bind tightly to ABL TK inactive form & prevent it from being activated (necessary for cell to become cancerous)
   D. Preclinical studies demonstrated that Gleevec strongly inhibited the growth in the lab of cells from CML patients & that the compound showed no harmful effects in animal tests
      1. In the very first clinical trial of Gleevec, all 31 CML patients went into remission after taking once-daily doses of the compound

Protein Adaptation and Evolution

I. Adaptations – traits that improve likelihood of an organism's survival in a particular environment
   A. Proteins are biochemical adaptations subject to natural selection & evolutionary change
   B. Best revealed by comparing homologous proteins from organisms living under very different environmental conditions —> show different adaptations
      1. Proteins in halophilic (salt-loving; up to 4 M KCl) archaebacteria have amino acid substitutions allowing them to maintain their solubility/function at very high cytosolic [salt]
      2. Halophilic version of malate dehydrogenase has its surface coated with aspartic & glutamic acid residues – these residues' carboxyl groups compete with salt for water interactions

II. The fundamental event in protein evolution is a random change in amino acid sequence
   A. Many, if not most, amino acid substitutions seen in proteins over time are neutral; these are changes that do not affect the fitness of the molecule
   B. A change from one nonpolar amino to another nonpolar amino in protein core will not usually have significant effect on either structure or function of molecule

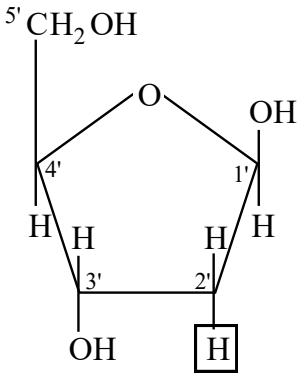
C. Homologous proteins from different organisms can have virtually identical shapes & folding patterns & strikingly divergent amino acid sequences
   1. The greater the evolutionary distance between 2 organisms, the greater the difference in the amino acid sequences of their proteins
   2. Sometimes, a few key aminos in particularly critical portion of protein is present in all organisms in which that protein has been studied
   3. In one comparison 226 globin sequences, only 2 residues were found to be absolutely conserved in all of them (one is histidine important in O_2 binding & release)
   4. These observations indicate that 2° & 3° structures change more slowly during evolution than primary structures

III. Not only does evolution different versions of proteins in individual organisms, but it has also produced different versions of proteins in individual organisms
   A. Several different versions of proteins with a given function (globin, collagen) are encoded by the human genome
      1. Usually, different versions of a protein (known as isoforms) are adapted to function in different tissues or at different stages of development
      2. Humans have 6 different genes encoding isoforms of the contractile protein actin; 2 are found in smooth muscle, 1 in heart muscle & 2 in virtually all other cell types
   B. Most proteins are members of much larger families (or superfamilies) of related molecules
      1. Genes encoding various members of a protein family are thought to have evolved from single ancestral gene that underwent a series of duplications to make 2 or more copies
      2. Over long periods of evolutionary time, sequences of the various copies diverged from one another to generate proteins with related (homologous) sequences
      3. Many protein families contain a remarkable variety of proteins that have evolved diverse functions
      4. Expansion of protein families is responsible for much of the protein diversity encoded in the genomes of today's complex plants & animals

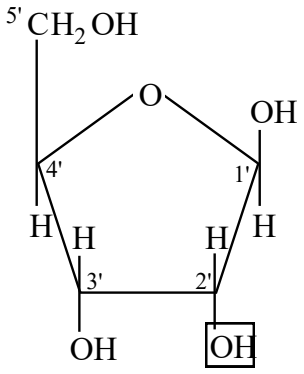
Nucleic Acids

I. Primarily involved in storage & transmission of genetic information; may also be structural or catalytic
   A. 2 types in living organisms – deoxyribonucleic acids (**DNA**) & ribonucleic acids (**RNA**)
      1. DNA – serves as genetic material of all cellular organisms; RNA plays that role in many viruses
      2. Information stored in DNA used to govern cell activities through formation of RNA messages
   B. Constructed as long chain (**strand**) of monomers called **nucleotides**
   C. Concentrate here on RNA as representative molecule; more complex, double-stranded DNA structure will be covered in Chapter 10

II. Both DNA & RNA are composed of nucleotides (phosphate + sugar + base) connected to form polymers (**polynucleotides**)
   A. Phosphate group (PO_4^-) - linked to 5'-carbon of sugar
   B. 5-carbon sugar - ribose or deoxyribose



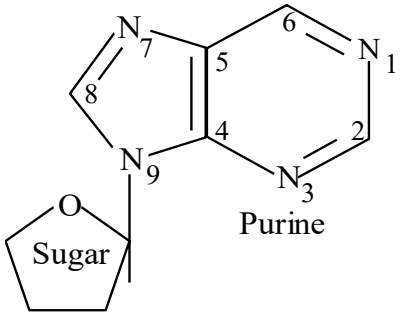
Deoxyribose

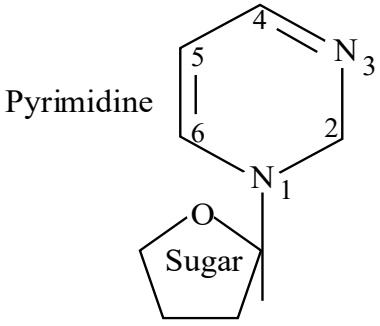


Ribose

C. Nitrogenous base (pyrimidine -1 ring or purine -2 rings) - rings contain nitrogen; linked to 1'-carbon of sugar (adenine, guanine, thymine [not in RNA], cytosine, uracil [not in DNA])
   1. Called nitrogenous bases because nitrogen atoms form part of the rings of the molecules
   2. **Purines** in RNA & DNA are **adenine** & **guanine;** larger structure, consisting of 2 rings
   3. **Pyrimidines** in RNA are **cytosine** & **uracil**; in DNA, **uracil** replaced by **thymine**, a pyrimidine with an extra methyl group attached to the ring; smaller, single ring structure

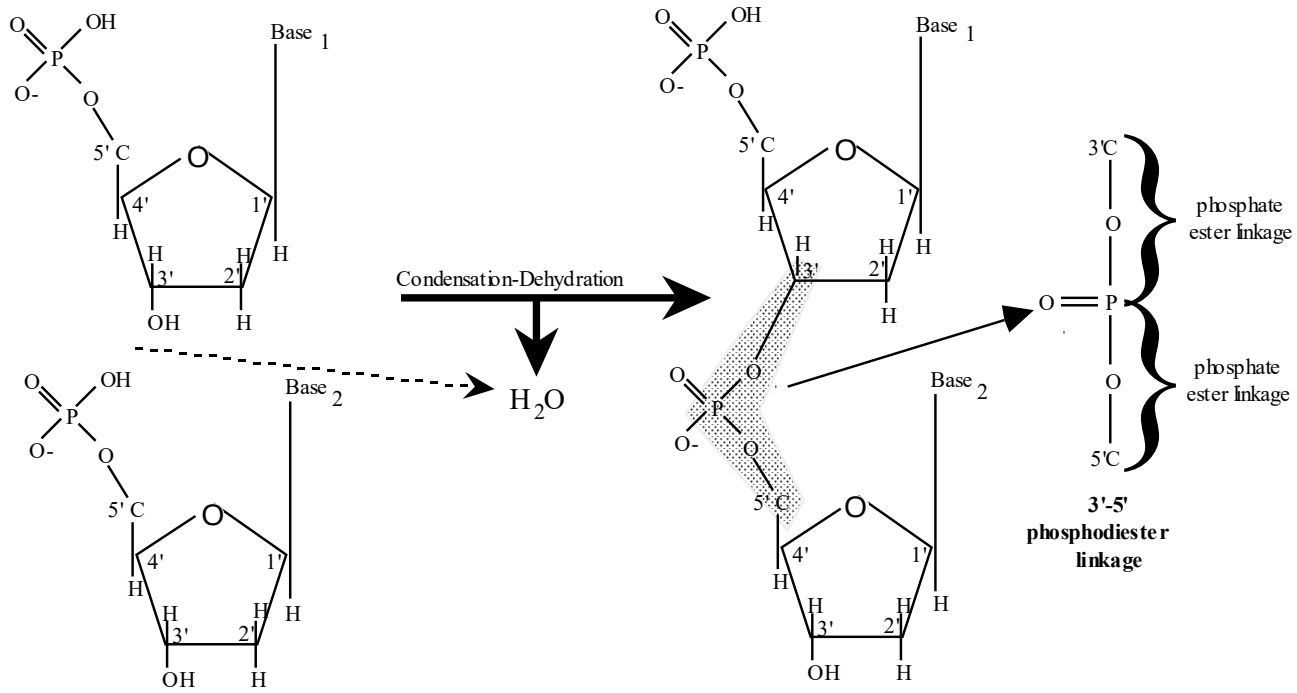
Generalized Stucture of Purines and Pyrimidines



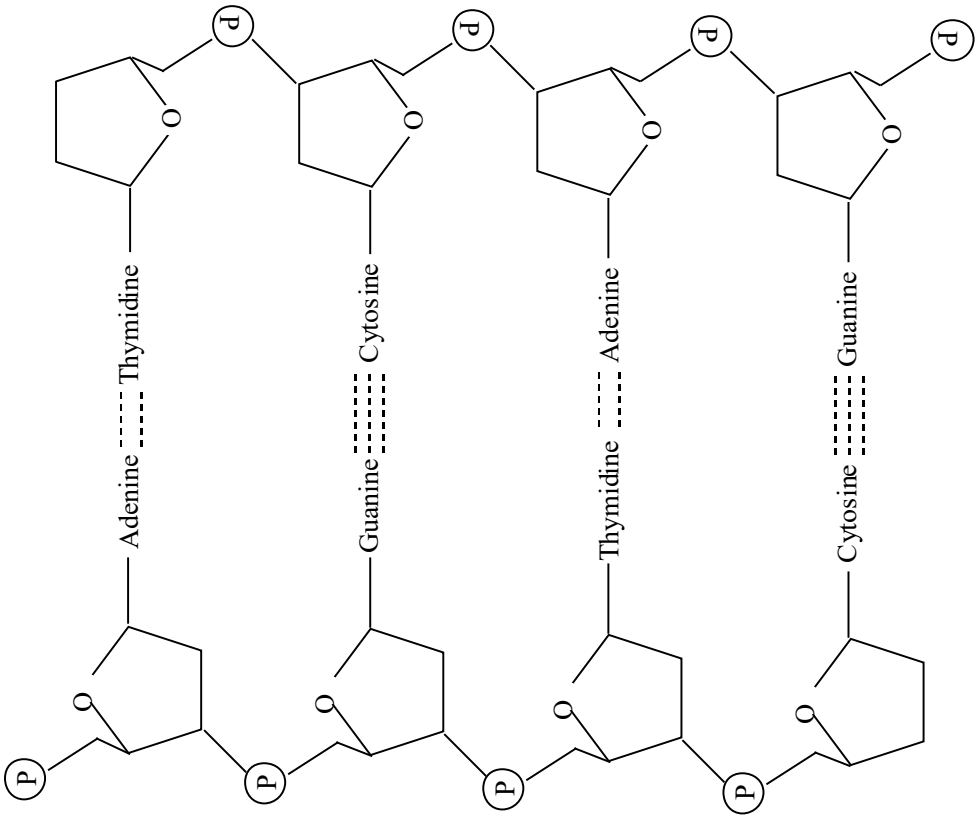


III. The sugar & nitrogenous base together form a **nucleoside**, so nucleotides of RNA strand are known as **ribunucleoside monophosphates**
   A. The phosphate is normally linked to the 5'-carbon of the sugar
   B. The nitrogenous base is attached to the sugar's 1'-carbon

IV. Monomers polymerize when sugar 3'-OH is linked by ester bond to 5'-phosphate of next nucleotide in chain (**3' - 5' phosphodiester linkage**); so nucleotides are joined by sugar-phosphate linkages
   A. Phosphates in backbone attached to 2 sugars by ester linkages (phosphorus atom linked to 2 oxygen atoms, one from each of the 2 adjoining sugars)
   B. Have hydrophilic, charged backbone (repeating P & sugar units) & nitrogenous bases as side groups
   C. Bases largely hydrophobic due to ring structure



V. General structure and function of DNA/RNA
   A. Sequence of bases determines specificity of DNA/RNA & encodes hereditary information for synthesis of proteins
   B. RNA is usually single-stranded, but can fold back on itself to form frequent double stranded regions with local H-bond pairing (just like in DNA) & complex 3° structures
      1. With DNA, it directs & carries out protein synthesis, but ……
      2. Some RNAs do not carry genetic information but can be structural and/or enzymatic (rRNA)



VI. Types of RNA - ribosomal RNA (rRNA), transfer RNA (tRNA), messenger RNA (mRNA)
- A. rRNAs do not carry genetic information but serve as structural scaffolds to which ribosome proteins can attach; recognize & bind various soluble components used in protein synthesis
- B. Some RNAs have catalytic activity (**ribozymes** or **RNA enzymes**); double-stranded regions held together by H bonds between the bases; also hold together DNA strands
  1. One of rRNAs of large ribosomal subunit catalyzes reaction by which amino acids are joined during protein synthesis); hammerhead ribozyme – cleaves its own RNA strand
  2. Recent x-ray crystallography studies of small RNAs show 3° structures that achieve a level of structural complexity approaching that of proteins; cannot be done with larger RNAs
- C. Adenosine triphosphate (ATP) is an RNA nucleotide - source of energy for biochemical reactions (most energy used at any given moment in any living organism is derived from it)
  1. Guanosine triphosphate (GTP) is also of enormous importance in cell activities
  2. GTP binds to a variety of proteins (G proteins) & acts as switch to turn on their activities

VII. An RNA World - as proteins & DNA have garnered most of headlines, RNA relegated primarily to a role as messenger, an intermediate in the flow of genetic information from DNA to protein
- A. Recent discovery that RNAs can catalyze essential chemical reactions (ribozymes) has changed that
- B. Has led to speculation that early in evolution, there were no proteins & no DNA
- C. Probably RNA performed double duty, serving as genetic material & catalyzing necessary reactions – later in evolution, these jobs turned over to DNA & proteins

The Formation of Complex Molecular Structures

I. Proteins fold into 3D shape by self-assembly or with help of a few nonspecific "chaperones" in cell

A. Amino acid sequence is primary determinant of ultimate 3D shape of protein (largely self-assemble into most stable arrangement); so do larger structures like cell organelles
B. Evidence that assembly process is self-directed is demonstration that it can occur outside cell (*in vitro*) under physiologic conditions when only macromolecules involved in structure are present

II. Assembly of Tobacco Mosaic Virus (TMV) particles – self-assembly can occur *in vitro* under physiologic conditions when only the macromolecules involved in assembly are present
A. Heinz Fraenkel-Conrat & Robley Williams, California-Berkeley (1955) – demonstrated that TMV particles are capable of self-assembly
1. Consist of one long RNA molecule (~6600 nucleotides) wound within helical capsule made of 2130 identical protein subunits
2. Purified TMV RNA & protein separately & mixed them -> got mature, infective particles
B. The proteins & RNA of TMV contain all the information necessary for particle formation

III. Ribosomes also made of RNA & protein; several different types of RNA & lots of different proteins & two subunits of different size & highly irregular in shape
A. Large (or 50S) subunit of bacteria contains 2 RNA molecules & ~32 different proteins
B. Small (or 30S) ribosomal contains 1 RNA molecule & 21 different proteins

IV. Masayasu Nomura et al., Wisconsin (1966) – reconstituted complete, fully functional 30S subunits; mixed 21 purified proteins of small subunit & purified small subunit rRNA from *E. coli*
A. Subunit self-assembly occurs in sequential step-by-step manner paralleling *in vivo* process
1. Incorporation of individual proteins alters growing particle, making it more receptive to attachment of new proteins
2. At least one small subunit protein (S16) works only in ribosome assembly, its deletion from reconstitution mix greatly slowed assembly but did not block functional ribosome formation
3. Many other small subunit proteins function primarily to stabilize assembled structure
4. Ribosomal reconstitution process takes up to 2 hrs at 50°C *in vitro* but a few minutes in bacterium as low as 10°C
5. Accessory factors like chaperones that help in protein folding may be involved that investigators did not have
B. Large ribosomal subunits of *E. coli* also reconstituted in next decade

V. Eukaryote ribosomes need transient associations with proteins that do not end up in final particle
A. Also about half of nucleotides are removed from large ribosomal RNA precursor
B. Information for self-assembly *in vitro* is actually gone after excision of these nucleotides

THE HUMAN PERSPECTIVE: FREE RADICALS AS A CAUSE OF AGING

I. Why do humans have a maximum life span of ~100 years while chimpanzees only live about half this length of time?
A. Many biologists believe aging results from gradual accumulation of damage to bodily tissues
1. The most destructive damage probably occurs to DNA
2. DNA alterations lead to faulty genetic messages that promote gradual cellular deterioration
B. How does cellular damage occur and why is it more rapid in chimpanzees?

II. Atoms are stabilized when their shells are filled with electrons
A. Electron shells consist of orbitals that hold a maximum of 2 electrons

1. Atoms or molecules that have orbitals containing a single unpaired electron tend to be highly unstable (**free radicals)**
2. Free radicals may be formed when a covalent bond is broken such that each portion keeps one-half of the shared electrons **or**
3. They may be formed when an atom or molecule accepts a single electron transferred during an oxidation-reduction reaction (ex.: H_2O —> $HO^{\cdot}$ [hydroxyl radical] + $H^{\cdot}$)

B. Free radicals are extremely reactive & can chemically alter many types of molecules (proteins, nucleic acids, lipids; formation of OH radicals may be major reason for skin damage by sun

III. Denham Harman (Univ. of Nebraska, 1956) – proposed aging results from tissue damage caused by free radicals; no significant interest because biologists/doctors not familiar with free radicals

IV. Joe McCord & Irwin Fridovich (Duke Univ., 1969) – discovered enzyme superoxide dismutase (SOD) whose sole function was the destruction of the superoxide radical ($O_2^{\cdot -}$)

A. Superoxide free radical is formed when molecular oxygen picks up an extra electron
1. SOD catalyzes the following reaction: $O_2^{\cdot -} + O_2^{\cdot -} + 2H^+$ —> H_2O_2 (hydrogen peroxide) + O_2
2. H_2O_2 is potentially reactive oxidizing agent (often used as disinfectant & bleaching agent); normally destroyed in cell by the enzymes catalase or glutathione peroxidase
3. If it is not destroyed rapidly, hydrogen peroxide can break down to hydroxyl radicals, which can attack cellular macromolecules

B. Subsequent research has revealed that superoxide radicals are formed in cells during normal oxidative metabolism & SOD is present in cells of diverse organisms, from bacteria to humans
1. Animals possess 3 different versions (isoforms) of SOD: a cytosolic, mitochondrial & extracellular isoform
2. It is estimated that 1 – 3% of oxygen taken into human mitochondria is converted to H_2O_2 rather than water, the normal end product of respiration
3. Mutant bacteria & yeast lacking SOD are unable to grow in presence of oxygen
4. Mice lacking mitochondrial SOD (SOD2) cannot survive more than about a week after birth
5. 2005 - mice genetically engineered so mitochondria have elevated levels of H_2O_2-destroying enzyme catalase live 20% longer than untreated controls
   a. First demonstration that enhanced antioxidant defenses can increase life span of mammal

V. SOD importance as factor in aging controversial – Harman's free radical & aging hypothesis makes certain predictions

A. Animals with longer life spans might make fewer free radicals, have a better capacity for destroying them or be better able to repair cellular damage resulting from their reactions

B. Evidence – growth of mouse & human fibroblasts in culture under conditions of standard (20%) & reduced (3%) oxygen levels
1. Mouse fibroblasts (connective tissue cells) grown under reduced O_2 conditions suffered ~1/3 as much DNA damage
2. They also underwent many more cell divisions before they stopped dividing than the same cells grown at normal O_2 levels
3. Mouse fibroblasts cultured in 20% O_2 suffered 3 times as much oxidative DNA damage as human fibroblasts cultured under the same conditions
4. Human cells seem to be much better than mouse cells in preventing and/or repairing oxidative DNA damage

C. Evidence - animal life spans can be increased by restricting calories present in diet
1. 1930s – mice maintained on very strict diets typically live 30 – 40% longer than their littermates fed diets of normal caloric content

2. Studies of their metabolic rates have produced contradictory data but…….
3. There is general agreement that animals fed calorie-restricted diets exhibit a marked decrease in $O_2^{\cdot -}$ & H_2O_2 production, which could explain their increased longevity
4. Long term studies currently in progress on rhesus monkeys to see if they can live longer & healthier lives when maintained on calorie-restricted diets
   a. Not conducted long enough yet to determine if their maximum life span (normally ~40 years) is increased, but…..
   b. They have lower blood levels of glucose, insulin & triglycerides, which makes them less prone to age-related disorders like diabetes & coronary artery disease
5. Lowered blood-insulin levels may be particularly important in promoting longevity
   a. Nematode & fruit fly studies suggest that reducing insulin-like hormone activity in these invertebrates can dramatically increase their life spans

VI. Studies are being done on antioxidants, substances that are able to destroy free radicals
   A. Common antioxidants in body are glutathione, vitamins E & C, beta-carotene (orange pigment in carrots & other vegetables) – they destroy free radicals & may prove highly beneficial in diet
   B. Studies on rats & mice have not provided convincing evidence that they retard aging or increase maximum life span
   C. An antioxidant receiving considerable interest is resveratrol (a polyphenolic compound found at high concentration in the skin of red grapes)
      1. It is widely believed that resveratrol is responsible for the health-related benefits attributed to red wine
      2. Rather than scavenging for free radicals, resveratrol appears to act by stimulating an enzyme (Sir2) that serves as a key player in promoting longevity

THE HUMAN PERSPECTIVE: PROTEIN FOLDING CAN HAVE DEADLY CONSEQUENCES

I. Paper published in *Lancet* (April 1996) caused alarm in Europe; described study of 10 persons who had Creutzfeld-Jakob disease (CJD)
   A. It is a rare, fatal disorder that attacks brain, causing loss of motor coordination & dementia
      1. Can occur as inherited disease that runs in certain families or as sporadic form that appears in individuals who have no family history of the disease
   B. Unlike almost every other heritable disease, CJD can also be acquired
      1. Until recently, persons who had acquired CJD had been recipients of organs or organ products donated by a person with undiagnosed CJD
      2. *Lancet* article reported on CJD acquired from contaminated beef that infected individuals had eaten years before instead from another person
      3. Contaminated beef derived from cattle raised in England; they contracted a neurodegenerative disease that caused animals to lose motor coordination & develop demented behavior
      4. Disease has come to be known as "mad cow disease"
   C. Patients who got CJD from contaminated beef distinguished by several criteria from those who suffer from the classical forms of the disease
   D. To date, ~150 people have died of CJD acquired from contaminated beef; number of deaths declining; on surface, suggests epidemic over, but public health officials have reasons for concern
      1. Studies of tissues removed during surgeries in England indicate 1000s of people are likely to be infected with disease without exhibiting symptoms
      2. Even if these people never develop clinical disease, they remain potential carriers who could pass CJD to others through blood transfusions

3. At least 2 people, are believed to have contracted CJD after receiving blood from a donor harboring the disease
4. Underscores need to test blood for presence of responsible agent

II. How can same disease be both inherited (caused by faulty gene) & infectious (traced to infectious agent? – answer has emerged gradually over past several decades

A. Started with observations by D. Carleton Gajdusek in 1960s on strange malady that once afflicted native population of Papua, New Guinea
   1. Showed that islanders were contracting a fatal neurodegenerative disease (kuru) during funeral ritual in which they ate brain tissue of recently deceased relative
   2. Brain autopsies of kuru patients who died had distinct pathology (spongiform encephalopathy) in which certain brain regions were riddled with microscopic holes (vacuolations)
   3. Tissue resembled a sponge

B. It was soon shown that brains of kuru-suffering islanders were strikingly similar in microscopic appearance to the brains of those suffering from CJD
   1. Raised question – does brain of CJD sufferer with inherited form contain an infectious agent?
   2. Gajdusek (1968) – injected extracts prepared from brain biopsy of dead CJD victim into suitable lab animal —> animal developed spongiform encephalopathy similar to kuru or CJD
   3. Concluded that extracts contained an infectious agent presumed at the time to be a virus

C. Stanley Prusiner (Univ. of Cal., SF, 1982) – reported that, unlike viruses, CJD-causing agent lacked nucleic acid & instead was composed solely of protein; called the protein a **prion** (protein only)
   1. Protein only hypothesis initially met with skepticism, but subsequent studies (Prusiner & others) provided overwhelming support
   2. Prion protein was first thought to be an external agent, a virus-like particle lacking nucleic acid

D. However, prion protein was soon shown to be encoded by gene (*PRNP*) in cell's own chromosomes
   1. Gene is expressed in normal brain tissue & encodes protein designated PrP^{C} (prion protein cellular) that resides at surface of nerve cells; its precise function is a mystery
   2. Modified version of protein called PrP^{Sc} (prion protein scrapie) is present in brains of humans with CJD; unlike normal PrP^{C}, it accumulates within nerve cells (forms aggregates that kill cells)

E. In purified states, PrP^{C} & PrP^{Sc} have very different physical properties
   1. PrP^{C} remains as monomeric molecule that is soluble in salt solutions & readily destroyed by protein-digesting enzymes
   2. In contrast, PrP^{Sc} molecules interact with one another to form insoluble fibrils that are resistant to enzymatic digestion
   3. Based on differences, one might expect distinct variation in their amino acid sequences, but this is not so
   4. Instead, they can have identical amino acid sequences but differ in way they fold to form 3D protein molecule
   5. PrP^{C} molecule consists almost entirely of α-helical segments & interconnecting coils; the core of a PrP^{Sc} molecule consists of β sheet
   6. PrP can be converted from soluble, protease-sensitive conformation into insoluble, protease-insensitive aggregates in vitro by simply changing the conditions in test tube

III. How can mutant polypeptide that might be less stable & more likely to fold into abnormal PrP^{Sc} conformation act as an infectious agent?

A. Presently thought that abnormal prion molecule PrP^{Sc} can bind to normal PrP^{C} protein & cause the normal protein to fold into the abnormal form
   1. This conversion can be shown to occur in test tube: addition of PrP^{Sc} to preparation of PrP^{C} can convert PrP^{C} molecules into the PrP^{Sc} conformation

2. It is thought that appearance of abnormal protein in body (whether due to rare misfolding event in case of sporadic disease or by exposure to contaminated beef) starts chain reaction
3. Chain reaction involves gradual conversion of normal cell proteins to abnormal prion form

B. Precise mechanism by which prions lead to neurodegeneration remains unclear

IV. CJD is rare disease caused by protein with unique infective properties, but Alzheimer's disease (AD), on other hand, is common disorder that strikes up to 10% of people ≥65 & 40% who are ≥80

A. Persons with AD exhibit memory loss, confusion & loss of reasoning ability

B. CJD & AD share a number of important features

1. Both are fatal neurodegenerative diseases that can occur in either inherited or sporadic form
2. Like CJD, AD sufferer brain contains fibrillar deposits of insoluble material called amyloid
3. In both CJD & AD, toxic fibrillar deposits result from self-association of polypeptide composed predominantly of β sheet

C. There are also many basic differences between the 2 diseases

1. The proteins that form the disease-causing aggregates are totally unrelated
2. The parts of the brain that are affected are distinct
3. The protein responsible for AD does not act like an infectious agent (it is nontransmissable)

D. Most evidence suggests that AD is caused by production of molecule called **amyloid β-peptide** (*Aβ*), which is originally part of a larger protein called **amyloid precursor protein** (*APP*)

1. APP spans the nerve cell membrane
2. Aβ peptide is released from APP molecule after cleavage by 2 specific enzymes, β-secretase & γ-secretase
3. Length of Aβ peptide is somewhat variable; predominant species has length of 40 amino acids (Aβ40), but a minor species with 2 additional hydrophobic residues (Aβ42) is also produced
4. Both peptides can exist in soluble form consisting predominantly of α helices, but Aβ42 has tendency to spontaneously refold into different conformation with considerable β-pleated sheet
5. Aβ42 version of protein has greatest potential to cause AD; it tends to self-associate to form small complexes (oligomers) as well as large aggregates that are visible as fibrils in EM
6. Not settled, but recent evidence suggests that soluble oligomers are most toxic to nerve cells, rather than insoluble aggregates
7. Oligomers appear to attack synapses that connect one nerve cell to another & eventually the death of the nerve cells

E. People who suffer from inherited form of AD carry mutation that leads to increased production of Aβ42 peptide

1. Overproduction of Aβ42 can be caused by mutations in *APP* gene or in genes (*PS1, Ps2*) that encode subunits of γ-secretase
2. Individuals with such mutations exhibit symptoms of disease at early age, typically in 50s

V. Develop animal model for AD that mimics human disease & use animals to test effectiveness of potential therapies; up until 1995, there was no AD animal model

A. Brains of aging mice show no evidence of amyloid deposits found in humans

1. In 1995, researchers found they could create mouse strain that developed amyloid plaques in the brain & performed poorly at tasks that require memory
2. Created strain by genetically engineering mice to carry a mutant human *APP* gene, one responsible for causing AD in families
3. These genetically engineered (transgenic) mice have been invaluable for testing potential AD therapies

B. Dale Schenk, et al. (Elan Pharmaceuticals, 1999) – found that formation of amyloid plaques in mice carrying mutant human *APP* gene could be blocked

1. Could be done by repeatedly injecting animals with molecule that causes problem, the aggregated Aβ42 protein
2. In effect, researchers had immunized (vaccinated) the mice against the disease
3. When young (6-week-old) mice were immunized with Aβ42, they failed to develop the amyloid brain deposits as they grew older
4. When older (13-month-old) mice whose brains already contained extensive amyloid deposits were immunized with Aβ42, significant fraction of the fibrillar deposits was cleared out of NS
5. Even more important, the immunized mice performed better than their non-immunized littermates on memory-based tests

C. These results combined with the fact that the animals suffered no ill effects led government regulators to approve Phase I clinical trials of the Aβ42 vaccine
1. This is the first step in testing a new drug or procedure in humans & usually comes after years of preclinical testing on cultured cells & animal models
2. Phase I trials are done on a small number of subjects & are designed to monitor the safety of the procedure rather than its effectiveness against the disease
3. None of subjects in 2 separate Phase I trials of Aβ vaccine showed any ill-effects from injection of amyloid peptide

D. Success allowed them to go to Phase II clinical trials with larger group of subjects; designed to get measure of procedure/drug effectiveness (a randomized, double-blind, placebo-controlled study)
1. Patients randomly divided into 2 groups treated similarly except that one group given curative factor being tested & other group given placebo (inactive substance with no therapeutic value)
2. Double-blinded means that neither researchers nor patients know who is receiving placebo

E. Phase II trial enrolled 350 people (US, Europe) diagnosed with mild to moderate AD
1. After 2 injections of synthetic β-amyloid (or placebo), 6% of subjects experienced potentially life-threatening brain inflammation; most were successfully treated with steroids
2. Trial was discontinued due to serious side effect, but those receiving vaccine were still monitored & there is reason for optimism
3. Autopsies on several patients from trials who have since died indicate that amyloid plaques present in certain brain regions had largely disappeared
4. Suggests that antibodies made in response to immunization had entered patients' brains & induced desired effect, just as in original mouse studies
5. More importantly, tests on >30 patients from study indicate that vaccine may have dramatically slowed disease progression

F. Present goal is to develop safer immunization strategies – safest approach currently being investigated is injection of antibodies directed against Aβ that have been produced outside body
1. Known as passive immunization because person does not produce the therapeutic antibodies themselves; already proven capable of restoring memory function in transgenic mice
2. If strategy is effective in Phase I & II clinical trials, it will enter Phase III trial, which is usually last step before government approval
3. Phase III typically employs large numbers of subjects (1000 or more at several research centers) & compares effectiveness of new treatment with standard approaches

VI. 2 therapeutic strategies proposed for AD prevention; emerged as result of epidemiologic studies, which try to correlate certain disease outcome with particular activity in members of large population

A. Individuals who have taken certain anti-inflammatory medications like ibuprofen or cholesterol-lowering medications (statins) have a markedly reduced incidence of AD

B. Both types of drugs are already approved for human use & have been taken regularly by tens of millions of people, which makes them ideal therapeutic agents

C. Large-scale, controlled trials to test their ability to prevent AD are under way

VII. Epidemiological studies also suggest that AD onset is delayed in people who continue to participate in physical and/or mentally challenging activities

VIII. Other approaches are being considered for AD treatment; given the variety of potential therapies being tested, there is some reason for optimism that AD may soon be treatable
- A. Implantation into the brain of cells that secrete NGF, a protein that has been shown to prevent neurodegeneration in animal models
- B. Compounds that inhibit the enzymatic activity of either β- or γ-secretase, which should decrease the production of Aβ42
- C. Compounds that bind to the soluble Aβ peptide & prevent it from aggregating or forming fibrils
  1. One such compound (Alzhemed) has shown much promise in slowing or preventing cognitive decline in patients with mild AD
- D. Compounds that inhibit ACAT, an enzyme involved in cholesterol metabolism, whose activity may lead to increased Aβ production
- E. Compounds, such as Clioquinol, that remove (chelate) zinc & copper ions
  1. Both of these ions are normally present within amyloid plaques & are thought to promote plaque formation

EXPERIMENTAL PATHWAYS: CHAPERONES – HELPING PROTEINS REACH THEIR PROPER FOLDED STATE

I. F. M. Ritossa (1962) – Italian biologist studying development of fruit fly *Drosophila* reported a curious finding
- A. When temperature at which fruit fly larvae were developing was raised from normal 25°C to 32°C, a number of new sites on giant chromosomes of larval cells became activated
  1. The giant chromosomes of these insect larvae provide a visual exhibit of gene expression
  2. Results suggest that elevated temperature induced new gene expression; confirmed a decade later with characterization of several proteins appearing in larvae after temperature elevation
- B. This response, the **heat shock response** was not confined to fruit flies, but can be initiated in many different cells from virtually every type of organism (bacteria to plants & mammals)
  1. The heat-shock proteins (hsps) produced during response were found not only in heat-shocked cells, but also at lower concentration in cells under normal conditions

II. Multisubunit structures (bacterial ribosome, tobacco mosaic virus particle) can self-assemble from purified subunits
- A. 1960s – demonstrated that proteins that make up bacteriophage particles also possess remarkable ability to self-assemble, but can't form complete, functional virus particle by themselves *in vitro*
  1. Phage assembly experiments in bacteria cells showed that phages require bacterial help
- B. 1973 – found that certain mutant bacteria strain (*GroE*) did not support assembly of normal phages
  1. Depending on the type of phage, phage particle head or tail was assembled incorrectly
  2. Suggested that a protein encoded by the bacterial chromosome participated in viral assembly, even though this host protein was not a component of the final virus particles
  3. Since it did not evolve as an aid for viral assembly, bacterial protein needed for phage assembly had to play some role in bacterial cell's normal activities, but precise role was obscure
- C. Later studies – *GroE* site on bacterial chromosome actually contains 2 separate genes, *GroEL* & *GroES* that encode 2 separate proteins, GroEL & GroES
  1. Under EM, purified GroEL protein appeared as cylindrical assembly consisting of 2 disks

2. Each disk was composed of 7 subunits arranged symmetrically around central axis

D. Several years later – study on pea plants hinted at existence of similar assembly-promoting protein in plant chloroplasts
   1. Rubisco is large chloroplast protein that catalyzes reaction in which CO_2 molecules taken up from atmosphere are covalently linked to organic molecules during photosynthesis
   2. Rubisco comprises 16 subunits: 8 small subunits (molecular mass of 14,000 daltons) & 8 large subunits (55,000 daltons)
   3. Large Rubisco subunits, synthesized in chloroplast, are not present in independent state, but associated with huge protein assembly consisting of identical 60,000 dalton (60 kDa) subunits
   4. Researchers considered possibility that complex formed by large Rubisco subunit & the 60-kDa-polypeptide was an intermediate in the assembly of a complete Rubisco molecule

E. Separate study on mammalian cells – revealed existence of proteins that appeared to assist assembly of multisubunit proteins
   1. Like Rubisco, antibody molecules consist of a complex of 2 different types of subunits, smaller light chains & larger heavy chains
   2. Like large Rubisco subunits, heavy chains of antibody complex become associated with another protein not found in the final complex
   3. This protein, which associates with newly synthesized heavy chains, but not with heavy chains that are already bound to light chains, was named binding protein (BiP)
   4. BiP was subsequently found to have a molecular mass of 70,000 daltons (70 kDa)

F. 1986 – 2 lines of investigation (heat-shock response & proteins that promote protein assembly) came together
   1. Shown that very prominent heat shock response protein, heat-shock protein 70 (hsp70) due to its molecular mass, was identical to BiP, the protein implicated in antibody molecule assembly

III. Even before heat-shock response was known, protein structure was known to be sensitive to temperature

A. A small rise in temperature could cause delicate proteins to begin to unfold
   1. Unfolding exposes hydrophobic residues previously buried in protein core
   2. Hydrophobic residue patches on protein surfaces attract each other like fat droplets
   3. Thus, when cell is heat shocked, soluble proteins are denatured & form aggregates

B. 1985 report showed that after temperature elevation, newly synthesized hsp70 molecules enter cell nucleus & bind to nuclear protein aggregates
   1. They then act like molecular crowbars to promote disaggregation
   2. Due to their role in assisting protein assembly by preventing undesirable interactions, hsp70 & related proteins were named **molecular chaperones**

IV. Soon demonstrated that bacterial heat-shock protein GroEL & Rubisco assembly proteins in plants are homologous proteins

A. The 2 proteins share same amino acids at nearly half of >500 residues in their respective molecules
   1. The fact that the 2 proteins (both members of Hsp chaperone family) have retained so many of the same amino acids reflects their similar & essential functions in the 2 cell types

B. At this point, it was thought that their primary function was to mediate the assembly of multisubunit complexes like bacteriophage particles & Rubisco

C. This view was changed by experiments studying molecular chaperones in mitochondria
   1. Newly made mitochondrial proteins made in cytosol have to cross outer mitochondrial membranes in unfolded, extended, monomeric form
   2. Mutant was found that altered the activity of another Hsp chaperone family member that resided in mitochondria

3. In cells containing mutant chaperone, proteins that were transported into mitochondria failed to fold into their active forms
4. Even proteins consisting of single polypeptide chain failed to fold into native conformation
5. This changed the perception of chaperone function from a notion that they assist assembly of already-folded subunits into larger complexes to idea that they assist polypeptide chain folding

V. Such results indicated presence in cells of at least 2 major molecular chaperone families: the Hsp70 chaperones (BiP) & the Hsp60 chaperones (chaperonins; Hsp60, GroEL, Rubisco assembly protein)
   A. 1979 – GroEL shown to be huge molecular complex of 14 polypeptide subunits arranged in 2 stacked rings resembling a double doughnut
   B. 15 years later – 3D structure of GroEL complex was determined by X-ray crystallography; it revealed presence of central cavity within GroEL cylinder
   C. Later studies showed the cavity was divided into 2 separate chambers, one at each end of complex
      1. Each chamber was situated within the center of one of the GroEL complex rings & was large enough to enclose a polypeptide undergoing folding
   D. EM studies also provided information about the function of a second protein, GroES, which acts in conjunction with GroEL
      1. GroES, like GroEL, is a ringlike protein with 7 subunits arrayed symmetrically around a central axis
      2. GroES, however, consists of only one ring & its subunits are much smaller (10.000 daltons) than those of GroEL (60,000 daltons)
      3. GroES is seen as a cap or dome that fits on top of either end of a GroEL cylinder
      4. Attachment of GroES to one end of GroEL causes dramatic conformational change in GroEL protein that markedly increases the volume of the enclosed chamber at that end of complex
   E. Importance of conformational change revealed in X-ray crystallographic studies by Arthur Horwich & Paul Sigler (Yale)
      1. Binding of GroES cap is accompanied by 60° rotation of the apical domain of the subunits that make up the GroEL ring at that end of the GroEL cylinder
      2. GroES attachment does more than trigger a conformational change that enlarges the GroEL chamber
      3. Before GroES attachment, the wall of the GroEL chamber has exposed hydrophobic residues that give the lining a hydrophobic character
      4. Nonnative polypeptides also have exposed hydrophobic residues that become buried in the interior of the native polypeptide
      5. Since hydrophobic surfaces tend to interact, the hydrophobic lining of the GroEl cavity binds to the surface of nonnative polypeptides
      6. Binding of GroES to GroEL buries the hydrophobic residues of the GroEL wall & exposes a number of polar residues, thereby changing the chamber wall character
      7. Thus, a nonnative polypeptide that had been bound to the GroEL wall by hydrophobic interactions is displaced into the space within the chamber
      8. If polypeptide has not reached its native conformation by the time it is ejected, it can rebind to the same or another GroEL, & the process is repeated
   F. Proteins usually seen as highly specialized molecules with large impact on very narrow range of events; most enzymes accelerate reaction rates greatly, but only for one or a few related reactions
      1. GroEL, in contrast, can promote the folding of a broad spectrum of unrelated proteins; described as a "folding machine for many but a master of none"
      2. Structure of GroEL-GroES is a compromise that allows chaperonin to help a large variety of proteins to some degree without helping any proteins to a great degree
   G. In one study, site-directed mutagenesis was used to modify a key residue, Tyr71 of GroES, whose side chain hangs from ceiling of folding chamber

1. Due to its aromatic ring, tyrosine is a modestly hydrophobic residue
2. If Tyr71 replaced by "+" or "-" charged amino acid, resulting GroEL-GroES variant had higher ability to help the folding of a specific foreign protein, green fluorescent protein (GFP)
3. But substitutions for Tyr71 that improved GroES-GroEL ability to increase GFP folding made the chaperonin less competent to help its natural substrates fold
4. Thus, as chaperonin became more & more specialized to interact with GFP, it lost its general ability to assist folding of proteins having an unrelated structure
5. Suggests that individual amino acids in the folding chamber wall may participate somehow in the folding reaction
6. Thus, chaperonins may do more than simply provide a passive chamber in which proteins can fold without outside interference

H. Molecular chaperones do not convey information for folding process but instead prevent proteins from veering off correct folding pathway & finding themselves in misfolded or aggregated states
   1. The 3D structure of a protein is determined by its amino acid sequence

LECTURE HINTS

Basic Biological Chemistry

Some students of cell biology are distressed to discover that an understanding of cell biology requires a command of (at least) some basic chemical principles. This distress most often stems from a fear (oftentimes a long-standing one) of chemistry that can make these students refractory to the information you are trying to convey. Thus, it is important to present these concepts in a less threatening manner so that the students will have a better chance of understanding the principles. If the students begin to grasp the concepts, they will, in most cases, gain confidence and eventually lose at least some of the fear that is hampering their learning.

Chemical Bonding

The first of the principles that students of cell biology must master is that of chemical bonding. It has been useful to explain ionic bonds and covalent bonds in the following way. Chemical bonds arise as a way of filling the outer electron shells of atoms participating in the bond. Full outer shells confer enhanced stability on the atoms. Ionic bonds occur as a result of valence *electron transfer* from one atom to another, e. g., from Na to Cl (as in table salt), creating Na^+ and Cl^- ions that attract one another. On the other hand, covalent bonds involve a *sharing of valence electrons*. There are two types of covalent bonds: polar and nonpolar covalent bonds. Polar covalent bonds result from an unequal sharing of electrons that causes the more electronegative atom, like the oxygen atom in water, to acquire a partial negative charge since the electrons spend more time in its vicinity. At the same time, the other atom acquires a partial positive charge since this same electron is not spending as much time in its vicinity (the H atoms of water). Nonpolar covalent bonds involve an equal sharing of electrons as in molecular hydrogen and CH_4 (methane). Consequently, there is no separation of charge across this bond and the bond is nonpolar. This situation can be illustrated as follows:

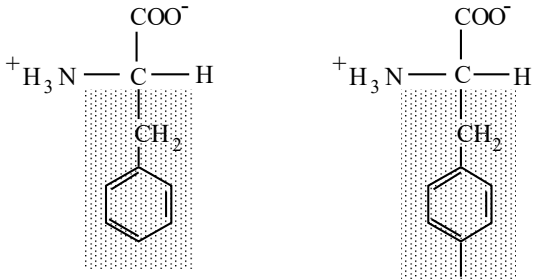
complete e^- transfer	unequal sharing of e^-	equal sharing of e^-
ionic bond	**polar covalent bond**	**nonpolar covalent bond**
full charge	partial charge	no charge

The formation of these bonds can be considered to lie along a continuum from complete electron transfer (ionic bonds) to an equal sharing of electrons (nonpolar covalent bonds). One might well think of

unequal sharing of electrons as an incomplete transfer of electrons and equal sharing of electrons as essentially no transfer of electrons. It depends on your viewpoint.

Hydrophobic vs. Hydrophilic

Perhaps the most difficult concept for students to grasp has been the difference between hydrophilic and hydrophobic molecules and how these molecules or portions of molecules may be recognized. I have attempted more complex chemical arguments that have met with limited success. However, for some students, simplifying the definition does relatively little to improve matters. You could tell students to consider straight carbon chains or ring structures without attached hydrophilic groups (hydroxyls, sulfhydryl groups, amine groups, carboxyl groups, etc.) as earmarks of hydrophobic molecules or sections of molecules. Using the amino acids phenylalanine and tyrosine as examples, emphasize the effect of one hydroxyl group on a largely hydrophobic structure. Phenylalanine and tyrosine are considered to be, respectively, hydrophobic and polar, uncharged amino acids by virtue of their R group chemistry (the shaded area in the drawing below).



Phenylalanine
Nonpolar Amino Acid

Tyrosine
Polar, Uncharged
Amino Acid

The R groups of both amino acids contain the same typically hydrophobic carbon-containing ring. However, tyrosine contains, in addition, a single hydrophilic hydroxyl group that alters its properties enough to justify its classification as a polar, uncharged amino acid. By applying some simple rules, a student can determine whether a molecule has hydrophobic and/or hydrophilic properties. The presence of a carbon-containing ring or chain of carbons without characteristically hydrophilic groups indicates a hydrophobic molecule. The most common hydrophilic groups in living cells are amino, carboxyl, hydroxyl, and sulfhydryl groups. A molecule possessing both hydrophilic and hydrophobic substituents is said to be amphipathic (*amph* - meaning on both sides).

Demonstration

SOLUBILITY TOYS

I demonstrate the differential solubility of hydrophobic and hydrophilic materials by using souvenirs that invariably show up in gift shops at the beach. The simplest of these toys consists of a hollow plastic cube containing two liquids with a plastic surfer that manages to ride the waves at the interface between the two fluids. The lower fluid is blue due to the presence of a blue dye that is undoubtedly a hydrophobic molecule soluble in other hydrophobic materials, but insoluble in water. Consequently, no matter how hard one shakes the cube, the blue dye will never leech out into the water that makes up the upper liquid in the cube. You can shake the cube as vigorously as you wish, but the oil and the dye along with it will separate from the water above after a short time. There are more elaborate toys like this on the market, some of which appear to contain both hydrophilic and hydrophobic dyes. For example, the Colorfall by the

Carlisle Company of Carson City, Nevada (702-246-7822; PO Box 21029, Carson City NV 89721) contains water colored with a hydrophilic yellow dye. Two chambers within the device contain mineral oil and the dyed water. The mineral oil in the two chambers contains dyes of different colors. The device clearly demonstrates the tendency of oil and water to remain separate and the solubility of hydrophobic substances in other hydrophobic substances. The dyes do not leech into materials of dissimilar chemistry. Students can also observe the fusion of drops of oil as they encounter one another in both toys, although the Colorfall is better suited to this purpose. This can aid the students in understanding how membranes can fuse during exocytosis. Either type of toy serves as an effective demonstration of the differential solubilities of hydrophobic and hydrophilic substances.

Acids, Bases, Buffers and pH

Another aspect of chemistry that is likely to cause problems is the comprehension of acids and bases. An acid is traditionally described as a molecule that can donate protons (H^+ ions), while a base is described as a molecule that can accept protons. These definitions, while certainly correct, can be troublesome for students when they are presented with circumstances in which a base can donate electrons and an acid can accept them. For instance, at pHs below 9 - 10, amino groups, the major biological bases, act like classic bases as described above. However, at pH values above 10, an amino group actually tends to lose its extra proton; that is, under these extreme conditions of pH, a base behaves like an acid. Conversely, the carboxyl group, the major biological acid, acts like an acid at pH values in excess of 2 - 3. At more acidic pHs, this chemical group tends to acquire and retain the proton it usually loses; under this circumstance, this normally acidic group acts like a base. It is perhaps best to remind students that the definition for acids and bases was developed within a frame of reference at pH 7 where an acid behaves like an acid and a base like a base. It is wise to stress that an acid loses its protons at pH's below 7, while bases lose their protons at pH's above 7 as illustrated below:

In other words, both acids and bases can gain and lose protons. The difference between them is the pH at which this happens. The ionization of amino and carboxyl groups under different pH conditions is of paramount importance to protein structure since charges on such groups, when altered, can affect the three dimensional shape of proteins.

In my experience, students do not have much difficulty comprehending buffers. I do, however, try whenever possible to get students to apply what they are learning to common everyday experiences. If I do it enough, I hope that they may get the idea and make the same sorts of connections on their own. A prime example of this refers to a well-known pain reliever. Most students are aware that because of its acidity (and its ability to inhibit COX-1 as described in this chapter in Karp and above in the outline), aspirin can exacerbate stomach and intestinal troubles in certain individuals. I ask the students to think back and come up with a product that is advertised as a remedy for this particular problem. One of them usually comes up with Bufferin; I then ask them if they know the origin of the name. Eventually, students (perhaps with some prodding) realize that buffer is added to the aspirin to neutralize its acidity and prevent acid build-up in the stomach. This also serves as an example of another practice I employ as much as possible. I urge my students to think about names. Why is Bufferin called Bufferin? What do the prefixes and roots in "exothermic" and "endergonic" mean? Instead of memorizing terms and their definitions, students can learn to dissect them and use their roots and prefixes as devices for

remembering their meanings. To that end, wherever possible I point out these roots and prefixes and their meanings.

Macromolecules

Names and Nomenclature: Memorization vs. Understanding

In the discussion of macromolecules, a number of problems crop up that seem to obstruct student understanding. A conflict exists between the tendency of students to memorize facts about macromolecules and the use of previously acquired knowledge to extend their understanding about these molecules. A trivial example of this phenomenon is the difficulty students often seem to have remembering the names of the various bonds involved in holding together macromolecules. For example, many students cannot remember that the bonds that connect monosaccharide monomers in polysaccharides are called glycosidic bonds. Instead of realizing the connection between glycosidic bonds and the root glyco- (*sweet*), many students simply memorize the term. Biology majors, who form the majority of students in our cell biology course, have been repeatedly exposed to the meaning of this root; if they realized this connection, memorization (at least in its classic sense) would seem to be virtually unnecessary. This does not mean that memorization is a tool that cannot be used. The ester linkage (named by German chemist L. Gmelin) that arises from a reaction between an acid and an alcohol with the elimination of water, does not have an association with a root that would facilitate connecting the compounds involved with the name. Here memorization makes sense. Once the meaning of "ester" has been memorized, however, the names of other bonds derived from it can be remembered by using roots. The phosphate ester (P—O—C) linkage by which a phosphate group is attached to a phospholipid resembles an ester (C—O—C) with one of the carbon atoms replaced by a phosphorus atom. In DNA and RNA, the bond that holds together successive nucleotides is the 3'- 5'-phosphodiester bond. The atoms involved in this linkage can be lined up as follows: 3'C—O—P—O—5'C. Clearly, these are two sequential phosphate ester bonds connected by the middle phosphorus atom (phospho**di**ester). The numbers simply identify the carbons of the ribose or deoxyribose sugar participating in the interaction.

When memorization in place of understanding occurs, students often forget the material by the time the next exam rolls around; we call this a "core dump". Some students are seemingly unaware that most courses build on information throughout the semester. If they assume that they can forget previously tested material, they often do not realize that they are severely decreasing their chances of understanding subsequent topics. Realizing this, I emphasize as much as possible the importance of making associations rather than pure memorization. I also frequently stress, in as humorous a manner as possible, that principles covered earlier in the semester will reappear time and time again during the semester and that "dumping" them after the test on the material has been administered is inadvisable.

Carbohydrates

Describe different types of polysaccharides and emphasize the differences in their structure and function (see Chapter outline and table above). For example, mention the presence of α (1-> 4) glycosidic linkages in glycogen, amylose and amylopectin and stress that the branch points in amylopectin and glycogen involve α (1-> 6) glycosidic linkages. Point out the ability of mammals to digest these molecules, while another polymer of glucose, cellulose, cannot be digested by mammals; tell them that the reason for this seeming oddity stems from the glucose monomers of cellulose being connected by β (1-> 4) glycosidic linkages instead of α (1-> 4) glycosidic linkages. The shapes of these two polysaccharides are significantly different especially in the region of these bonds. In glycogen and starch, each glucose monomer is oriented in the same direction. In contrast, the β– linkages in cellulose cause each glucose monomer to be upside-down relative to its two immediate neighbors. This

significantly alters the shape of the polysaccharide in the area of the linkages. The enzyme(s) that hydrolyzes the α (1-> 4) glycosidic linkages cannot recognize the analogous regions in cellulose and, hence, it cannot be digested in mammals like ourselves. You can then ask students to explain how mammals that don't possess the enzyme to digest cellulose can use cellulose as a food source (enteric bacteria digest the cellulose for them). Usually, they come up with the correct answer. I also have them consider the relevance of cellulose indigestibility in humans as it relates to the necessity for roughage or fiber in the human diet.

Lipids

It is important to emphasize that the lipids, while considered macromolecules, differ from the others (proteins, nucleic acids and carbohydrates) since they are not formed by polymerization. Emphasize the structures of the triglyceride and phospholipid building blocks: glycerol and fatty acids. Glycerol is viscous and seems greasy. Yet it is really hydrophilic. I ask my students the reason for that. Usually, they note the three hydroxyl groups and come up with the right answer. They are also often able to recognize the hydrophobic (the chain of carbons) and hydrophilic (the terminal carboxyl group) regions of fatty acids. Once the assembly of these building blocks into triglycerides has been described, ask why the product of this assembly is completely hydrophobic. This occurs, of course, because the hydrophilic portions of both molecules are obliterated during triglyceride formation.

This is also a good time to address the issue of saturation and unsaturation in the fatty acids of triglycerides. Once again, examples from the students' common experiences are valuable; they have all encountered animal fat and salad oil but have probably never been aware of their structures and the effect of those structures on their properties. Fats that are solid at room temperature contain predominantly fatty acids that are saturated (no double bonds). As a result, these substances can pack more closely together when present in higher quantities since their carbon chains are straight. On the other hand, lipids that contain largely unsaturated (with one or more double bonds) fatty acids are fluid and classed as oils since their carbon chains have kinks at the position of each double bond. This makes closer packing unlikely and explains the liquid nature of oils. Remind students that they are inundated by advertisements lauding the virtues of polyunsaturated fats like those from plants, the dangers of animal fats in their diets and the user-friendliness of soft margarine. It is instructive to ask students to compare butter that contains animal fats with soft margarine as both are removed from the refrigerator. Butter is, of course, hard and soft margarine much softer. This clearly illustrates the difference between saturated and polyunsaturated fats.

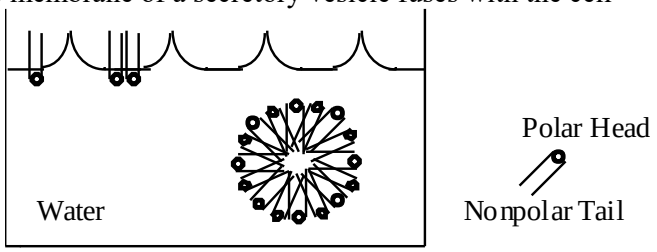
Contrast the hydrophobicity of triglycerides to the amphipathic nature of phospholipids. Since these lipids are built of a negatively charged phosphate group (to which other hydrophilic groups normally attach) along with the glycerol and fatty acids found in triglycerides, the resultant molecule possesses a polar head and nonpolar tail consisting of two fatty acids. Such a molecule has a split personality that leads to interesting structural and functional implications. Indeed, they form a significant portion of the cell membrane lipid bilayer. The behavior of these molecules in water serves as a good example of these implications (see Analogy box below).

ANALOGY

The Behavior of Phospholipids in Water

If phospholipids are spread on the surface of a water solution as they would be in a bowl of Grandma's chicken soup, they will arrange themselves so that their polar heads are in contact with the water while their hydrophobic tails project into the air above. If you push separate puddles of floating phospholipid

together with your spoon, these puddles will fuse into one in order to expose as little phospholipid to water as possible. This behavior is similar to what occurs when secretory vesicles approach the cell membrane during exocytosis. The membrane of a secretory vesicle fuses with the cell



membrane, opening the vesicle to the extracellular space and releasing its contents. If the soup is stirred vigorously, some of these lipids are driven beneath the surface of the water, exposing the hydrophobic tails to water. As phospholipids encounter each other in this underwater environment, they are able to enhance their collective stability by forming spheres with the hydrophilic heads facing and interacting with the surrounding water, while the hydrophobic tails interact with each other in the center of the structure away from the water. Such a structure is called a micelle. If there are many phospholipids around, these structures can get larger in which case they form a double-layered structure with two layers of phospholipid tails in the center and each surface exposed to water and covered with hydrophilic head groups, a structure called a liposome. Micelles are quite useful in the shower. Imagine that you have just spent the day mowing a few acres of grass without the luxury of a riding mower. You are as dirty as it gets and you decide to take a shower to get squeaky clean. After stepping into the stream and wetting yourself down, you soap up and create an appreciable lather. This lather contains micelles. Since dirt is often oily, the micelles will pick up the hydrophobic dirt molecules and since hydrophobic materials are soluble in other hydrophobic materials, the dirt will be attracted into the region of the micelle occupied by the phospholipid tails. When you step back into the shower spray, the water carries the micelles and the greasy dirt they contain off your body and down the drain since the outer surface of the micelles interacts well with the water.

Proteins

Structure and Chemistry of Amino Acids

Proteins are the most varied of macromolecules in structure and versatility of function. Emphasize the varied roles played by proteins in living cells and then proceed to a discussion of their structure. Macromolecular structure is often complicated for students to grasp, especially when emphasis is placed on the levels of protein structure. This is another area where students are tempted to memorize the material. Naturally, I start with the amino acids. Some instructors feel it is important to have their students memorize the structures of R groups. I don't feel that this is particularly advantageous, especially since the students will usually be asked to do this when they take a Biochemistry course. I tell my students that they will not be required to memorize the structure of the amino acid R groups. The resultant classwide sigh of relief is startlingly audible. Since R groups are responsible for the different chemical properties of individual amino acids, I do require the students to recognize the chemical nature of the R groups. If students are shown an amino acid, I expect them to know whether it is polar or nonpolar, polar, charged or polar, uncharged. Most students will be able to do this, given what they have already been taught about hydrophobicity and hydrophilicity.

Polymerization of Amino Acids

Polypeptides, polysaccharides and polynucleotides form through polymerization. This is a concept that can, at times, be difficult for students to grasp. They have trouble understanding how such seemingly

different molecules can be hooked together in an orderly fashion. By stressing that amino acids have both a constant portion (the α-carbon, the amino group and carboxyl group) and a variable portion (the R group), it is easy to make the point that the peptide bond is formed by interactions between the constant portions of each amino acid. This naturally allows the bonds to form by the same process no matter which two of the twenty amino acids are being connected. It is not unlike a freight train that has a wide variety of cars, all of which are fitted with identical couplers.

Emphasize that during polymerization, the monomers are connected by condensation reactions with the release of water as a byproduct. Also stress that the introduction of a water molecule across the bond leads to the breaking of the bond, a reaction called **hydrolysis** (hydro - *water*; lysis - *loosening*).

Levels of Structure in Proteins

The levels of structure in proteins are often a source of confusion for students. Perhaps the biggest problem is the tendency of some students to confuse the structural levels in proteins with those in nucleic acids. This is partially due to the structural level names of the two groups of macromolecules and the similarities in their definitions. Perhaps the most common error on exams that I have administered is confusion about the monomer building blocks for proteins and nucleic acids. Many students have said that proteins are polymers of nucleic acids or that amino acids constitute the building blocks of polynucleotides. This occurs even if a special effort is made to warn them that such errors are common.

It is valuable for students to learn the levels of structure of proteins; spend a significant amount of time on the topic. Emphasize the definition of each level, the bonds involved in maintaining the structure and the role that each level plays in the overall structure and function of polypeptides. *Primary structure* is the sequence of amino acids in the polypeptide from the N-terminal end (with its free amino group) to the C-terminal end (with its free carboxyl). Stress that the primary structure determines all of the higher levels of structure. There are a number of examples of this, but the classic example of sickle cell anemia hemoglobin is hard to beat. A change in one amino acid in a chain of 146, the β chain of hemoglobin, alters the higher levels of structure enough to cause sickle cell anemia. Sickle cell hemoglobin crystallizes under conditions of low oxygen tension leading to the disease symptoms. The sixth amino acid in the normal chain, which is a glutamic acid (polar, charged amino acid) is changed to a valine (a hydrophobic amino acid) in the mutant hemoglobin β chain. Due to the chemical differences of the component amino acid R groups, the normal formation of the higher levels of structure is disrupted.

Secondary structure is structure caused by interactions between adjacent or nearly adjacent areas of the backbone primarily in the region of the peptide bonds. It is stabilized by H bonds and comes in two varieties: the α-helix and the β-pleated sheet. Both of these structures make the maximal number of H bonds possible within the area of the backbone participating in the interaction and are, thus, equally stable. The major difference between them is that the H bonds in an α–helix are oriented parallel to the molecular axis, while the H bonds in a β-pleated sheet are oriented perpendicular to the axis. The regular arrangement of amino acids in the α–helix is a result of the rigidity of the peptide bond. This rigidity is, in turn, the product of resonance between the carbonyl group and its adjacent peptide bond. Electrons from the doubly bonded oxygen of the carbonyl occasionally move to the peptide bond. This prevents rotation around the bond as is typical of double bonds. Illustrate this concept to students using molecular models, pencils and Styrofoam balls or simply your fingers. The resultant rigidity leads to a structure called the amide plane within which lay the atoms of the polypeptide backbone from the central carbon of one amino acid to that of the next in line. The amino acid chain thus resembles successive, rigid planar structures joined by a single bond about which they can rotate.

Analogy

The Slinky Analogy

In space, the α-helix resembles a Slinky with the rungs of the Slinky corresponding to the polypeptide backbone, while the H bonds would connect successive rungs and thus be parallel to the axis of the molecule running through the center of the Slinky. The H bonds in an α–helix connect every fourth amino acid (i. e., amino acid 1 is H bonded to amino acid 5, etc.) as the chain makes one full turn around the helix every 3.6 amino acids. This brings, for example, the imino group (N-H) of amino acid 5 beneath the carbonyl group (C=O) of amino acid 1 so that an H bond can form.

Analogy

The Jacob's Ladder Analogy

The nearest example of the amide plane from common experience that comes to mind is a Jacob's Ladder, a child's toy. It consists of planar slats connected to each other by flexible straps. If you hold the top slat with the rest of the toy suspended below it, you can tilt it. Suddenly, the slat below it will be released and it will swivel downward causing the slat below it to be released. This chain reaction continues to the bottom of the ladder. If one lays this toy on its side and tries to pull it into as tight a circle as possible, there is a limit to how tight a circle can be formed. This is due to the planar slats that constrain the structure from forming too tight a circle. A similar constraint prevents the α–helix from having fewer than 3.6 amino acids per turn.

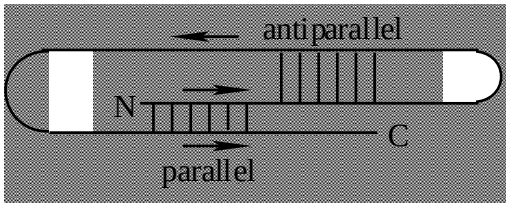
Keratin from hair is an example of a protein that contains lots of α–helix. It is extremely strong and extensible. If enough force is applied to hair to exceed the strength of the collective H bonds in the α–helices of the hair, they will separate and the polypeptide backbones will break, tearing the hair. Otherwise, once the force is removed, the helix will return to its original shape like a Slinky.

Image

The Analogy of Superman's Hair

When I talk about keratin and α-helices, I am reminded of reading Superman comic books as a child. Occasionally in those stories, visits to the Superman museum were illustrated. One of the displays at the museum consisted of a hair generously contributed by the Man of Steel to which was tied a one ton weight. Obviously, Superman's hair contains more H bonds than normal human hair.

The β-pleated sheet, as its name suggests, resembles a pleated sheet. It is strong as well as flexible, but not extensible. Fibroin, the major protein constituent of silk, has large amounts of β-pleated sheet. The H bonds that join adjacent parts of the backbone are oriented perpendicular to the molecular axes. If the two chains connected run in the same direction, the β-pleated sheet is called parallel. In contrast, if the H-bonded chains run in opposite directions, the structure is named antiparallel. This is illustrated in the drawing below:



Image

The Use of Silk During World War II

When I was a kid, I would often frequent a local Army - Navy surplus store. The store sold maps of China printed on silk priced at 50 cents. I bought a couple. The owner of the store, obviously trying to make the sale, told me why they were used. He said that pilots would take these maps with them during bombing missions over Japan. Since they could not land on the aircraft carriers from which they had taken off, they would continue over Japan and land in China. The maps were handy because they could be removed from pockets, quickly perused and then jammed quickly back into a pocket. Folding the maps carefully was unnecessary because of the strength and flexibility of the silk (fibroin). If the maps were paper, like those we can get from an auto club, even if there was time to fold them carefully, they would eventually tear. If they were crumpled up like the silk, they would be irretrievably wrinkled. Furthermore, there was always a chance that the pilots could land in an area of China occupied by the Japanese. The military would not want these maps to fall into enemy hands. To prevent this, the maps could be rolled into a little ball and swallowed. Since they are protein, they would be digested and unavailable to the enemy. This could not be easily done with a paper road map, which consists of polysaccharide fibers that could not be digested due to their β-glycosidic linkages.

Tertiary structure results from interactions between R groups within the same polypeptide chain. It involves hydrophobic interactions, H bonds, van der Waals forces, ionic bonds and a type of covalent bond, the disulfide linkage. The disulfide links form when two cysteine residues are moved into close proximity by protein folding. Cysteine residues have at the end of their R group a sulfhydryl (-SH) group. When two of these groups approach one another, the H atoms leave and a link forms between the two sulfur atoms, hence the name disulfide link. The disulfide links stabilize the folded structure.

Folding of the protein chains begins before their synthesis is completed. Since it involves adjacent or nearly adjacent regions of the polypeptide backbone, secondary structure forms first after the primary structure is laid down. This introduces some twists and turns into the chain, changes the shape of the molecules to some degree, and is followed by interactions between R groups that serve to fold the molecule even further. Eventually, the protein is folded into its proper three-dimensional shape. Once this has occurred, the disulfide linkages form and the molecule becomes fully functional. All the information needed by proteins to attain their proper shape is encoded within the protein sequence and many can fold completely without any help whatsoever. However, some proteins require the help of molecular chaperones or chaperonins to fold up. They speed up the folding of some molecules that would fold up much more slowly on their own. Furthermore, as the proteins fold into their final tertiary conformation, hydrophobic amino acid residues tend to end up in the center of the protein while the polar amino acids tend to wind up on the protein's outer surface exposed to water. When proteins fold, they do so in steps. After the first step, there are clues in the new structure to the next step. This process continues step-by-step, until the protein is completely folded. Each of these steps is called a nucleation state.

Analogy

The Garden Shears - Hedge Clippers Analogy

Garden shears and hedge clippers have a safety feature that keeps them closed when they are stored in your garage. It is a latch on one handle of the shears that fits over a post on the other handle. When the latch is engaged, the handles are secured and the shears cannot be opened until the latch is released. If the shears are open, it is impossible to close the latch over the post since the two of them are just not close enough. On the other hand, if the shears are closed, the latch can be engaged. This is analogous to what happens in proteins. The disulfide links do not form until the protein has essentially attained its final folded structure. The cysteine residues that will participate in the disulfide linkages are then close enough for the bonds to form.

Analogy

The Jig-Saw Puzzle Analogy
The folding of a protein into its final three-dimensional shape (its tertiary structure) is a lot like a jigsaw puzzle. When most people assemble a jigsaw puzzle, they look first for the corner pieces with two straight sides and then for other pieces with one straight side. These pieces are then assembled into the frame of the puzzle (the first nucleation state). Once the frame is completed, it provides clues for the next group of pieces to be laid in. These provide another set of assembly cues and so on until the puzzle is completed.

Quaternary structure involves interactions between R groups. All of the bonds that participate in tertiary structure can participate in quaternary structure. The only difference is that the interactions are between R groups on different polypeptide chains in quaternary structure, not between R groups on the same chain as in tertiary structure.

Nucleic Acids

The good news about nucleic acids is that many of the general features are understood by the bulk of Cell Biology students. This may be due to increased exposure to these basic principles in high school and introductory level courses in college or to the inherent elegant simplicity of A pairing with T and G with C. They are also familiar with the idea of the double helix, although confusion does arise once they have been introduced to the α-helix. They do, however, tend to get a little dim when the finer points of function are introduced.

General Structure of Nucleotide Monomers

Introduce the students to the components of nucleotides (phosphate + 5-carbon sugar + nitrogenous base) and nucleosides (5-carbon sugar + nitrogenous base). Once again, I do not require my students to memorize the structure so that they can draw it. I am content if they can look at a drawing and recognize the sugar, base and phosphate group of a nucleotide. Emphasize the negative charge on the phosphate group making it hydrophilic. When discussing the sugars, point out the important elements of their structure. Remind the students that sugars are generally hydrophilic. Stress the roles of each carbon in the sugar. Nitrogenous bases are attached to the 1'-carbon of the 5-carbon sugar. The 2'-carbon can be used to identify the sugar as ribose or deoxyribose, with the presence of an oxygen atom below the plane of the ring at that position indicating ribose and its absence indicating deoxyribose. The 3'- and 5'-carbons participate in the bonds that attach adjacent nucleotides. The 4'-carbon connects the 3'- and 5'-carbons. The nitrogenous bases are a bit harder for some students to grasp. Point out that these structures are composed of carbon-containing rings that contain an occasional nitrogen. Ask the students what this reveals about nucleotide chemistry and, if necessary, guide them to the answer (hydrophobic due to the ring structure). Tell them that the bases will occasionally have a group protruding from them that is capable of engaging in H bonds ($-NH_2$, $=O$). This, of course, means that bases are partially hydrophilic as well and, therefore, amphipathic. Once again, I do not expect the students to memorize the structures of bases, but I do expect them to be able to recognize the purines (adenine, guanine) with their two rings and distinguish them from the pyrimidines (thymine, cytosine, uracil) with their single ring structures. To aid them in remembering the differences between pyrimidine and purine structure, suggest that they associate the smaller name (purine) with the larger structure and the larger name (pyrimidine) with the smaller structure. Remind them that they may distinguish DNA and RNA by the presence of uracil in RNA replacing thymine (only found in DNA).

Biochemical analyses by Erwin Chargaff revealed that the number of adenine bases in a given DNA sample was equivalent to the amount of thymine bases and that the amount of cytosine was equal to the

amount of guanine. Furthermore, he found that the G+C/A+T ratio differs from organism to organism, while the pairing rules hold for all organisms. As an aside, I have found two books about the search for DNA structure and the origins of molecular biology enlightening: The Double Helix by James Watson and The Eighth Day of Creation by Horace Freeland Judson, a somewhat broader study.

Polymerization of Nucleotides and the Double Helix

Individual nucleotide monomers, whether RNA or DNA, are joined together by condensation reactions between the 5'-phosphate group of one nucleotide and the 3'-OH group on the sugar of another nucleotide. The resultant bond is called a 3'-5' phosphodiester bond. Break down the name of this bond to illustrate how seemingly intimidating lengthy words can be dissected to extract their meaning. The bond connects the 3' and 5' carbons of successive nucleotides in the chain. It is a combination of two phosphate esters which accounts for the rest of the name. The resultant chain has a 5' -> 3' polarity. In DNA, these polynucleotide chains usually come in pairs and they are antiparallel. A base in one chain pairs with the complementary base in the opposite chain. For example, a thymine nucleotide in one chain will pair with an adenine nucleotide in the opposite chain and a cytosine nucleotide will pair with a guanine nucleotide. This complementary relationship also explains the ability of DNA to replicate, an essential feature of the genetic material and its ability to pass the genetic code on to complementary RNA molecules that carry the code to the site of protein synthesis. The double stranded DNA that forms as a result of the pairing resembles a spiral staircase in shape. The phosphate-sugar backbone, which is hydrophilic, will form the railings of the staircase. The paired nitrogenous bases represent the actual stairs. The hydrophilic backbone essentially covers the outside surface of the molecules, while the hydrophobic paired bases are sheltered from water in the interior of the helix. They stack one on top of the other by way of hydrophobic interactions. As in proteins, the arrangement of hydrophilic groups on the outside and hydrophobic groups on the inside contributes much stability to the double helix.

The regularity of the width of the polynucleotide is also striking. Ask your students how they might have guessed about the lack of variation in the width even if they had not been told about it. Sometimes, a student will realize that the number of rings that fit across the helix at the site of each base pair is always three, since purines (2 rings) always pair with pyrimidines (1 ring).

Levels of Structure in Nucleic Acids

The levels of structure of polynucleotides, as mentioned previously, are often confused with those of polypeptides and it is important to stress the definitions of each level. *Primary structure* of polynucleotides simply refers to the nucleotide sequence from one end of the molecule to the other (usually from the 5' to 3' end). This applies to both DNA and RNA polynucleotides. Mention the effort to sequence the entire human genome and raise issues about the cost, potential benefits and potential disadvantages of this mammoth and now completed effort.

Students will frequently get secondary structure in proteins and nucleotides confused. This is invariably due to the similarity in the names of the α-helix and the double helix, which students easily mix up. I have not found a good way around this problem other than warning students each semester of the confusion that has plagued their predecessors. *Secondary structure* in DNA is the double helix. In RNA, secondary structure refers to local areas within the single RNA strand where intrastrand pairing can and does occur. It is useful to expose students to the likelihood of this happening.

Tertiary structure in polynucleotides is easy to define, but sometimes difficult for students to visualize. It is defined as any further coiling or twisting of the polynucleotide chain above the level of the double helix. The most common example of this is called supercoiling.

Analogy

The Telephone Cord Analogy

The easiest way to describe supercoiling to students is by using an example they are sure to have experienced, the telephone cord. If you have a phone cord, especially a long one on your kitchen phone, you will often walk aimlessly around the kitchen while talking on the phone. When you hang the phone up, you will usually discover that your helical phone cord (analogous to the double helix although it has only one "strand") will be hopelessly twisted. This extra twisting above and beyond that of the regularly coiled phone cord is analogous to supercoiling or tertiary structure in nucleic acids.

Analogy

The Balsa Wood Airplane with Rubber Band Analogy

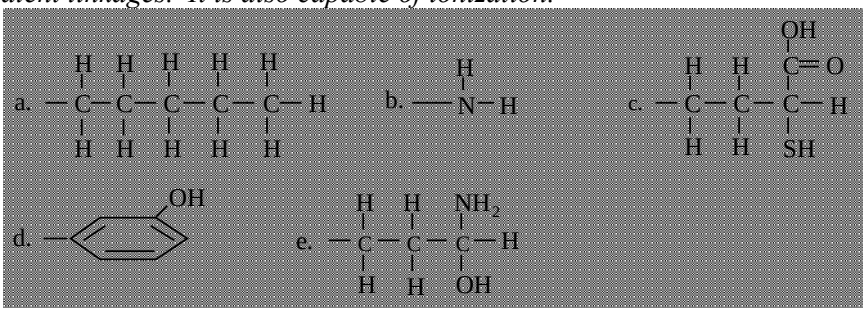
Another toy from childhood also demonstrates supercoiling. When we were kids, many of us at one time or another had balsa wood airplanes with propellers powered by rubber bands. We would turn the propeller that turns the rubber band and begin to twist it. At first, the rubber band would adopt a conformation reminiscent of the double helix. However, as we turned the band tighter and tighter, the rubber band would become supercoiled. This would be reflected by irregular bumps along the length of the rubber band. Interestingly, this supercoiling can be relieved by releasing the tension being applied to the propeller. The same sort of thing happens to relieve supercoiling in DNA. An enzyme (topoisomerase) will cut one of the strands of DNA and allow the supercoiling to be relieved before reconnecting the severed ends of the chain.

RNA Structure and Function

Briefly summarize the structure and functions of the different major types of RNA: ribosomal RNA (rRNA), transfer RNA (tRNA) and messenger RNA (mRNA). Remind the students of the differences between RNA and DNA summarized above, including the fact that RNA is usually single-stranded. Also mention here and elaborate in later lectures on the idea that RNA can sometimes serve as an enzyme, a ribozyme.

CRITICAL THINKING QUESTIONS

1. Which of the groups below is capable of only hydrophobic interactions? Explain your answer. *a. is capable of only hydrophobic interactions. It contains no ionizable or hydrophilic groups.* Which is capable of only hydrophilic interactions? Explain your answer. *b. is capable of only hydrophilic interactions since it has no component with a long carbon chain or a carbon-containing ring and no nonpolar covalent linkages. It is also capable of ionization.*



2. You treat a partially purified preparation of protein with a reagent that breaks bonds between sulfur atoms. Which level(s) of protein structure are likely to be affected the most? *Both the tertiary and quaternary levels of structure would be affected since those levels are the only ones in which disulfide bonds are prominent.*

3. Not all proteins are able to renature. Some proteins when exposed to heat or some other denaturing treatment are irreversibly denatured. What is an example of such a protein? *Egg white protein and yolk are examples of proteins that are irreversibly denatured by heat.*

4. You are working with an enzyme altase that you denature in the presence of urea. If altase were denatured no further by the addition of mercaptoethanol, what would that suggest to you about the enzyme? *The enzyme probably contained no disulfide linkages since mercaptoethanol breaks such linkages.*

5. Would all proteins be likely to require exposure to mercaptoethanol in order to accomplish full denaturation? If not, what trait would a protein that did not require mercaptoethanol possess?
Not all proteins would require mercaptoethanol to accomplish full denaturation. If a protein has no disulfide linkages, it probably would not require mercaptoethanol for full denaturation.

6. An enzyme is placed in a solution containing urea. Assuming that this protein contains no disulfide linkages, is it reasonable to suspect that it will be totally denatured by the treatment? *Placement in a urea solution should totally denature the enzyme, especially since there are no disulfide linkages.* How could you know that the enzyme has, in fact, been denatured? *If there are extensive hydrophobic interactions between enzyme R groups, total denaturation may be difficult to accomplish. If the enzyme activity disappears, there is a good chance the enzyme has been denatured.* Why does the urea denature the tertiary structure of the enzyme? *Urea breaks up the tertiary structure by interfering with hydrophilic interactions like H bonds.*

7. Which of the following tripeptides would be most likely to be soluble in an organic (hydrophobic) solvent like benzene: N - phenylalanine - alanine - glycine - C, N - leucine - alanine - lysine - C, N - proline - phenylalanine - leucine - C, N - arginine - lysine - proline - C, N - glutamate - aspartate - glycine - C? Explain your answer. *N - proline - phenylalanine - leucine - C would be most soluble in a hydrophobic solvent. All three amino acids are classed as nonpolar amino acids and could be soluble in benzene. In the other tripeptides, at least one of the amino acids does not belong to the nonpolar class.*

8. What level of structure in DNA would be disrupted by a reagent that breaks apart hydrogen bonds? *Secondary structure would be disrupted, because it is held together by hydrogen bonds. Hydrogen bonds are also involved in tertiary and quaternary structure. Thus, such a reagent would also disrupt these levels of structure in areas where H bonds are involved.*

9. DNA is isolated from two different species. Both DNA samples are found to be the same size. One of the DNA samples has a G+C/A+T ratio of 2.0 and the other 2.5. Which DNA sample has a higher G+C content? *The second sample has the higher G+C content since the ratio for that sample is the largest.* Which sample contains the smallest number of H bonds between strands? *The first sample contains a larger amount of A+T. Since A-T base pairs make only 2 H bonds, while G-C base pairs make three, the sample with the most A-T base pairs would have the fewest H bonds.* Which DNA sample would be easiest to denature? *The second sample would be easiest to denature since it is held together with the smallest number of H bonds.*

10. Mammals lack the enzyme that hydrolyzes cellulose. Yet many mammals are herbivores and they eat grass and other plant material for nutrition. How can this be, given that they cannot digest the food they are eating? *While these animals lack the enzyme that digests cellulose, bacteria that reside within their digestive tracts possess it. There is a symbiotic relationship between the two organisms.*

The herbivores seek out and eat the grass; the bacteria in their digestive tract digest it. What they don't use, the herbivore does.

11. You are a crew member on the starship Enterprise. Your responsibilities include investigation of biological life forms. You take out your tricorder after landing on the planet Yamihere and find a number of organisms, all of which contain DNA that follows the nitrogenous base pairing rules you are familiar with on Earth. For one species, the following relationships hold for the organism's DNA.

moles of adenine = 8 $\qquad \frac{A + T}{G + C} = 2$

How many moles of guanine are present? *4*

How many moles of thymine are present? *8*

How many moles of uracil are present? *0 (no uracil in DNA)*

You isolate DNA from another organism living on the surface of Yamihere and find that it contains all the bases normally found in DNA, but does not obey the pairing rules. Can you explain these strange results? *One possible explanation is that the DNA is single-stranded.*

12. What are some possible explanations for the branched structure of glycogen? *First, branching allows more efficient storage of energy. More glucose monomers can be stored in a smaller space. Second, branching creates more free ends on the structure. This would allow glycogen to be disassembled more rapidly when free glucose is needed and would also allow quicker assembly when glycogen is being constructed.*

13. Scientists have sequenced proteins by using specific proteases to "clip" a purified protein preparation between two specific amino acids, thus forming a number of moderately sized fragments; they have used acid hydrolysis to produce smaller fragments. Each fragment can then be sequenced by breaking the moderate fragments into dipeptides that are easily sequenced. The fragments below are obtained after the initial enzymatic cleavages. Can you deduce the sequence of the original polypeptide? (**HINT:** the original cleavages at specific locations differ depending on which proteolytic enzyme was used to create each fragment; this causes an overlap in the fragments' sequences.) The final polypeptide should have 18 amino acid residues.

N - ala - ala - gluN - aspN - met - C
N - iso - pro - aspA - try - thr - C
N - met - cys - leu - lys - phe - arg - aspA - C
N - aspN - met - cys - leu - lys - C
N - aspA - try - thr - phe - tyr - ala - ala - C

N- iso - pro - aspA - try - thr - phe - tyr - ala - ala - gluN - aspN - met - cys - leu - lys - phe - arg - aspA - C

14. Many so-called temperature-sensitive mutations have been discovered in a wide variety of organisms. These are proteins that are non-functional at higher temperatures, while, at lower temperatures (often just a few degrees lower), they function normally. For example, the coloration patterns in Siamese Cats arise from a temperature-sensitive mutation. An enzyme required for the synthesis of dark pigment is unable to function in areas close to the body where normal physiological temperatures prevail. However, at the tips of the ears, paws, the tip of the tail and other extremities

where the temperature is slightly lower, the enzyme works correctly and dark pigment is produced. What is happening at the molecular level that explains this? *In warmer areas of the organism, the temperature is just high enough to denature the enzyme in question. Since it is denatured, it will not work properly and dark pigment will not be produced in those areas.*

15. Which of the following tripeptides would be most likely to be soluble in an organic (hydrophobic) solvent like benzene? Explain your answer.

a. N - phenylalanine - alanine - glycine – C
b. N - leucine - alanine - lysine - C
c. N - proline - phenylalanine - leucine - C
d. N - arginine - lysine - proline - C
e. N - glutamate - aspartate - glycine – C

Answer c. N - proline - phenylalanine - leucine - C would be most soluble in a hydrophobic solvent. All three amino acids are classed as nonpolar amino acids and could be soluble in benzene. In the other tripeptides, at least one of the amino acids does not belong to the nonpolar class.

HUMAN PERPECTIVES QUESTIONS: FREE RADICALS AS A CAUSE OF AGING

1. What kinds of conditions can cause free radicals? *Free radicals may form when a covalent bond is broken such that each atom that had participated in the bond retained one of the two shared electrons that comprised the bond. They may also form when an atom or molecule accepts a single electron transferred during an oxidation - reduction reaction. Water, for example, can be converted into free radicals when exposed to solar radiation.*

2. Why are free radicals capable of altering molecules, such as proteins, nucleic acids and lipids? *They are extremely reactive which makes them well suited for chemically altering these molecules.*

3. Carcinogens are often identified as such by their ability to act as mutagens, i.e., their ability to alter DNA. In early tests, bacteria were used as indicators of mutagenicity. Later, the test was altered by exposing potential carcinogens to mammalian tissue extracts, e.g., liver, before exposure to bacteria. It was found that chemicals testing negative in the first test gave positive results in the second test.
   a. What is a possible explanation for this? *Mammalian liver contains enzymes that can convert some chemicals from nonmutagens to mutagens. Therefore, these chemicals are not mutagenic unless they are chemically altered by mammalian liver.*

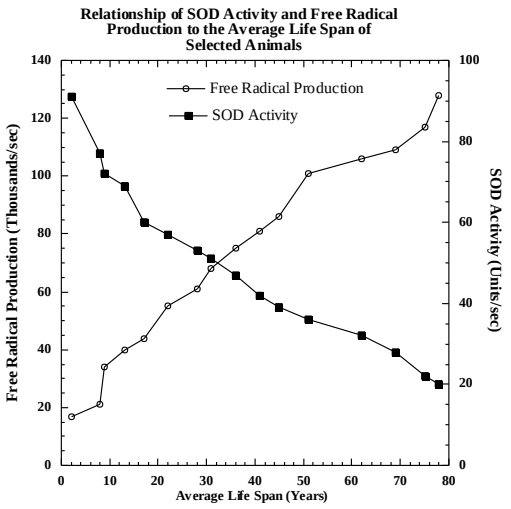
   b. Some investigators maintain that antioxidants in the diet will help reduce the number of free radicals and thus cancer and other diseases that plague human beings. Yet studies on cells growing in culture indicate that addition of antioxidants to the culture medium does not increase the growth ability of cells. What is a possible explanation for this? *While the cells may not be growing any better, they may be surviving longer.*

4. What is some specific evidence that demonstrates the importance of superoxide dismutase in getting rid of superoxide free radicals? *Mutant bacteria and yeast cells that lack SOD activity are unable to grow in the presence of oxygen. Furthermore, mice that lack the mitochondrial version of the enzyme (SOD2) are not able to survive more than a week or so after birth.*

5. Why might an organism that had functional SOD but mutant catalase and/or glutathione peroxidase be at a disadvantage? *SOD converts two superoxide free radicals and two hydrogen ions into hydrogen peroxide and oxygen. Hydrogen peroxide is also a highly destructive substance and without catalase and glutathione peroxidase, the organism would be less able to get rid of it.*

6. a. It has been hypothesized that aging results from tissue damage caused by free radicals. What are free radicals? *Free radicals occur when atoms or molecules have orbitals containing a single unpaired electron. They are highly unstable and extremely reactive chemical groups and are produced during normal metabolic processes. They can chemically alter many types of molecules, including proteins, nucleic acids and lipids; they may also damage tissues.*

b. Some time later, the enzyme superoxide dismutase (SOD) was discovered. The sole function of this enzyme is the destruction of the superoxide free radical. A connection between free radicals and aging has not been firmly established, but a few predictions assuming their involvement in aging have been made. Below are graphs depicting hypothetical data collected testing some of these hypotheses. Interpret the results of each graph. What do they tell you about the effect of free radicals on aging and the role of enzymes that neutralize free radicals? Consider each graph separately. Do not try to combine the results to form a coherent model.



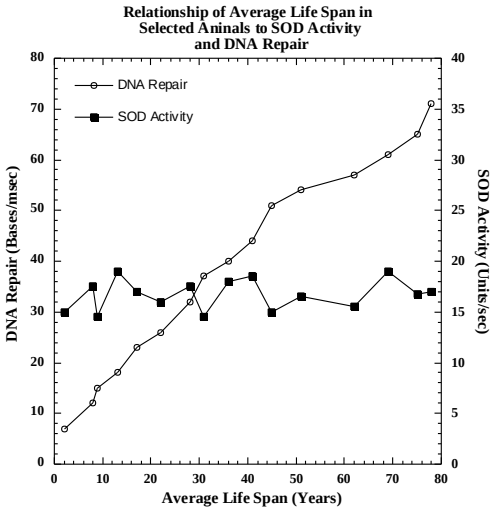


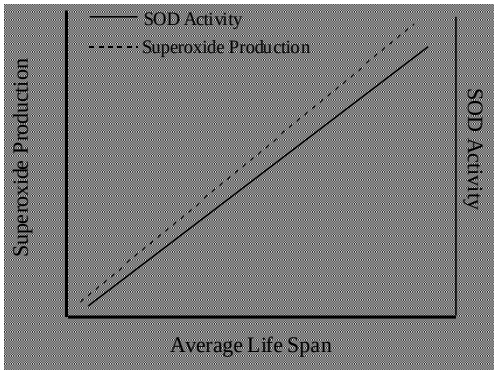
Figure 1. *Animals that live longer have higher SOD activity and correspondingly lower free radical production. This suggests that higher levels of free radical production correlate with shorter life spans. In addition, the ability to destroy the superoxide free radical reflected in higher levels of SOD activity correlates closely with longer life span. Therefore, the graphs suggest that organisms with higher SOD activity and/or lower free radical production will live longer.*

Figure 2. *This graph suggests that increased SOD activity has no influence on the life span of the organisms being monitored. However, an ability to repair DNA more efficiently appears to give organisms a better chance at having longer lives. Some of the DNA damage being repaired by the elevated repair enzymes may be caused by the chemical activity of free radicals.*

7. You isolate superoxide dismutase from two cell culture lines. One of the lines (SOD1) has a level of SOD activity similar to that found in liver, the tissue from which the cell line was originally obtained. The other cell line (SOD10) has elevated SOD activity. The enzyme in SOD10 is extremely efficient at converting the superoxide free radical to hydrogen peroxide. In a routine check of other critical enzyme activities, catalase was found to have activity levels that were severely depressed in SOD10,

while they appeared normal in SOD1. Observations of SOD10 reveal that this cell line cannot be maintained as easily as SOD1. SOD10 cells appear to die at an accelerated rate. What if anything can you conclude from these data? *While SOD10 is very efficient at neutralizing superoxide free radicals by producing hydrogen peroxide, the peroxide is toxic in its own right. SOD10 also has a relatively ineffective catalase, which detoxifies hydrogen peroxide. Thus, SOD10 builds up hydrogen peroxide rapidly but lacks the ability to neutralize it just as rapidly. The result is that these cells die at an accelerated rate.*

8. What would a graph similar to those above in question 6 look like if one could conclude from it that organisms that exhibit longer life spans also exhibit proportionately higher production of the superoxide free radical and correspondingly higher levels of SOD activity?



9. What happens to fruit flies that have been genetically engineered to produce large amounts of SOD? *They can live up to 40% longer than untreated controls.* Why do houseflies that are kept caged and unable to fly live longer than those allowed to fly? *Flying requires lots of energy and thus high metabolic rates. Thus, flies that are unable to fly have much lower metabolic rates and therefore require less oxygen. Consequently, they would be expected to produce fewer free radicals, which according to some would slow up aging.*

10. What are some common antioxidants found in the body? *Glutathione, vitamins E and C, beta-carotene (the orange pigment in carrots and other vegetables), and the parent compound for vitamin A.*

HUMAN PERSPECTIVES QUESTIONS: PROTEIN MISFOLDING CAN HAVE DEADLY CONSEQUENCES

1. What human disease was found to be similar to kuru in the brain abnormalities it caused? *Creutzfeld-Jakob disease (CJD) is similar to kuru.* What disease in sheep contributes its name to the abnormal prion molecule, PrP^{SC}? *The disease in sheep that contributes its name to the prion molecule is scrapie.* What have been the causes of outbreaks of acquired CJD? *Acquired CJD has been seen in recipients of organs and organ products that were donated by a person with undiagnosed CJD. Apparently, contaminated beef that the infected individuals had eaten years before has also been implicated as a cause of acquired CJD.*

2. What is spongiform encephalopathy? *This is a pathology in which certain brain regions are riddled with microscopic holes called vacuolations. It causes the tissue to resemble a sponge.*

3. When it was discovered that CJD could be acquired in addition to being inherited, why was it at first assumed that the infectious agent was a virus? *The infectious agent was found to pass through filters that retard the passage of bacteria. This is usually a characteristic of viral infections.*

4. How was it proved that CJD could be passed to another organism? *Extracts from the tissues of diseased individuals can be proved to be infectious if they transmit the disease to another individual. In the case of CJD, this was demonstrated across species with extracts from the brain biopsy of a human CJD victim causing disease in laboratory animals.*

5. An infectious agent is discovered that causes a particular disease. It has a relatively low molecular weight. Treatment with phenol or proteolytic enzymes, treatments that destroy proteins, render the infectious agent harmless, while treatment with nucleases and ultraviolet radiation, treatments that damage polynucleotides, has no effect. What is your interpretation of thee above data and why? *The sensitivity to protein-destroying treatments means that the agent contains protein and that the protein is important to the infectious process. The lack of effect of nucleic acid-destroying treatments suggests that nucleic acids are not important for infection and that the infectious agent is not a virus since nucleic acids are essential when viruses are responsible for an infection. The active part of the infectious agent above is clearly protein.*

6. How was it proved that the brains of patients suffering from CJD, an inherited disease, contain an infectious agent? *Carlton Gajdusek prepared extracts from a biopsy of the brain of a CJD victim. The extract was injected into a suitable laboratory animal. The animal developed a spongiform encephalopathy similar to that of kuru or CJD.*

7. Since replication is a property characteristic of nucleic acids, how might a prion, which lacks nucleic acids, "replicate" itself? *The mutant form of the protein in patients suffering from inherited CJD may act as a template that causes the conformation of the normal protein to convert to the abnormal form. The resultant two abnormal proteins could then convert two others, etc. The conversion of PrP^{C} to PrP^{Sc} has been accomplished in a test tube. Presumably, the appearance of the abnormal protein in the body, by whatever means, starts a chain reaction in which normal protein molecules in the cells are gradually converted to the abnormal prion form.* How can the inherited form of CJD be transmitted to another person? *A person who has the inherited form of CJD could transmit the disease to another person if they donate tissue or blood to a person who does not have the disease. The proteins in the donated tissue could then cause normal proteins in the recipient to shift conformation to the abnormal form. This could eventually lead to clinical CJD.*
8. How was kuru passed from one native of Papua-New Guinea to another? *During a funeral ritual, the mourners would eat the brain tissue of recently deceased relatives. If they had suffered from kuru, the disease could, and often would, be passed from the deceased relative to the mourners.*

9. What is the derivation of the name prion for the agent that can transmit diseases like CJD and kuru*? Disease transmission is by "**pr**ote**i**n **on**ly".*

10. Knockout mice are mice that have had one specific gene removed from their genome. This allows the role of the missing gene and its protein product to be assessed. Given this information, how would you explain the inability of mouse scrapie prions, which cause a malady similar to CJD, to cause the CJD-like disease scrapie in PrP knockout mice? *Since PrP knockout mice lack the PrP^{C} protein, there are no normal proteins in these mice to be converted to the mutant form; thus they do not develop the disease. With no PrP protein at all, normal or abnormal, the mice can still survive, since there appears to be no adverse effect if the protein is missing.*

11. What physical properties of the abnormal form of the PrP protein probably account for its ability to cause CJD? *Normal PrP (PrP^{C}) is a monomeric molecule that is soluble in salt solutions and readily destroyed by proteolytic enzymes. The abnormal PrP^{Sc} molecules are able to interact with each other to form insoluble fibrils that are resistant to enzymatic digestion.* What is odd about these differences given what is known about the structures of the two proteins? *The two proteins can have the same*

amino acid sequence but fold up differently to form significantly different three-dimensional structures. PrP^C consists almost entirely of α-helical segments and interconnecting coils. About 45% of a PrP^{cSc} consists of β-pleated sheet. The shift from a soluble, protease-sensitive conformation to an insoluble, protease-insensitive aggregate can be accomplished in vitro *by simply changing the conditions in the test tube.*

12. How does inherited CJD lead to the production of abnormal forms of PrP? *Under normal circumstances, the newly synthesized PrP polypeptide almost invariably folds into the PrP^C conformation. People with inherited CJD have a gene that encodes a mutant protein whose amino acid sequence is different from that of the normal protein. The mutant protein is presumed to be less stable in the PrP^C conformation than the normal version of the protein and more likely to fold into the abnormal β-pleated sheet-rich conformation. Once formed, the β-rich proteins produce aggregates, which lead to disease.*

13. In what ways are CJD and Alzheimer's disease similar? *Both are fatal neurodegenerative diseases that can occur in either an inherited or sporadic form. The brains of both CJD and Alzheimer's disease patients contain fibrillar deposits of an insoluble material. In both diseases, these toxic fibrillar deposits result from the self-association of a polypeptide composed primarily of β-pleated sheet.* What are the differences between the two diseases? *The proteins that form the disease-causing aggregates are completely unrelated. The parts of the brain that are affected are distinct and the protein responsible for Alzheimer's disease does not act like an infectious agent; it is nontransmissable.*

14. What surprising potential treatment for Alzheimer's disease has been demonstrated in a mouse animal model for the disease? *A group of investigators was able to create a strain of transgenic mice that developed amyloid brain plaques by introducing one of the mutant genes for human amyloid precursor protein (APP). They were able to block amyloid plaque formation by repeatedly injecting the animals with the same substance that causes the problem, the Aβ42 peptide. This caused the animals to produce antibodies against the peptides made in the brains of the mice by cleavage of the mutant APP protein. They were immunizing the animals against the disease. If the mice were injected when they were younger, they did not develop the amyloid deposits. If older mice whose brains already contained deposits were injected, many of the deposits were cleared out of the nervous system.*

15. What approaches other than immunization are being developed as treatments for Alzheimer's disease? *Drugs are being developed that inhibit the enzymes that cut the Aβ peptide out of the APP precursor, thereby reducing the production of the Aβ42 peptide. In an alternate approach, small peptides have been synthesized that can bind specifically to the α–□□□□□-rich version of the Aβ42 peptide and prevent it from refolding to the β–□□□□□□□-□□□□□-rich version. These peptides are called β-sheet breakers. They have a sequence of amino acids that is similar to a stretch of hydrophobic residues in the Aβ peptide that are involved in abnormal folding. The β-sheet breakers also contain a proline residue that inhibits the formation of a β-sheet. Injection of these β-sheet breakers into the brain of a rat with amyloid deposits blocks the continued formation of amyloid fibers and reduces the size of existing amyloid deposits. Because peptides are quickly destroyed in the body and generally unable to reach the brain, they are not likely to be very effective as drugs, but nonpeptide drugs with a similar structure have been designed and synthesized and may one day play a role in preventing Alzheimer's disease.*

EXPERIMENTAL PATHWAYS QUESTIONS

1. Why would heat shock proteins be present at low levels in cells or organisms raised at normal temperatures and then increase in number after a brief exposure to elevated temperatures? *At lower, normal temperatures, proteins would occasionally denature during the course of cell metabolism. Thus, a small number of these heat shock proteins would be enough to help them renature. At higher temperatures, proteins would probably denature at an accelerated rate. Therefore an increase in the number of the heat shock proteins will help the cell or organism survive the elevated temperature.*

2. When a protein denatures in a cell, what causes it to aggregate with other denatured cell proteins? *When proteins in the cell are denatured their interior hydrophobic amino acids are exposed to the aqueous cytoplasm of the cell. If the protein encounters other proteins that have become denatured in the same fashion, their hydrophobic amino acids are likely to aggregate rather than interact with the aqueous environment.*

3. How might a heat shock protein help a protein fold correctly in a cell under normal conditions? *Heat shock proteins promote polypeptide folding by binding to exposed hydrophobic patches on the surfaces of partially folded intermediates, preventing their aggregation with other proteins, which would be highly undesirable, since it would forestall the folding of the protein into its proper shape.*

4. You are working with the enzyme maltase that you obtain in a cell-free extract by homogenizing the cells that normally contain it. You heat the extract, a treatment known to denature the protein. When you cool the extract, the enzyme apparently renatures since most of its activity, which had disappeared when the extract was heated, returns rapidly. It appears that maltase is capable of self-assembly. Does this prove that maltase can renature without any assistance from other molecules like molecular chaperones? If not, describe an experiment that could prove that maltase is capable of self-assembly in the absence of molecular chaperones. *No, the experiment does not prove that the enzyme can renature without help from molecules like molecular chaperones. It is possible that some molecular chaperones were extracted along with maltase and that they aid the renaturation. If you purify maltase to homogeneity, there will be no other proteins present. If the completely purified protein renatures by itself after it has been denatured, one would conclude that no molecular chaperone is needed.*

5. The bacterial GroEL heat-shock protein and the plant Rubisco assembly proteins are said to be homologous proteins. What is it about these proteins that has caused them to be described as members of the same protein family and suggests that they have the same function? *The two proteins share the same amino acids at nearly half of the more than 500 amino acid residues in their respective molecules. The fact that they have retained so many of the same amino acids reflects their similar and essential function in the two types of cell.*

6. A molecular chaperone named GroEL is found in the cytoplasm of the bacterium *E. coli*. An homologous protein, the function of which is thought to aid in Rubisco assembly, is found in plant chloroplasts. Why would it not be surprising to find a protein so similar to GroEL, a prokaryotic protein, in these eukaryotic cell organelles? *According to the Endosymbiotic Theory, chloroplasts are derived from ingested photosynthetic bacteria that were not digested. It would not, therefore, be surprising to find a protein in chloroplasts that is similar to a protein found in prokaryotes like* E. coli.

7. Describe the evidence that convinced investigators that chaperones do not assist the assembly of already-folded subunits into larger complexes but that they instead assist polypeptide chain folding. *It has been known for some time that newly-made mitochondrial proteins produced in the cytosol have to cross the outer mitochondrial membrane in an unfolded, extended, monomeric form. During a study of molecular chaperones in mitochondria, a mutant was discovered that altered the activity of another member of the Hsp60 chaperone family that resided inside mitochondria. In cells containing*

this mutant chaperone, proteins that were transported into mitochondria failed to fold into their native conformation. Even proteins consisting of a single polypeptide chain failed to fold into its native conformation. This finding changed the perception of chaperone function from the notion that they assist assembly of already-folded subunits into larger complexes, to the current understanding that they assist polypeptide chain folding.

8. How do molecular chaperones help newly synthesized proteins cross membranes? *Usually, proteins must be in an extended, unfolded state to cross membranes. This is accomplished by their interaction with molecular chaperones.* What evidence suggests that molecular chaperones are necessary for this process? *Cells with defective molecular chaperones are unable to translocate polypeptides into membrane-bound organelles and the polypeptides accumulate in the cytoplasm of the cell.*

9. The GroEL complex is a chaperone protein that is shaped like a cylinder with a central cavity large enough to enclose a polypeptide undergoing folding. The central cavity is lined by a ring of hydrophobic residues. Why are non-native polypeptides able to bind to such a structure and what function(s) might that binding serve? *Since non-native polypeptides are denatured, their hydrophobic groups are exposed and will bind to the hydrophobic ring in the center of the GroEL cylinder. The binding of these proteins in the central cavity may remove a non-native polypeptide from an environment where it is likely to become aggregated. It may also unfold a polypeptide that has become misfolded, thus giving it a chance to refold properly.*

10. What is the apparent purpose of the GroES subunit binding to the GroEL cylinder when it is occupied by a non-native polypeptide? *GroES binds to the end of the GroEL cylinder and causes a conformational shift in the molecule. The shift elevates the cylinder wall and increases the volume of the enclosed chamber. In addition, the conformational shift buries the hydrophobic residues on the GroEL wall and exposes a number of polar residues. The non-native polypeptide is then released from the GroEL wall and displaced into the newly enlarged space where it can continue its folding in a protected environment. After about 15 sec, the GroES cap dissociates from the GroEL ring, and the polypeptide is ejected from the chamber. If the polypeptide has not attained its native conformation by the time it is ejected, it can rebind to the same or another GroEL, and the process is repeated.*

11. Why can it be said that a protein that folds inside the GroEL protein self-assembles? *It can be said to self-assemble since the GroEL chaperone only provides a safe location where folding can occur. The protein contains within its sequence all the information it needs to adopt its proper conformation.*

12. As many as 50% of the non-native soluble proteins of a bacterial cell can interact with GroEL. Given the fact that interactions between proteins are often highly specific, how is it possible that a single protein, like GroEL can bind to so many different polypeptides? *The GroEL binding site consists of a hydrophobic surface formed largely by two α-helices of the apical domain. This portion of the molecule is capable of binding virtually any sequence of hydrophobic residues that might be accessible in a partially folded or misfolded polypeptide. Comparison of the crystal structures of unbound GroEL and GroEL bound to several different peptides revealed that the binding site on the apical domain of a GroEL subunit can locally adjust its positioning when bound to different partners. This suggests that the binding site has structural flexibility that allows it to adjust its own shape to fit the shape of the particular polypeptide with which it has to interact.*

13. Describe an experiment that demonstrates that GroEL interacts with a broad spectrum of different proteins. *Ulrich Hartl and his colleagues incubated bacterial cells with labeled amino acids, lysed the cells and precipitated the GroEL-polypeptide complexes by adding anti-GroEL antibody. Several*

hundred different labeled polypeptides were found to be present in these immunoprecipitates, confirming that a great variety of newly synthesized polypeptides interact with GroEL.

ART QUESTIONS

1. Which atom(s) in Figure 2.1 are the least reactive and why? *Neon and argon are the least reactive atoms since they have full outer electron shells. Consequently, they are called inert (noble) gases.*

2. In Figure 2.2, you see a drawing of a salt crystal held together by ionic bonds. Are ionic bonds plentiful in living organisms? Explain your answer. *Ionic bonds are not very common in living organisms since their cells contain so much water, which interferes with such bonds. They can exist in areas of a living cell that restrict water, e.g., in the protein interior where hydrophobic R groups congregate.*

3. If H bonds are about 180 picometers long (Figure 2.4) and the strongest attraction between molecules participating in a single van der Waals interaction occurs when the molecules involved are separated by about 3.6 Å (Figure 2.6), which interaction is the strongest? *Since longer bonds are weaker as a general rule, the H bonds would be somewhat stronger than a single van der Waals interaction.*

4. a. Figure 2.6 exhibits the effect of distance on the attraction between two atoms. How would you describe the attraction or repulsion of two such atoms at a separation distance of 6 Å? *At a separation distance of 6 Å, there is virtually no attraction between two atoms.* What about at 4.5 Å? *At 4.5 Å, there is a slight attraction between two atoms.* How can you tell from this graph the distance at which the optimal attraction between the two atoms occurs? *The optimal attraction between the two atoms occurs at a distance of about 3.6 Å. This would be the separation distance on the x-axis corresponding to the point on the graph that projects farthest below 0 on the y-axis.*

   b. It is not unusual for a mutation to disrupt interactions such as these significantly. In these cases, one amino acid is often substituted for another. For example, a change in one amino acid in the hemoglobin β chains leads to the molecular shape changes that cause sickle cell anemia. How could such a change eliminate van der Waals interactions such as those illustrated in Figure 2.6? *A mutation making a significant change in the polypeptide chain, for example a hydrophobic amino acid residue exchanged for a polar, charged residue, would be likely to change the tertiary structure of the protein significantly. This might move normally adjacent parts of the molecule farther apart. If the distance between these two parts of the molecule were increased by 2-3 Å, the effect would be large enough to abolish the van der Waals attraction completely or nearly so.*

5. Figure 2.8 shows the interaction of a glucose molecule with water molecules via H bonds. What is the significance of such interactions with respect to glucose solubility? *In general, as the number of hydrogen bonds a molecule can make with water increases, the molecule's solubility will also increase.*

6. Figure 2.9 shows a ball and stick drawing of the structure of cholesterol. Are there any functional groups on this molecule that could allow even the slightest interaction with water? If so, what are they? *There is a hydroxyl group on the left side of the first ring in the figure. It could form H bonds with water.* If cholesterol is present in a membrane, which end is most likely to be directed to the outer surfaces of the membrane that are closer to the hydrophilic environments of the cell cytoplasm or the extracellular space? *The end with the hydroxyl group would be most likely to be exposed to the polar heads of the membrane phospholipids and the hydrophilic environments surrounding the membrane.*

7. a. Figure 2.10 demonstrates that water is released as a byproduct during condensation reactions and reintroduced across the same bond during hydrolysis reactions, resulting in breakage of the bond. If 1000 glucose molecules ($C_6H_{12}O_6$) were hooked together by condensation reactions, how many glycosidic bonds would be found in the resulting polymer? *999 glycosidic bonds.* After the formation of this polysaccharide, how many carbon, hydrogen and oxygen atoms are found in the polymer? *There would be 6,000 carbons, 12,000 hydrogens and 6,000 oxygens in 1000 glucose molecules. Since 999 glycosidic linkages would connect the 1,000 glucose molecules, 999 water molecules would be removed from the entire structure (1,998 hydrogens and 999 oxygens). Therefore, in the polysaccharide, there should be 6,000 carbons, 10,002 hydrogens, and 5,001 oxygens.*

   b. If a protein consists of 454 amino acids, how many hydrolysis reactions would be required to fully degrade the protein? *453 hydrolysis reactions.*

8. Some people are born with or can develop a condition known as lactose intolerance that causes them to suffer intestinal discomfort when they eat lactose-containing dairy products. This occurs because the lactose that can normally be metabolized and passed through the intestinal lining cannot do so in these individuals. Can you suggest an explanation for this? The structure of lactose is shown in Figure 2.16. *People affected by lactose intolerance lack the enzyme that breaks the bond between glucose and galactose, the two sugars that combine to form the disaccharide. Therefore, lactose remains in the intestine leading to the symptoms of lactose intolerance.*

9. In Figure 2.17, there are schematic drawings of glycogen, starch, and cellulose.
   a. Which of these polysaccharides would be most likely to allow the quickest release of glucose monomers during hydrolysis? *Glycogen.* Why? *Because it is branched and thus possesses more free ends from which glucose can be released.*
   b. Which polysaccharide(s) could be used as a fuel source by an organism that lacks enzymes that break α-glycosidic linkages but possesses enzymes that can break β-glycosidic bonds? *Cellulose.*
   c. What kind of bond is marked by the number 3 in the figure? *β (1-> 4) glycosidic bond.*
   d. What kind of bond is denoted by the number 1 in the figure? *α (1-> 6) glycosidic bond.* What feature of the molecule in question requires this bond? *This bond is required for branching.* What is (are) advantages of this feature? *The advantages are more efficient packing of more glucose residues in a smaller space and more free ends on the molecule to facilitate more efficient and rapid release of glucose monomers when they are needed.* If this bond could not be formed, what would the molecule look like? *Without this bond, the molecule would be linear with no branching. All monomer glucose units would be connected with α (1-> 4) glycosidic bonds.*

10. Which molecule in Figure 2.19 contains double bonds in at least some of its fatty acid chains? *Linseed oil.* Which molecule contains no double bonds in its fatty acids? *Tristearate.*

11. Is the phospholipid in Figure 2.22, saturated or polyunsaturated? How do you know? *It is saturated because the fatty acid tails contain no double bonds. Also, the chains are straight; they would be kinked if they were unsaturated. A kink would occur at each double bond.*

12. Which amino acid in Figure 2.26 would be most likely to form covalent bonds between two different polypeptide chains? *Cysteine.* Which amino acid would be most likely to be found at a kink in an amino acid chain? *Proline.*

13. In Figure 2.29, there is a scanning electron micrograph of a sickled red blood cell. If hemoglobin were isolated from this cell and others like it and subjected to chromatographic separation after

enzymatic digestion, one spot, representing a single peptide, would differ on the chromatograms of normal and sickle cell hemoglobin. How many amino acids are changed in the mutant form of hemoglobin to cause the difference in the two chromatograms? *The single affected peptide has a one amino acid difference between the normal and mutant forms of hemoglobin.*

14. Which of the structures shown in Figure 2.30 contains H bonds oriented perpendicular to the molecule's axis? *None of them. All of the structures are α-helix, the H bonds of which are oriented parallel to the molecule's axis.*

15. Which of the structures shown in Figure 2.31 contains H bonds oriented parallel to the molecule's axis? *None of them. All of the structures are β-pleated sheet, the H bonds of which are oriented perpendicular to the molecule's axis.*

16. The denaturation of ribonuclease is depicted in Figure 2.42. What role in denaturation is played by β-mercaptoethanol? *β-mercaptoethanol breaks disulfide bonds making denaturation occur more readily.*

17. In Figure 2.52a, what kind of nitrogenous base appears in the drawing? To which carbon of the nucleotide sugar is it attached? *It is a purine base, specifically adenine; it is attached to the 1'- carbon of the nucleotide sugar.* In Figure 2.52b, which end of the polynucleotide shown is the 5' end and which the 3' end? *The end of the polynucleotide nearest to the top of the page is the 5' end. The end nearest to the bottom of the page is the 3' end.*

17. Consult Figure 2.53. What is the difference between the pyrimidine nitrogenous bases uracil and thymine? *Thymine differs from uracil by a methyl group attached to the ring.*