Biopolymers
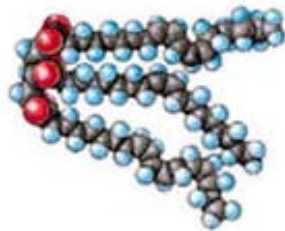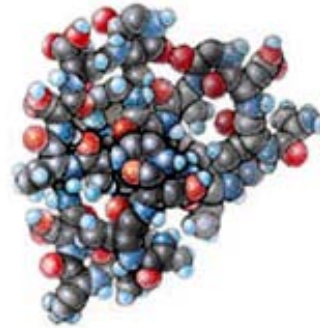

| Carbohydrate (starch) | Lipid (triacylglycerol) | Protein (enzyme) | Nucleic acid (DNA) |

We already surveyed lipids and their "self-assembly" of bilayer and vesicle polymers, during the section on water. We'll skip carbohydrates and turn to: Proteins and nucleic acids

Note: Your textbook covers nucleic acids in much more detail than proteins, so we will spend most time on proteins

Chemical evolution: The primary problem

Theory for origin of life by chemical evolution must explain following:

nuclei ➔ atoms ➔ molecules ➔ monomers ➔ polymers

It's the last step that is the problem, as we'll see.

But biological polymers are clearly special, with potential for multiple stages of hierarchical structure. How is this possible? ➔ Carbon-based molecules can bend, twist, and fold, *reversibly*. Consider protein shown below. "Polypeptide" on far left is already a polymer (amino acids are the monomer).
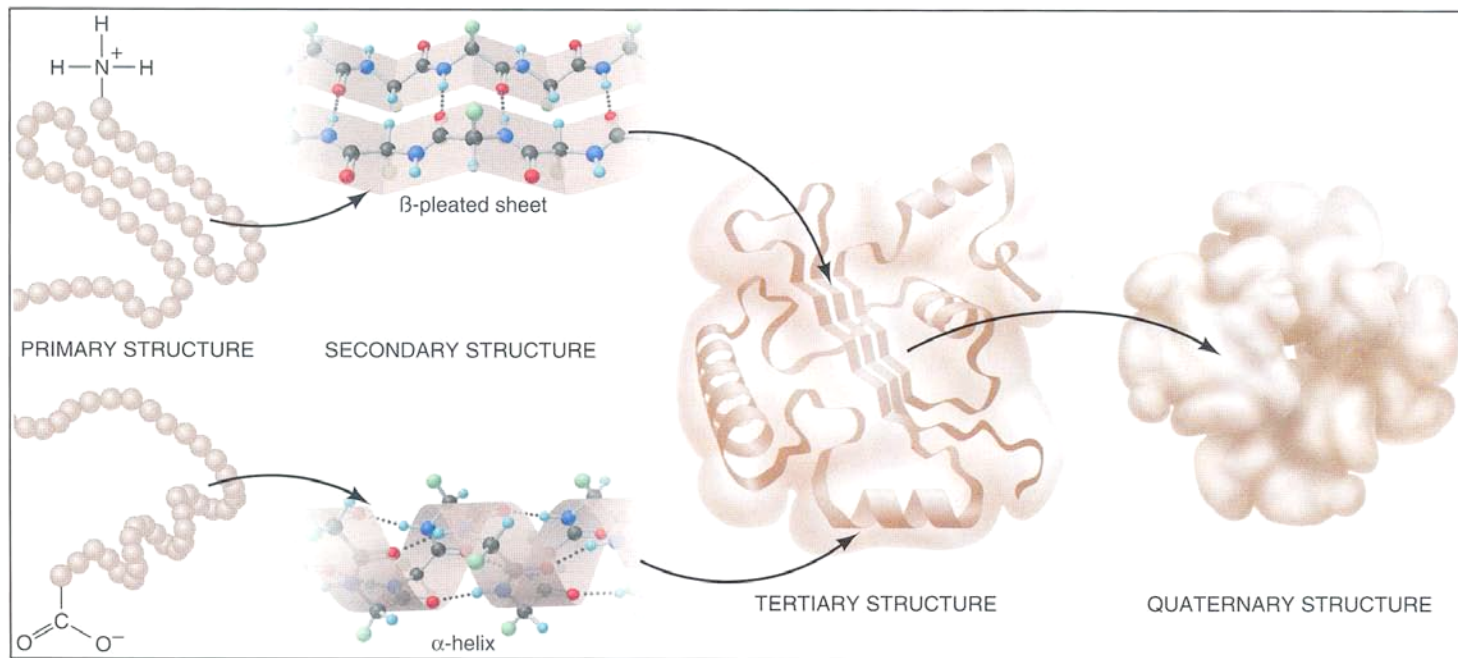
Premier example of the hierarchical levels of structure in biological polymers: Proteins
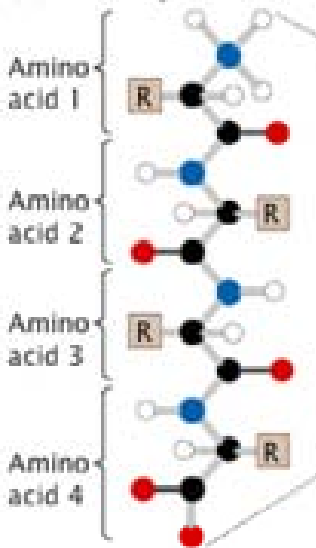

The primary structure of a protein is its sequence of amino acids.

(a) Primary structure

Interactions between amino acids cause the primary structure to fold into a secondary structure, such as this alpha helix.

(b) Secondary structure
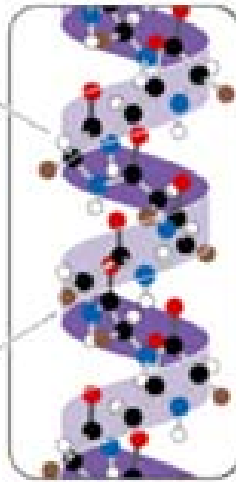
The secondary structure folds further into a tertiary structure.

(c) Tertiary structure

Two or more polypeptide chains may associate to create a quaternary structure.

(d) Quaternary structure

Amino acid 1
Amino acid 2
Amino acid 3
Amino acid 4
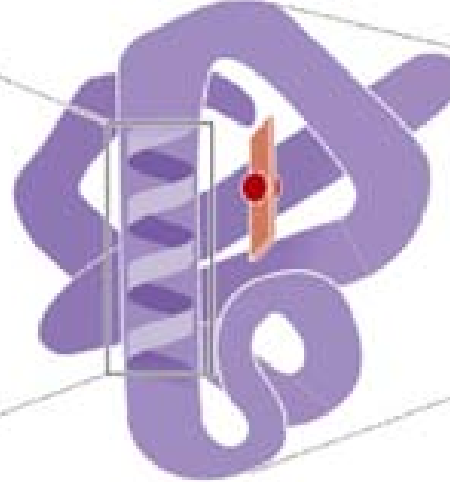
polypeptide = chain of amino acids = primary structure
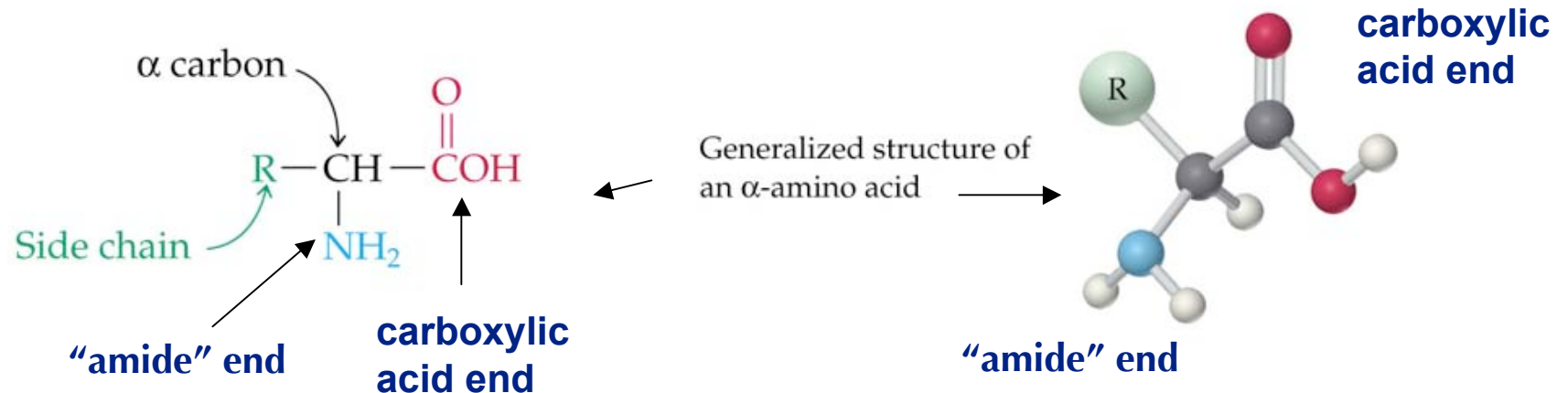
alpha helix secondary structure

Folded helix = tertiary structure

Several tertiary structures form the quaternary structure

The "**side chain**" ("R") is responsible for the specific functions of protein molecules, into which amino acids **polymerize** (join in a chain). Some are acidic, or charged, or polar (hydrophilic), or a number of other properties. As an example, a chain of amino acids (a **polypeptide**) in water twists in ways that keeps the water-loving (*hydrophilic*) amino acids on the outside of the resulting folded polymer. If charged, that spot on the polypeptide may be able to bond easily with other molecules, especially with other spots on the chain!

α carbon

Side chain

R—CH—COH

O

NH₂

"amide" end

carboxylic
acid end

Generalized structure of
an α-amino acid

carboxylic
acid end

R

"amide" end

Visualize in 3D: Are the upper and lower
amino acids the same molecule? Imagine
making a chain of them end to end.

Nineteen of the twenty common amino acids are *chiral* because they have four different groups bonded to the alpha carbon. Only glycine is *achiral* since it has two hydrogens attached to the alpha carbon and would have a plane of symmetry. (See next slides)

Chiral *Amino*, **Achiral** *Glycine*; *Propane*
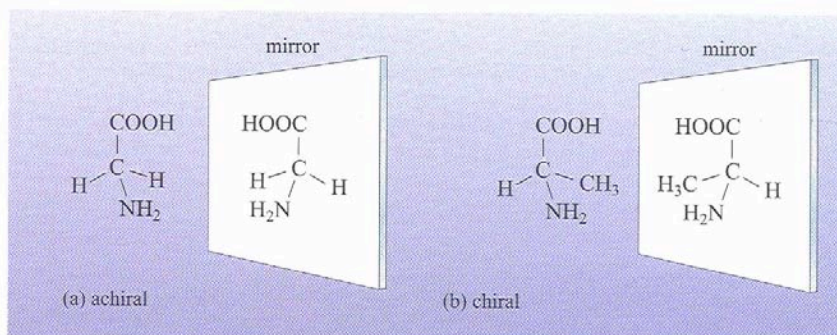

Mirror

(a) achiral   (b) chiral
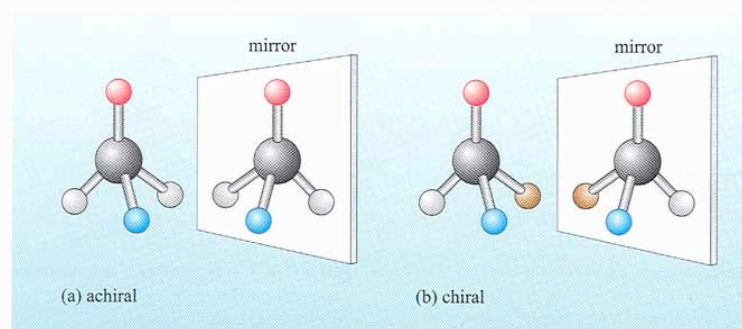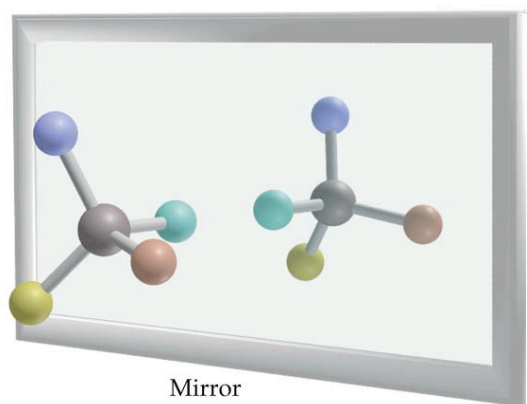
(a) An achiral molecule. Its mirror image can be superimposed on the original. (b) A chiral molecule. The mirror image cannot be superimposed on the original.


COOH

H — C — NH$_2$

CH$_3$
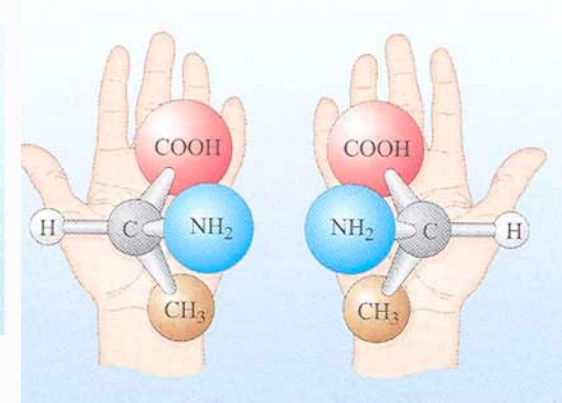
The chiral amino acid alanine can be present in left- and right-hand forms

A molecule with a carbon atom that is bonded to four different groups is **chiral** and is not identical to its mirror image. It thus exists in two **enantiomeric** forms.

(a) achiral   (b) chiral

(a) The achiral amino acid glycine. Its mirror image can be superimposed on the original. (b) The chiral amino acid alanine. The mirror image cannot be superimposed on the original.

"Right-handed"   "Left-handed"   Propane—achiral
D-alanine   L-alanine
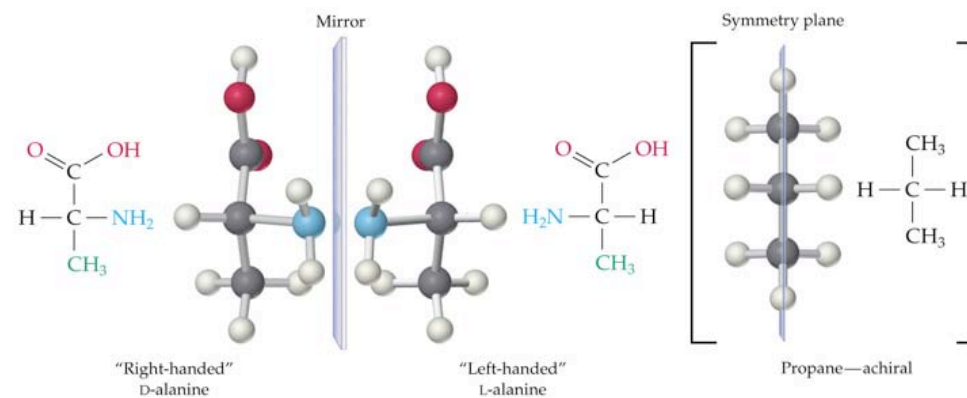
Alanine (2-aminopropanoic acid) has no symmetry plane and can therefore exist in two forms—a "right-handed" form and a "left-handed" form. Propane (not an amino acid), however, has a symmetry plane and is achiral.

"**Chiral**"= molecule whose mirror image configuration **cannot** be superimposed on the original. This is like a left and right hand: They are mirror images, but you can't superimpose a right glove on a left glove.

"**Enantiomer**" refers to one of the two forms of a chiral molecule, denoted L- and D-. An **achiral molecule** is identical to its reflection, so has only one form.

Another way to see it: Can you draw a plane through the molecule that divides it into identical parts? Alanine has no symmetry plane and can therefore exist in two forms—a "right-handed" form and a "left-handed" form. Propane (not an amino acid), however, has a symmetry plane and is achiral. The interesting molecules for origin of life are *chiral* i.e. *not* symmetric.

Mirror

Symmetry plane

"Right-handed"
D-alanine

"Left-handed"
L-alanine

Propane—achiral
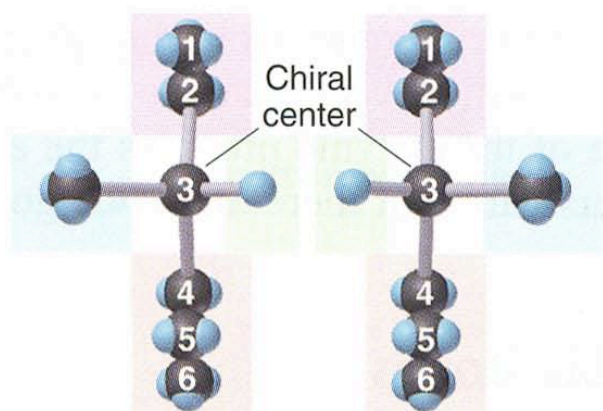
A Optical isomers of 3-methylhexane
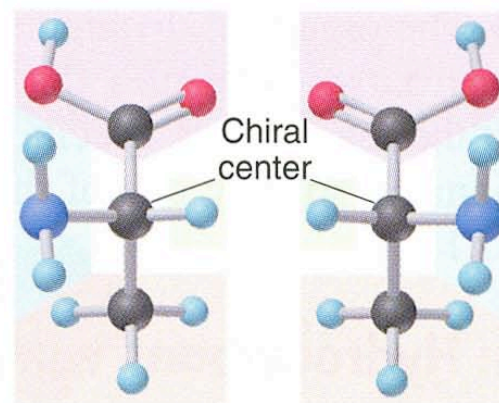
B Optical isomers of alanine

Figure 15.9 Two chiral molecules. A, 3-Methylhexane is chiral because C-3 is bonded to four different groups. These two models are optical isomers (enantiomers). **B,** The central C in the amino acid alanine is also bonded to four different groups.

Peptide Bond

(a)
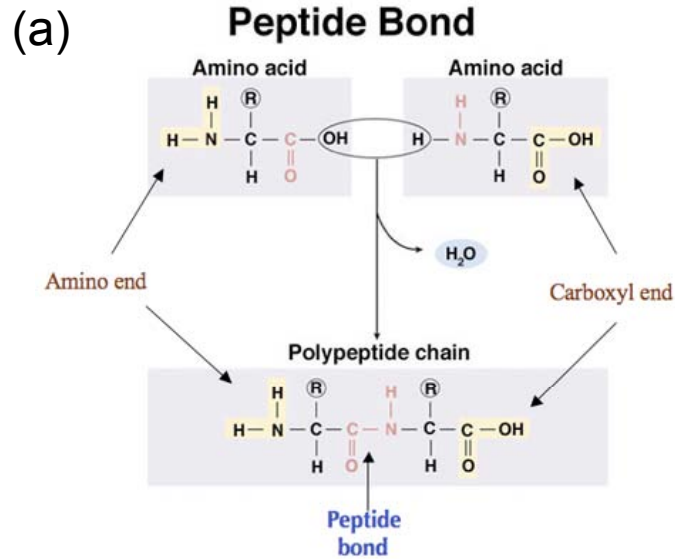
Peptide Bond

(a) A polypeptide is formed by the removal of water between amino acids to form peptide bonds. Each aa indicates an amino acid. R_1, R_2, and R_3 represent R groups (side chains) that differentiate the amino acids. R can be anything from a hydrogen atom (as in glycine) to a complex ring (as in tryptophan).

(b)

R groups project out from the C–N backbone
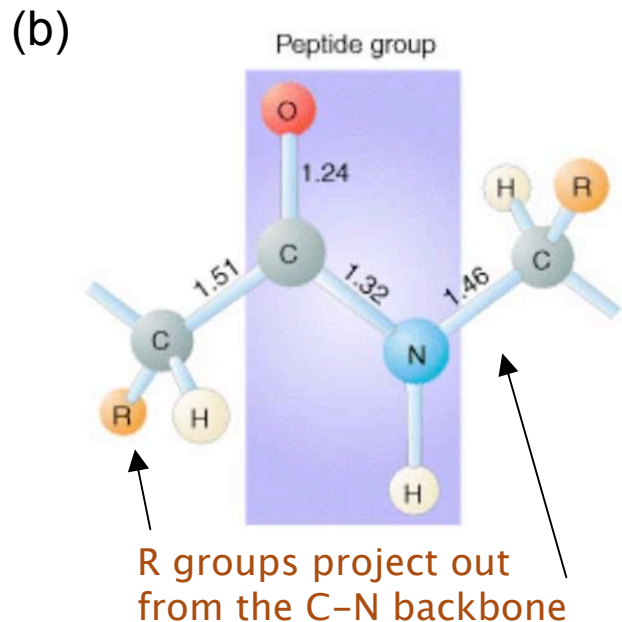
(b) The peptide group is a rigid planar unit with the R groups projecting out from the C—N backbone. Bond distances (in angstroms) are shown.

The key to protein folding is the variation in bond angles (through bending and rotating) that are possible for various amino acids.

NONPOLAR

CH₃
CH₃—CH
CH₂
H₃N⁺—C—COO⁻
H
Leucine

POLAR

OH
CH₂
H₃N⁺—C—COO⁻
H
Serine

SH
CH₂
H₃N⁺—C—COO⁻
H
Cysteine

IONIC

⁺NH₃
CH₂
CH₂
CH₂
CH₂
H₃N⁺—C—COO⁻
H
Lysine

COO⁻
CH₂
CH₂
H₃N⁺—C—COO⁻
H
Glutamic acid
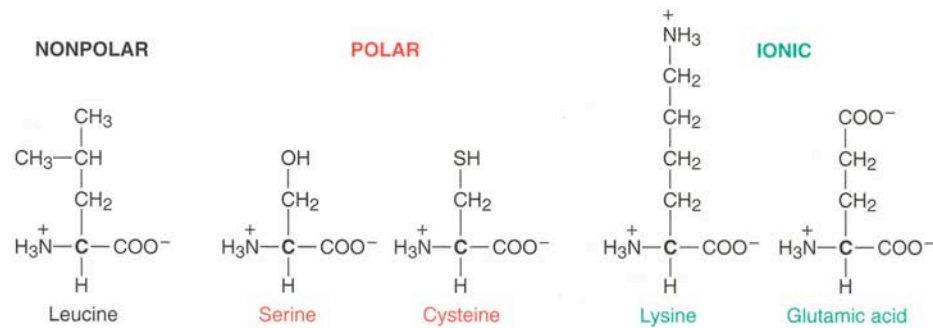
Look at these five amino acids and see that they have precisely the same structure at three of the four bonds with the alpha carbon, so **the variation in the *side chains* is responsible for the ways in which proteins can fold**. This is amazing because each amino acid is a representation of three "letters" of the genetic code, so **these side chains are the agents that convert a linear code into a functional three-dimensional structure.**
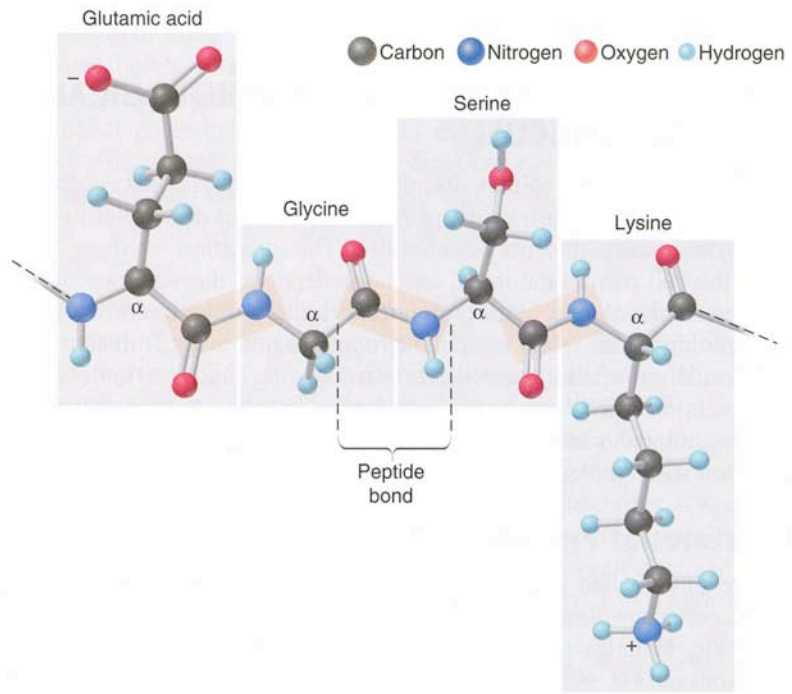
Glutamic acid

● Carbon ● Nitrogen ● Oxygen ● Hydrogen

Serine

Glycine

Lysine

α   α   α   α

Peptide bond

Peptide bonds hold monomers together in proteins. Three peptide bonds (orange screens) joining four amino acids (gray screens) occur in this portion of a polypeptide chain.

Note the repeating pattern of the chain: peptide bond--alpha carbon--peptide bond--alpha carbon-- and so on.

Also note that the side chains dangle off the main chain, or "backbone."

Secondary Protein Structure

Alpha helix
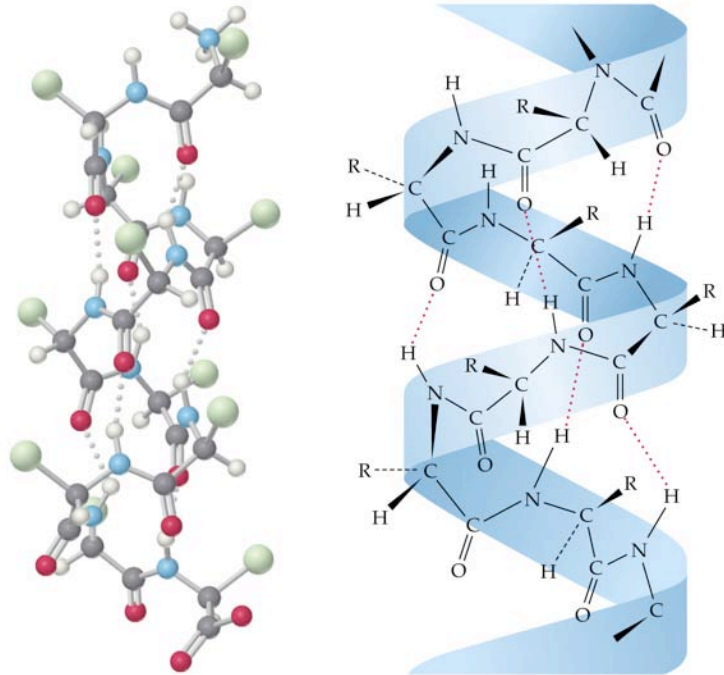
Beta sheet

The a-helical secondary structure of keratin. The amino acid backbone winds in a right-handed spiral, much like that of a telephone cord.
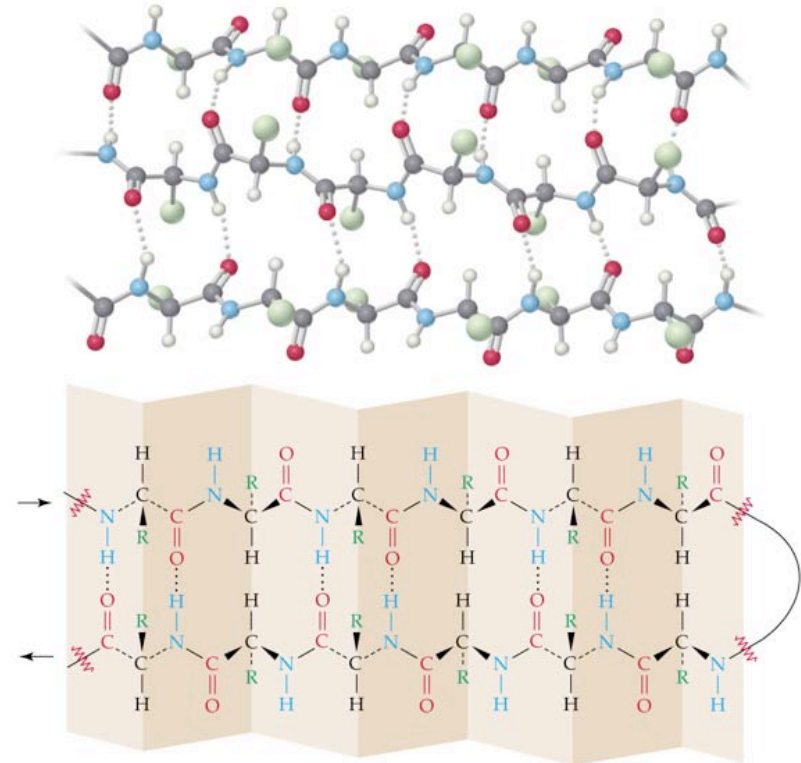
The b-pleated-sheet secondary structure of silk fibroin. The amino acid side chains are above and below the rough plane of the sheet. (Dotted lines indicate hydrogen bonds between chains.)

Secondary, tertiary structure of Myoglobin
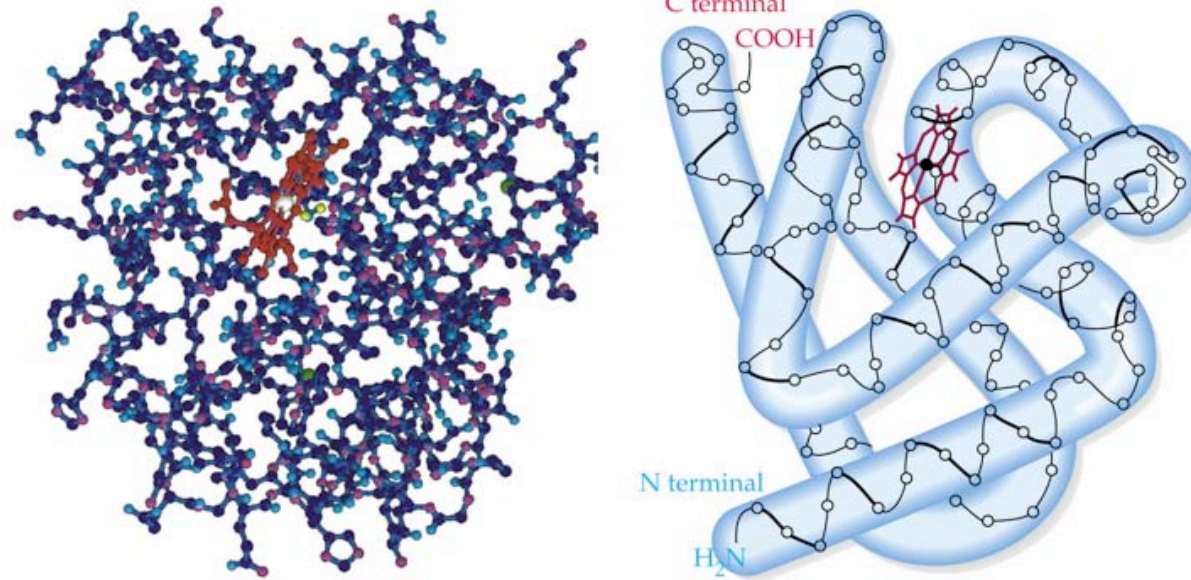
Secondary and tertiary structure of myoglobin, a globular protein found in the muscles of sea mammals. Myoglobin has eight helical sections.

As if proteins don't seem complex enough…
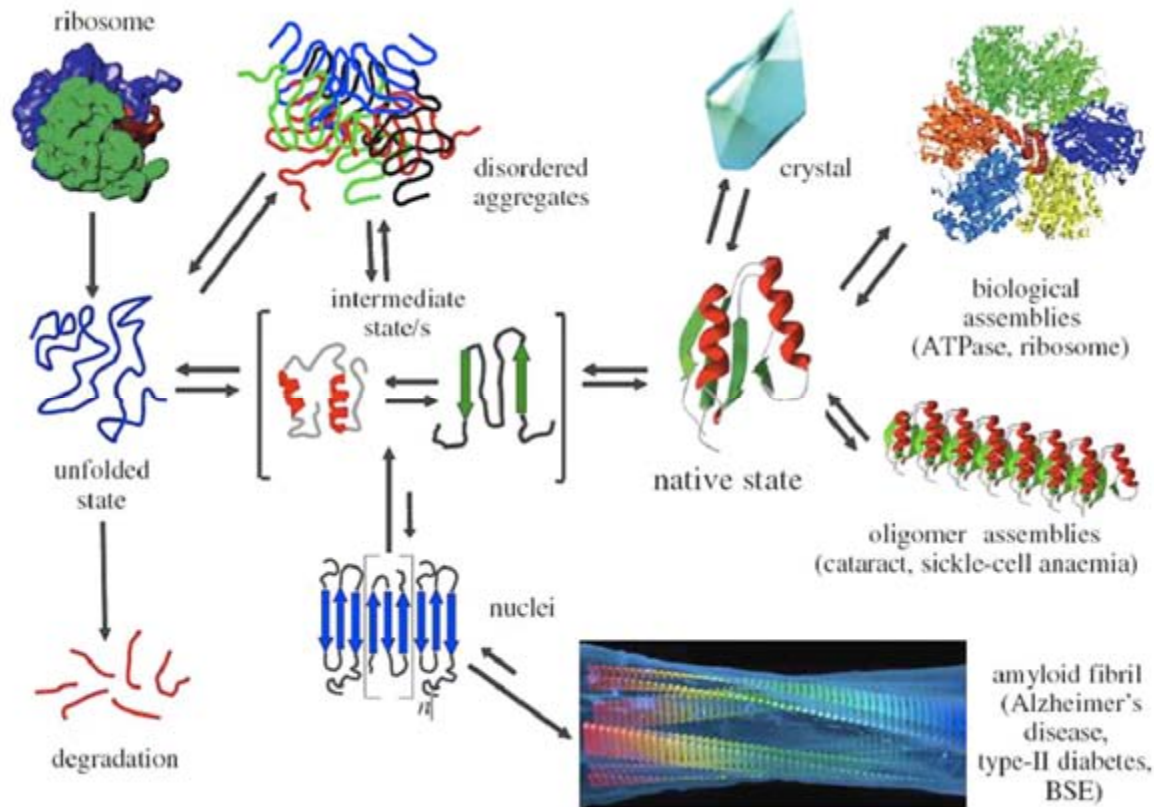
Consider how, without much extension to higher or lower levels of complexity, one biomolecule can be viewed in a large number of ways--we have no idea which of these might have been most important for the transition from nonlife to life.

Nucleic Acids: DNA, RNA

Tthese molecules are the basis for the genetic material of all life on Earth, and so are central for our speculations about life elsewhere. They consist of sequences of *nucleotides*, which are three chemical groups bonded together: one of four (or five) bases, a particular sugar, and a phosphate group. These nucleotides somehow became capable of linking up, or polymerizing, to form long sequences called nucleic acid, either single-stranded (RNA) or double-stranded (DNA).

1. a *nitrogenous* <u>base</u> – these 5 or 6 sided ring molecules are the "alphabet" of the genetic code. There are two classes, called <u>purines</u> (A, G; structurally they are two connected rings) and <u>pyrimidines</u> (C, T in DNA, C, U in RNA; structurally a single ring). These bases are illustrated below.

2. a 5-carbon <u>sugar</u>: (ribose for RNA, 2' -Deoxyribose for DNA – just ribose with one of the O atoms removed at the "2"' site of the sugar ring)

The joined base + sugar is called a *nucleoside*. Study the picture to the right:

3. a phosphate ~ (P + 4 O's), where P is phosphorus.

The base + sugar + phosphate is called a *nucleotide*. That extra phosphate group turns out to be very important, and many people think that phosphorus has several unique properties that make it an optimal (and maybe the only) choice for the third monomer of nucleic acids.
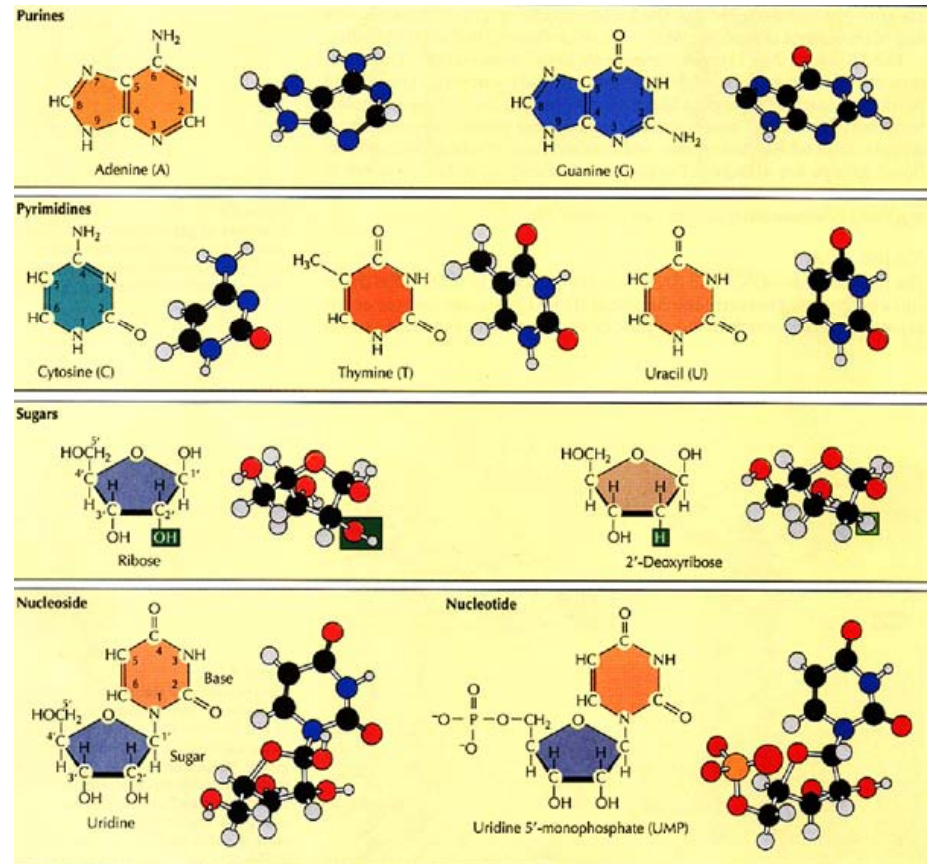
2. nucleotide

The base may be either a pyrimidine or a purine.

Base + Ribose or deoxyribose = Nucleoside + Phosphate = Nucleotide

Pyrimidines

Cytosine (C)   Thymine (T)   Uracil (U)

Purines

Adenine (A)   Guanine (G)

Nitrogenous base

Phosphate group

OH in RNA
H in DNA

Sugar

(a) Structure of a nucleotide
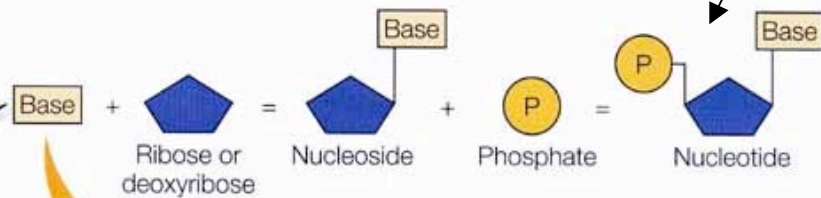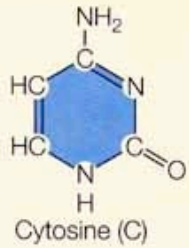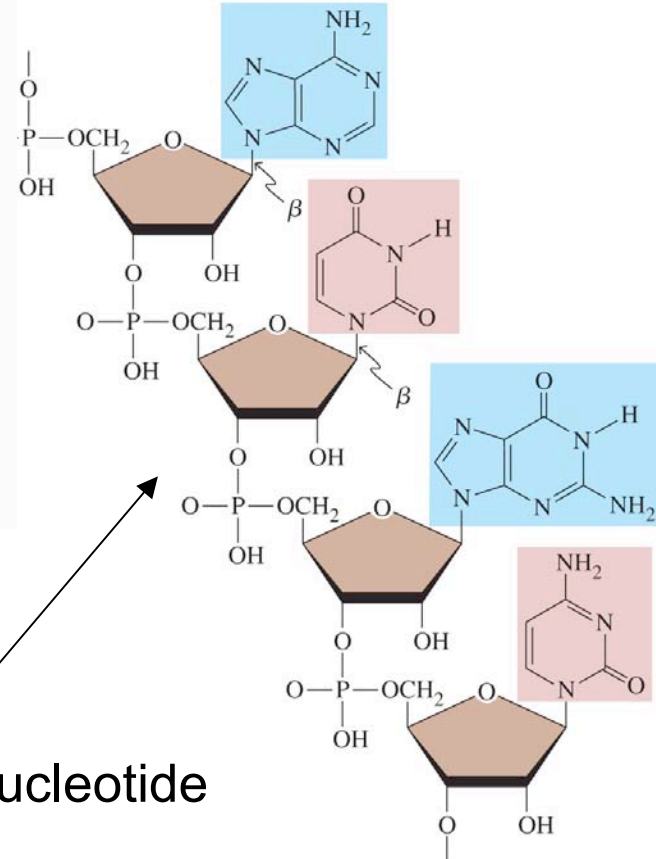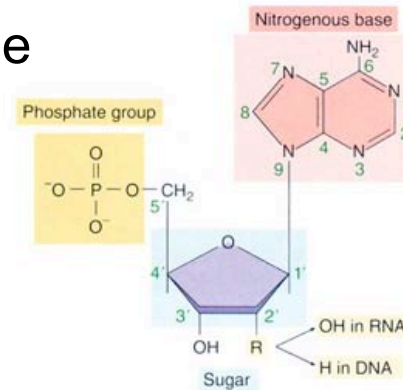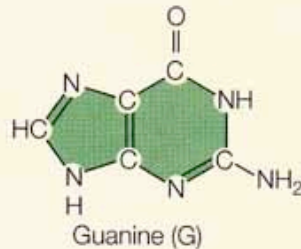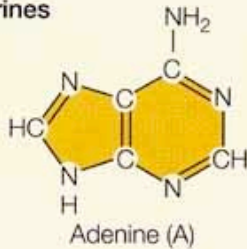
1. Bases: purines and pyrimidines
Pairing: Only (G,C), (A,T) can pair across the double helix

3. polynucleotide

Another view of a single strand of nuclei acid

FIGURE 3–15 Nucleotides and nucleic acids. A nucleotide (center box) has three parts: one to three phosphate groups (yellow), a sugar (brown), and a base (blue-green). The ring-shaped nitrogenous bases (lower box) are derived from one of two parent compounds, purine and pyrimidine. A nucleic acid polymer (far right box) consists of nucleotide monomers, linked with the phosphate group of one attached to the sugar of the next.

Polynucleotide chains for DNA, RNA
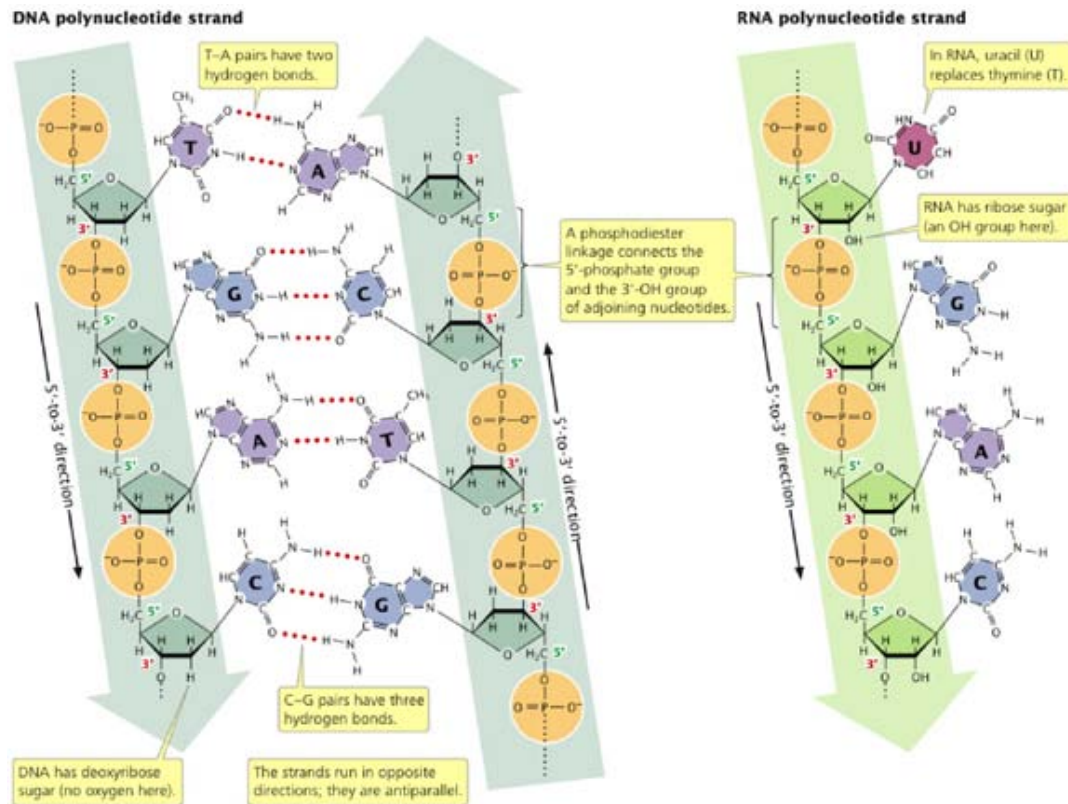
- DNA consists of two polynucleotide chains that are antiparallel and complementary, and RNA consists of a single nucleotide chain. Typical lengths of these sequences of nucleotides is ~ 10^5 atoms for RNA, and 10^{10} atoms in DNA. If you could straighten out DNA, would be mm (bacteria) to cm (vertebrates) long.

Complementary base pairing and stacking in DNA
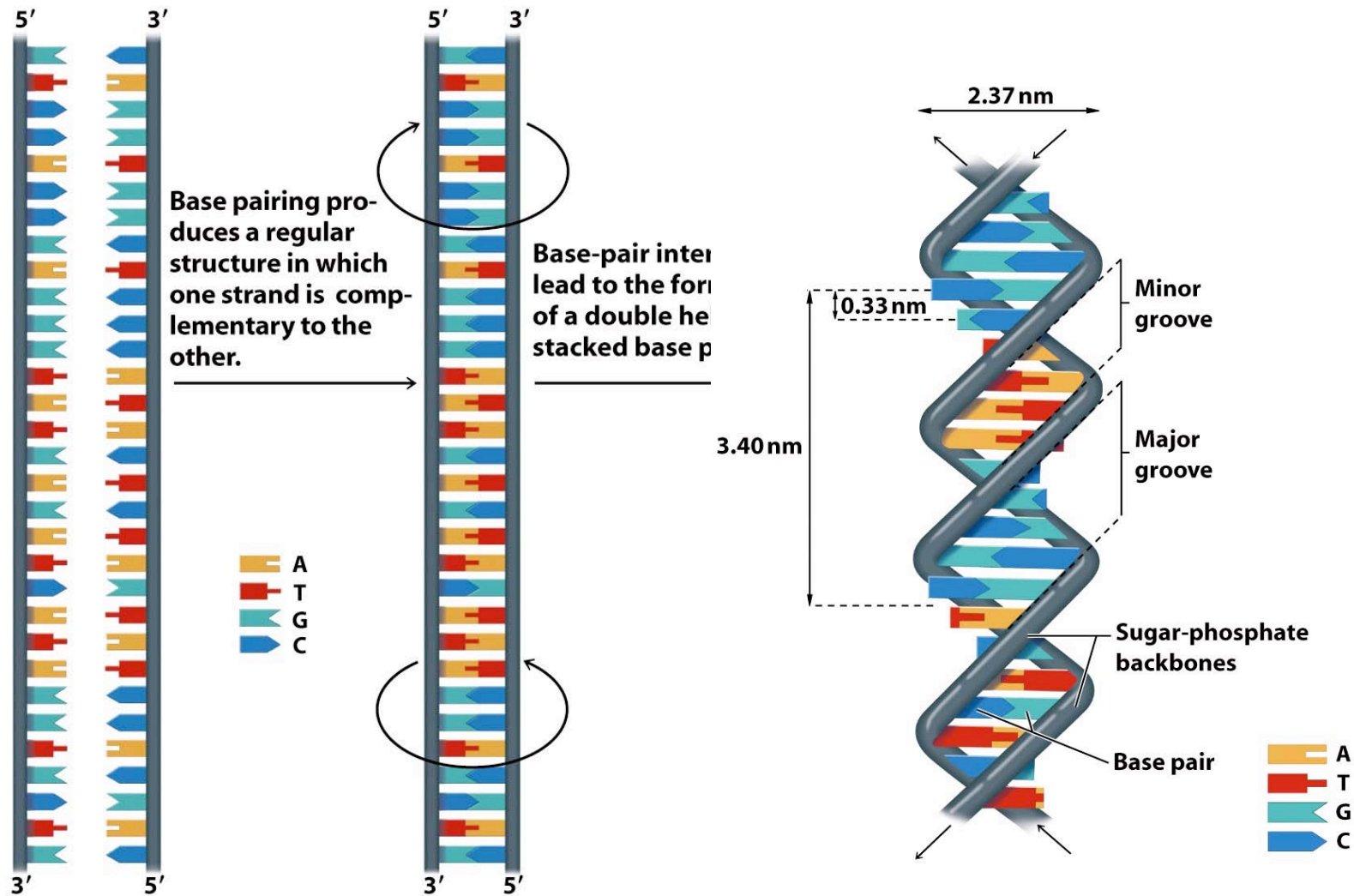
Three-dimensional structure of B-DNA. The sugar–phosphate backbone winds around the outside of the helix, and the bases occupy the interior. Stacking of the base pairs creates two grooves of unequal width, the major and the minor grooves.

In DNA the two strands are wound around each other, joined by **base-pairing** between each strand. The key feature of DNA is that each base can *only* be paired with its "complementary" base: A with G, C with T. Note that the bases are joined by hydrogen bonds (discussed earlier). Base-pairing is the key to replication in DNA.

Note that if it you could straighten out DNA, it would be about a mm (bacteria) to a cm (e.g. vertebrates) long! It might be impossible to find a compartment (cell membrane) large enough to house such a large information-carrying molecule. Instead, the double-helix coiling of the DNA allows it to be contained in a regions smaller than a *micro*meter. If DNA wasn't so coiled, there might never have been genetic material capable of carrying so much information.

**Again we see that part of the remarkable properties of complex biomolecules is their ability to attain varied and/or important shapes** (compare with proteins and lipids earlier).

(a)

5′ end
3′ end
HO

C   G
A   T
C   G

The DNA backbone is deoxyribose sugars...

...linked by phosphates.

G   C
T   A
C   G
T   A

|← 2 nm →| 3.4 nm

A   T
G   C
T   A
C   G

(b)

G   C
T   A   0.34 nm
C   G
T   A

Oxygen
Phosphorus
Hydrogen
Carbon in sugar–phosphate backbone

A   T
G   C

OH
3′ end

O
O—P=O
O⁻
5′ end

Bases

Minor groove
Major groove

Genetic code and replication process

translation dictionary = <u>genetic code</u>:

      a <u>codon</u> of 3 bases (out of 4) is a triplet code for specifying an amino acid,
e.g. in RNA (single strand):

```
   P---S---P---S----P----S
        |   |   |   |   |   |
        A   C   U  <-----a  codon
```
(Note: no. of possible codons = 4^3 = 64, which is greater than 20. Think about it!)

      A <u>gene</u> = sequence of codons long enough to specify a protein (~100-500 triplets long)

      In DNA, bases can't pair at random. Only A--T, G--C (base pairs).
When the 2 DNA strands unwind, each half can reproduce its partner exactly.

      <u>Messenger RNA</u> reads info (codons) from open DNA file.
Message taken to <u>ribosome</u> = assembly line for construction of proteins;
made of ~50 protein + RNA molecules. The needed amino acids are brought
to the ribosome by various <u>transfer RNAs</u>.

Can think of ``life" as a protein-making gene system.

Today's genetic code--Translates between sequence of letters to sequence of amino acids

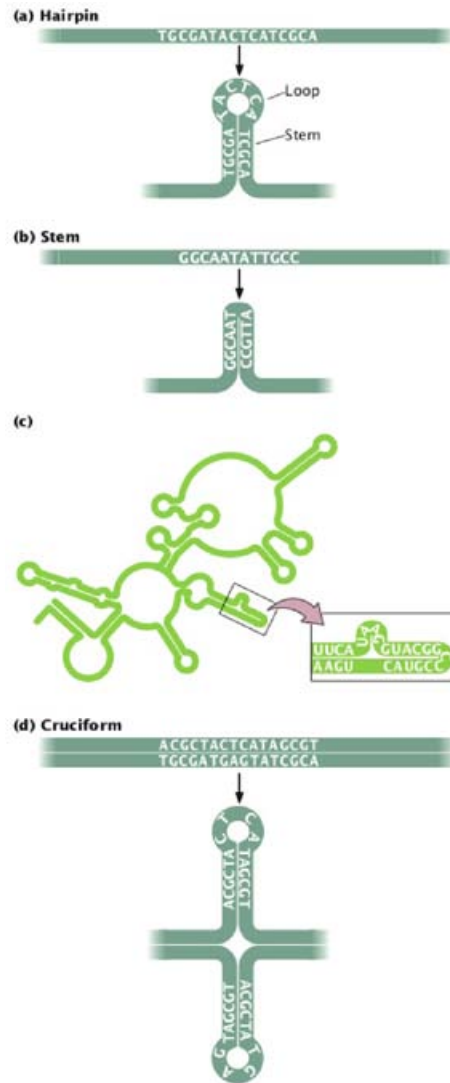|		Second Base of Codon				
		U	**C**	**A**	**G**	
First Base of Codon	**U**	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } UCA } Ser UCG }	UAU } Tyr UAC } UAA UAG	UGU } Cys UGC } UGA UGG Trp	U C A G
	C	CUU } CUC } CUA } Leu CUG }	CCU } CCC } CCA } Pro CCG }	CAU } His CAC } CAA } CAG } Gln	CGU } CGC } CGA } Arg CGG }	U C A G
	A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } ACA } Thr ACG }	AAU } Asn AAC } AAA } AAG } Lys	AGU } Ser AGC } AGA } AGG } Arg	U C A G
	G	GUU } GUC } GUA } Val GUG }	GCU } GCC } GCA } Ala GCG }	GAU } Asp GAC } GAA } GAG } Glu	GGU } GGC } GGA } Gly GGG }	U C A G

Third Base of Codon

Note: AUG is the start codon and UAA, UAG, and UGA are stop codons as highlighted in table.

DNA-protein system: Too complex for first life
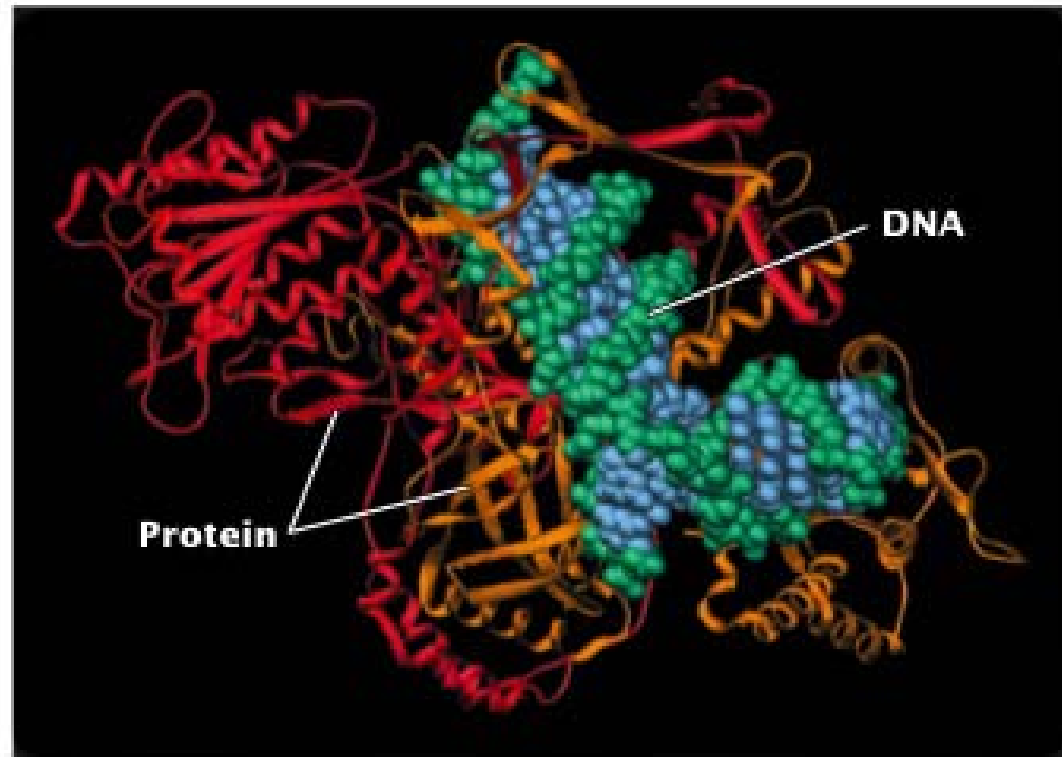
- The "chicken and the egg" problem is obvious: Neither DNA nor protein has any function without the other. Yet their symbiosis is far too complex to have arisen from "nothing."

➜ *So what preceded the DNA/protein system?*

RNA secondary structure

(a) Hairpin
TGCGATACTCATCGCA
— Loop
— Stem

(b) Stem
GGCAATATTGCC

(c)
UUCA GUACGG
AAGU CAUGCC

(d) Cruciform
ACGCTACTCATAGCGT
TGCGATGAGTATCGCA

- •　　Both DNA and RNA can form special secondary structures.
- • (a) A hairpin, consisting of a region of paired bases (which forms the stem) and a region of unpaired bases between the complementary sequences (which form a loop at the end of the stem).
- • (b) A stem with no loop.
- • (c) Secondary structure of RNA component of RNase *p* of *E. coli*. RNA molecules often have complex secondary structures.
- • (d) A cruciform structure.

These examples illustrate the potential for RNA to carry out the functions of both DNA (sequence bases) and protein (folding into characteristic structure according to sequence it carries)

When RNA molecules with this dual-function property were discovered (*ribozymes, autocatalytic RNA*), it became plausible, even likely that there was an "RNA world" before the present "DNA-protein world."

But what preceeded RNA?

Here are some suggested predecessors. They may be easier to produce, but this hardly solves the problem, since there is no evidence that these 'ancestors' could be dual-function like ribozymes.
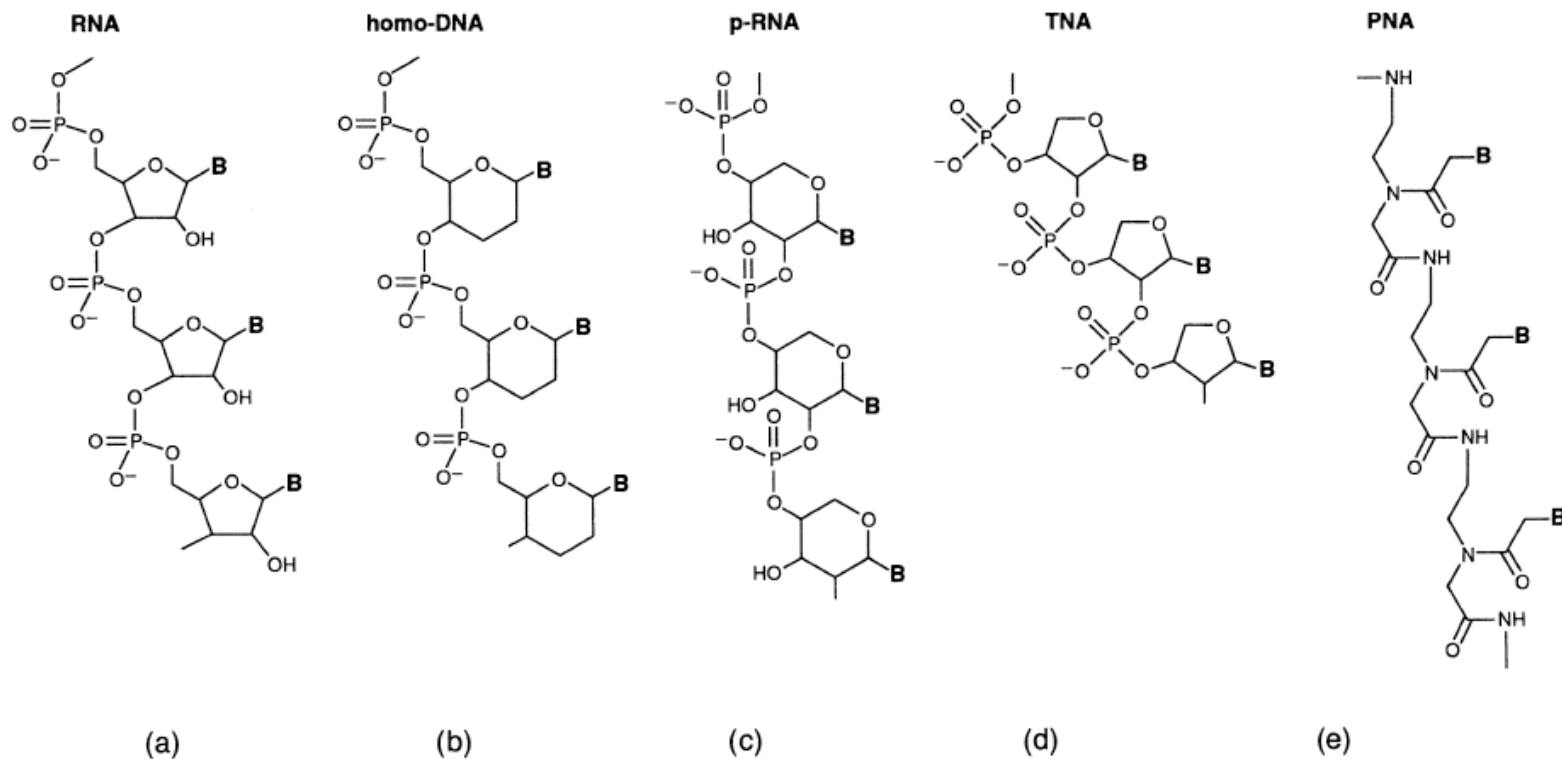

FIG. 9.   The structures of single strands of (a) RNA, (b) homo-DNA, (c) p-RNA, (d) TNA, and (e) PNA.

"Naked gene" or "random replicator" theories
(RNA world is an example)

1st need nucleosides (= base+ sugar)

　　　→heat sugar + bases + salts → <u>suggests drying tidepools</u>

　　<u>But</u> then must polymerize the nucleosides. Tough!

Spiegelmann: Qβ virus + enzyme + free nucleiotides→ "Spiegelman
　　　　　　　　　　　　　　　　　　　　　　　　　monster"

　　　(w/long RNA)

Eigen: enzyme + free nucleotides + salts → short RNA random replicator.

But both S. & E. started with proteins.

Orgel: RNA can form a double helix without any protein. But then stopped.

Cech *et al.*: <u>self-catalytic RNA</u>--RNA can cut up different RNAs, acting as an enzyme.
It can also join short RNAs into longer chains.
(Extremely influential result; gave rise to term "RNA World")

　　　1994: Joyce *et al.* made synthetic RNA that can copy itself (given the right
proteins).

　　　1997: Two studies in Jan.21 Proc.Nat.Acad.Sci. claim experimental evidence
related to enzymes that convert between RNA and DNA.

　　　2001 RNA shown to catalyze its own replication without enzymes.