Introduction to Bioinformatics:

Basic Concepts in Molecular Biology

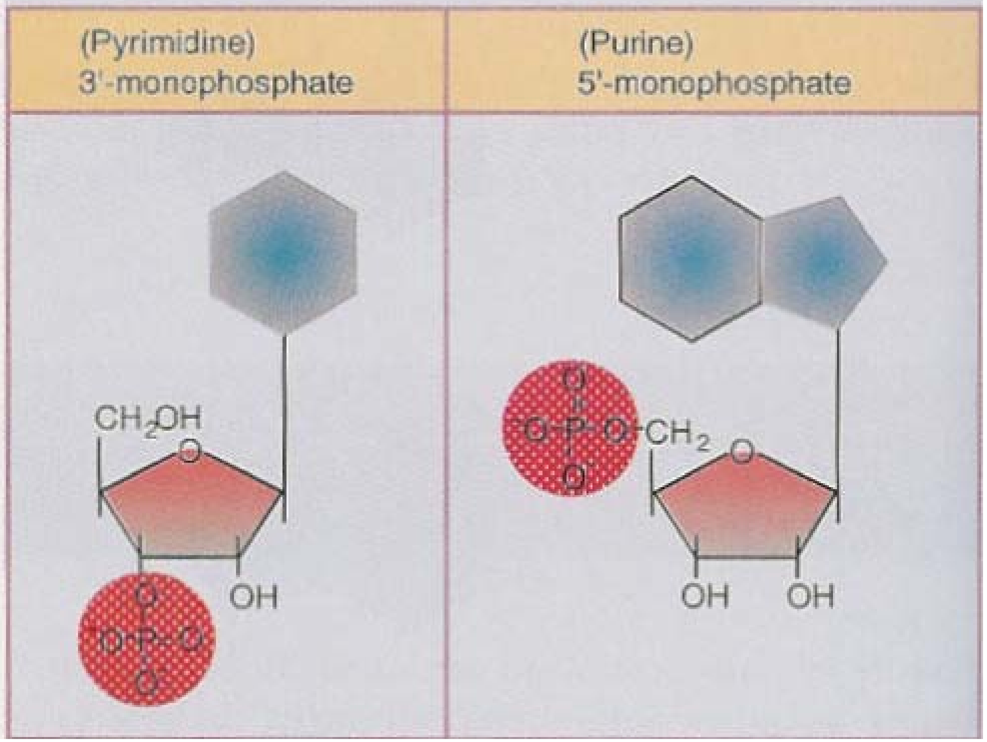
Peter K. Rogan, Ph.D.
Laboratory of Human Molecular Genetics
Children's Mercy Hospital
Schools of Medicine & Computer Science and Engineering
University of Missouri- Kansas City

http://www.sce.umkc.edu/~roganp

Nucleic Acids

- Monomers are (deoxy)ribonucleotides
- Linked by 3’ – 5’ phosphate backbone
- Chargarf’s rule f(A)=f(T) or f(U),f(G)=f(C)
- Base paring stabilized by stacking
- Antiparallel helices
- Replication is template directed

Figure 4.8 Nucleotides may carry phosphate in the 5' or 3' position.


(Pyrimidine)
3'-monophosphate

(Purine)
5'-monophosphate

Figure 4.4 Pyrimidines have a six carbon ring.

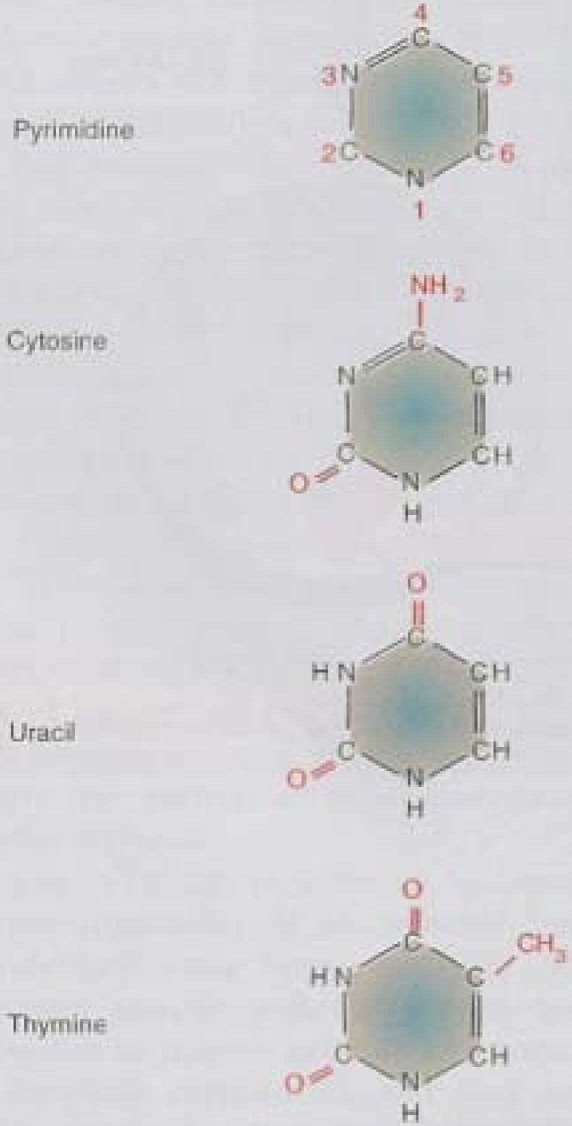
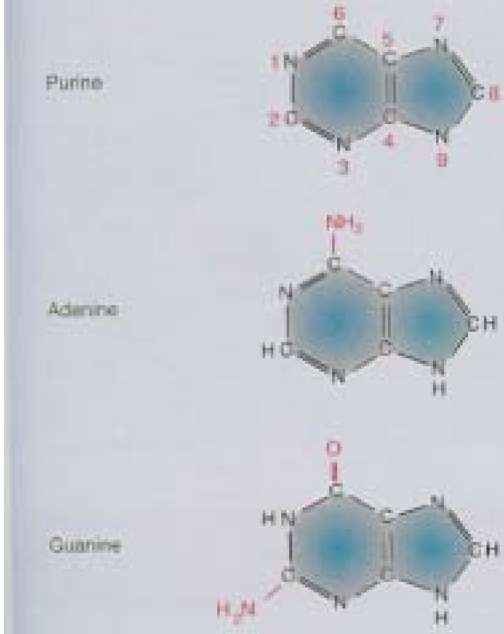
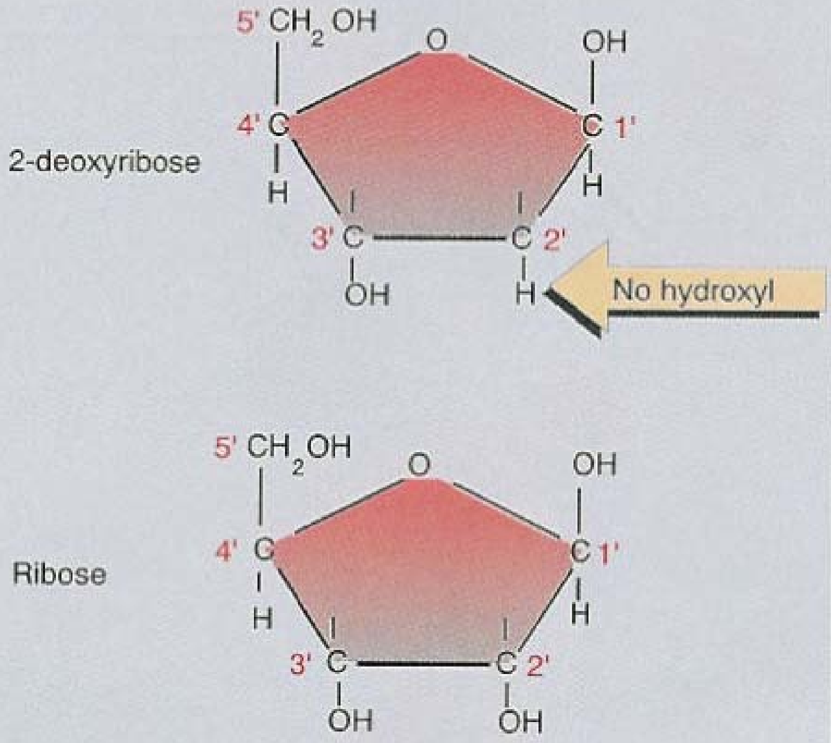


Figure 4.5 Purines consist of two joined carbon rings, with five and six members.



Figures abstracted from <u>Genes VI</u>, B. Lewin (Oxford University Press)

Figure 4.6 2-Deoxyribose is the sugar in DNA and ribose is the sugar in RNA. The carbon atoms are numbered as indicated for deoxyribose. The sugar is connected to the nitrogenous base via position 1'.



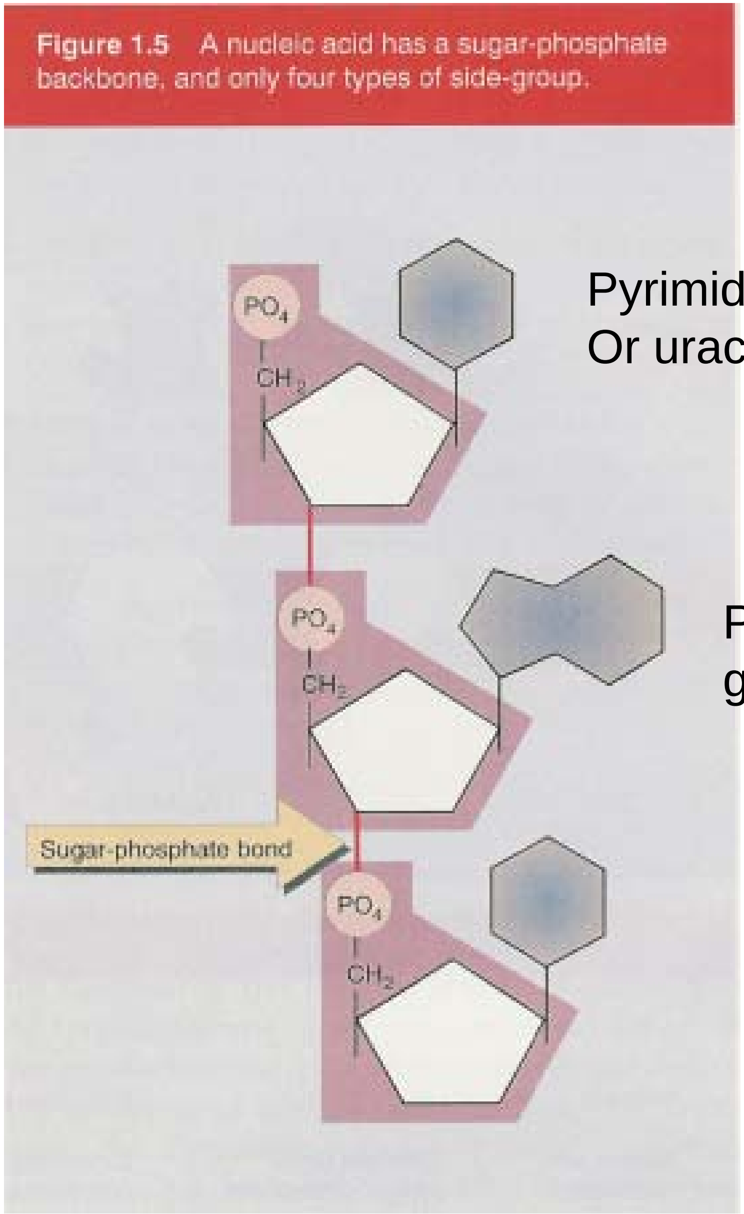


Figure 1.5 A nucleic acid has a sugar-phosphate backbone, and only four types of side-group.

Pyrimidine (cytosine, thymine
Or uracil [RNA])

Purine (usually
guanine or adenine)

Figure 4.7 A polynucleotide chain consists of a series of 5'-3' sugar-phosphate links that form a backbone from which the bases protrude.

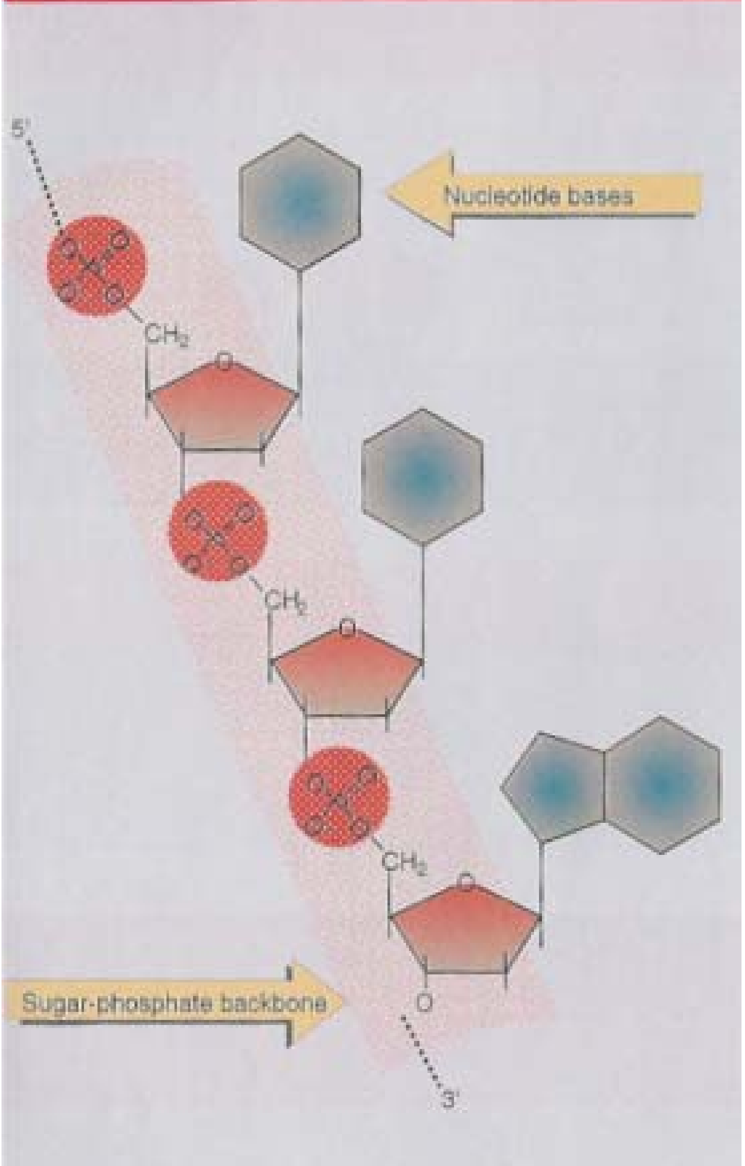
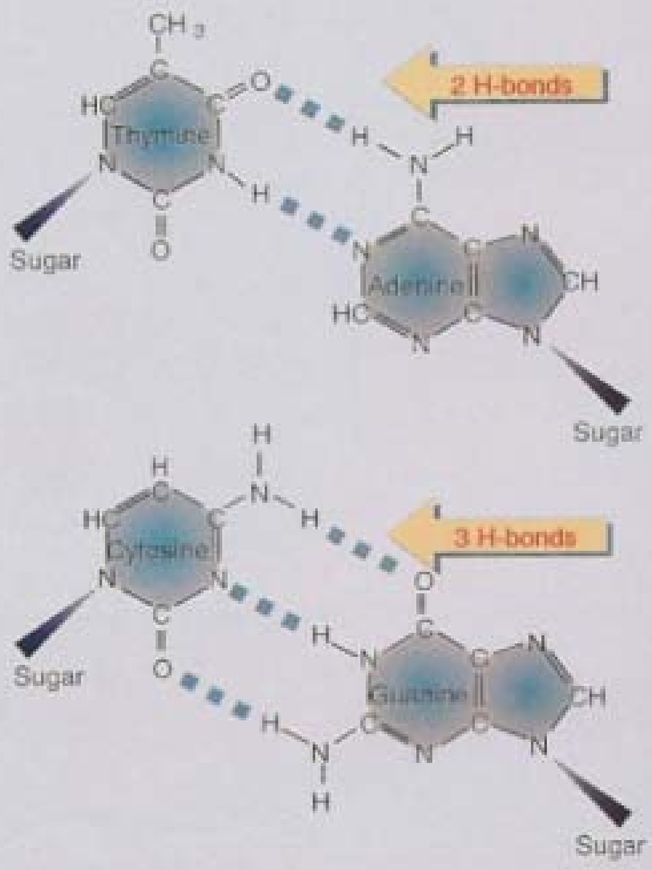


Figure 4.11 Complementary base pairing involves the formation of two hydrogen bonds between A and T, and of three hydrogen bonds between G and C. No other pairs form in DNA.



B- DNA has a right handed twist

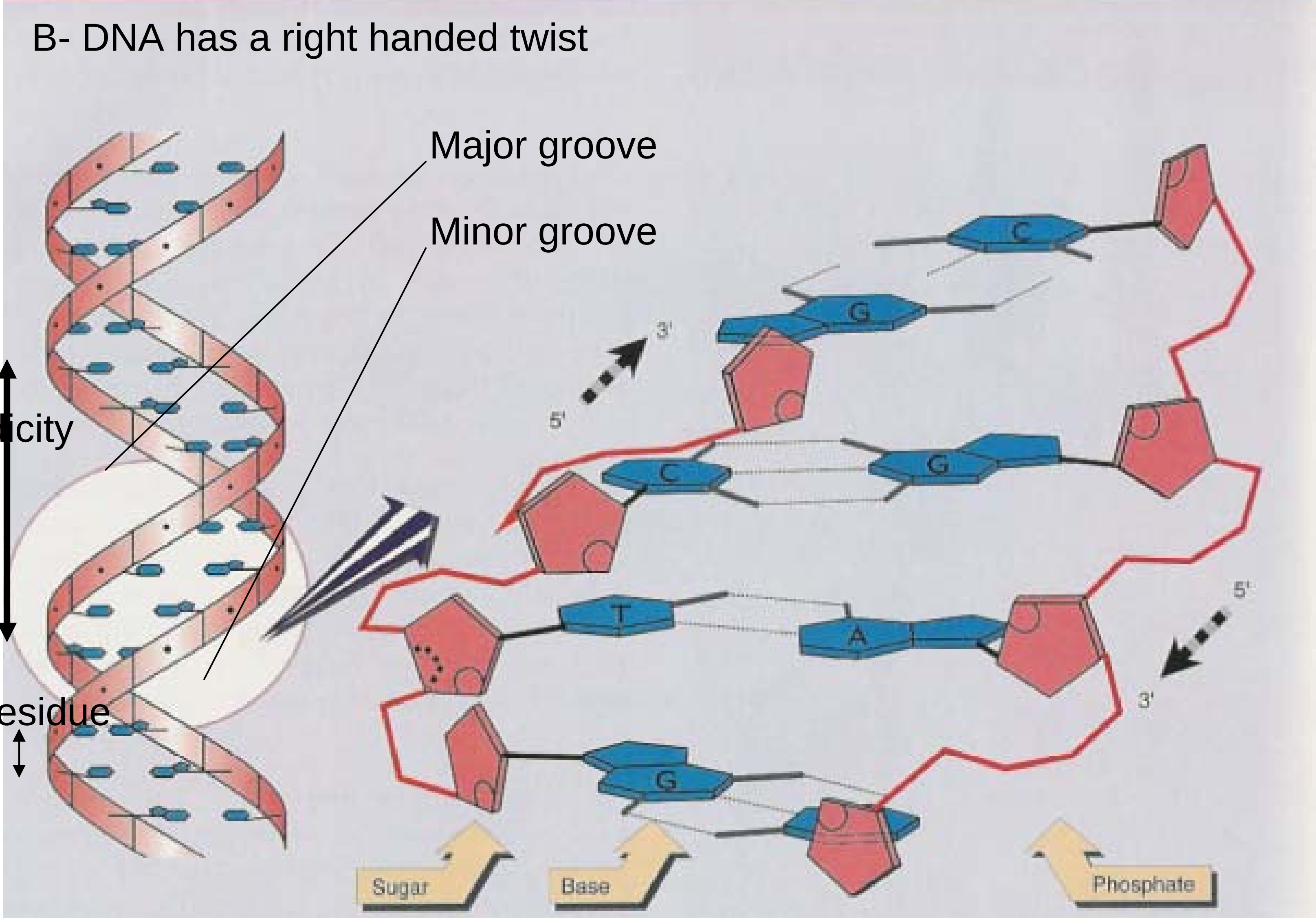


Figure 4.13 Flat base pairs lie perpendicular to the sugar-phosphate backbone.

Major groove

Minor groove

Periodicity
10 bp

Rise/residue
1.5 Å

Figure 4.12 The double helix maintains a constant width because purines always face pyrimidines in the complementary A-T and G-C base pairs.

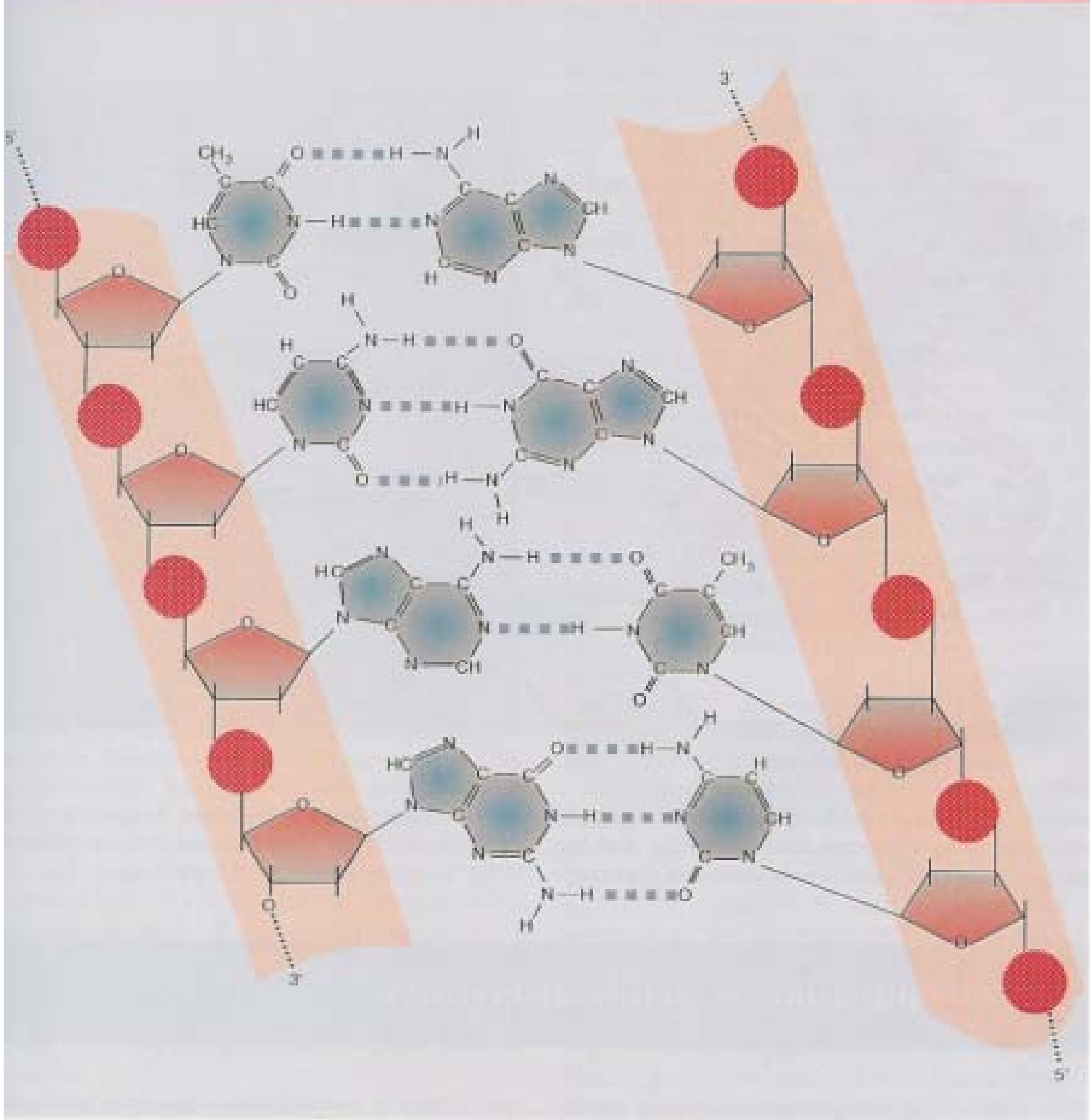


Figure 4.9 A nucleoside-5'-triphosphate has energy-rich phosphate bonds.

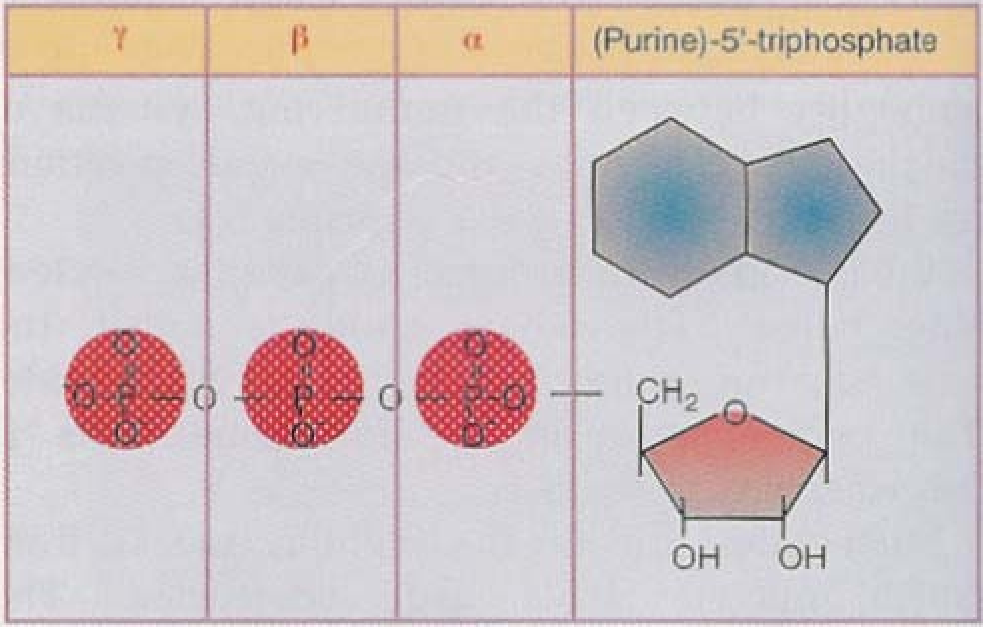
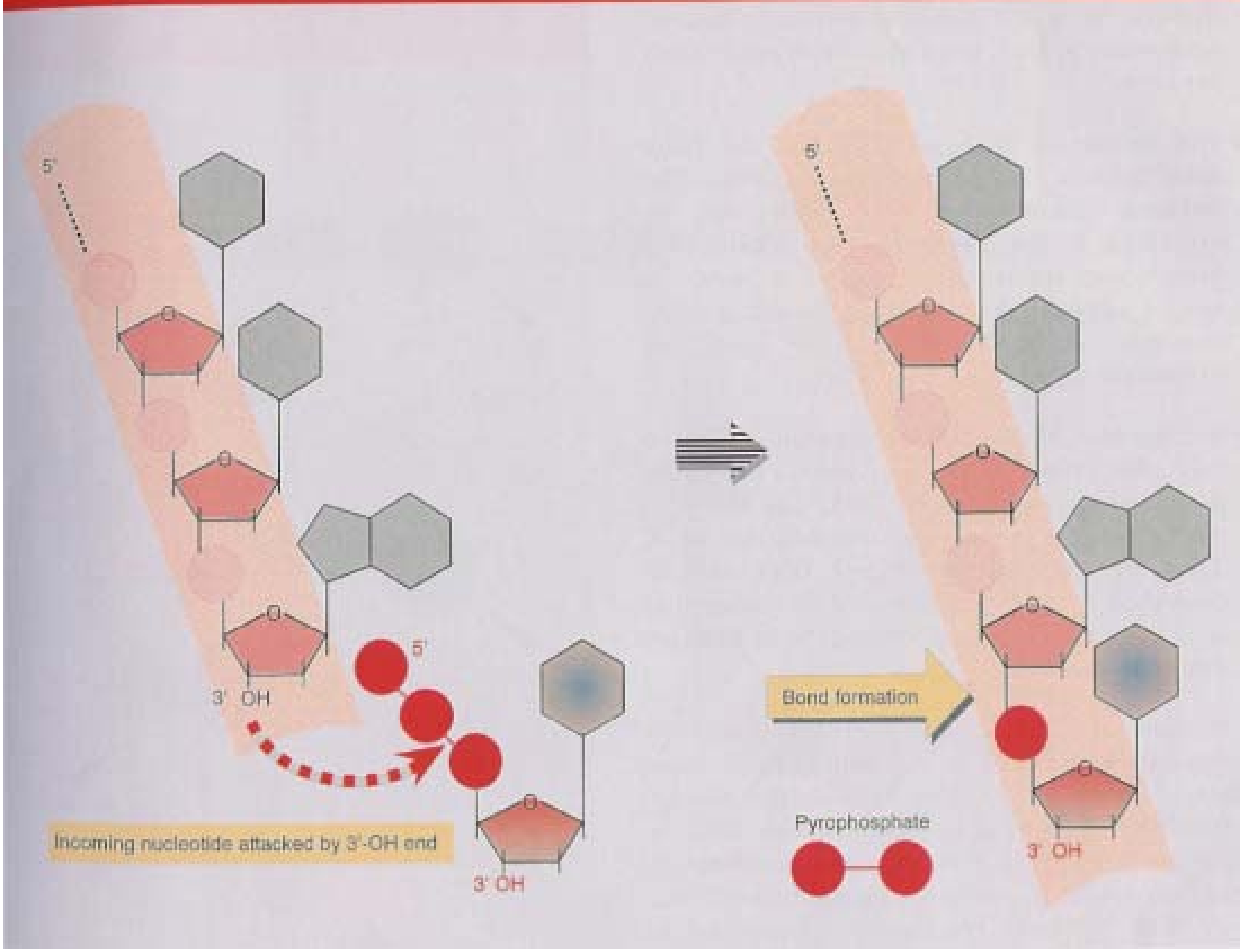
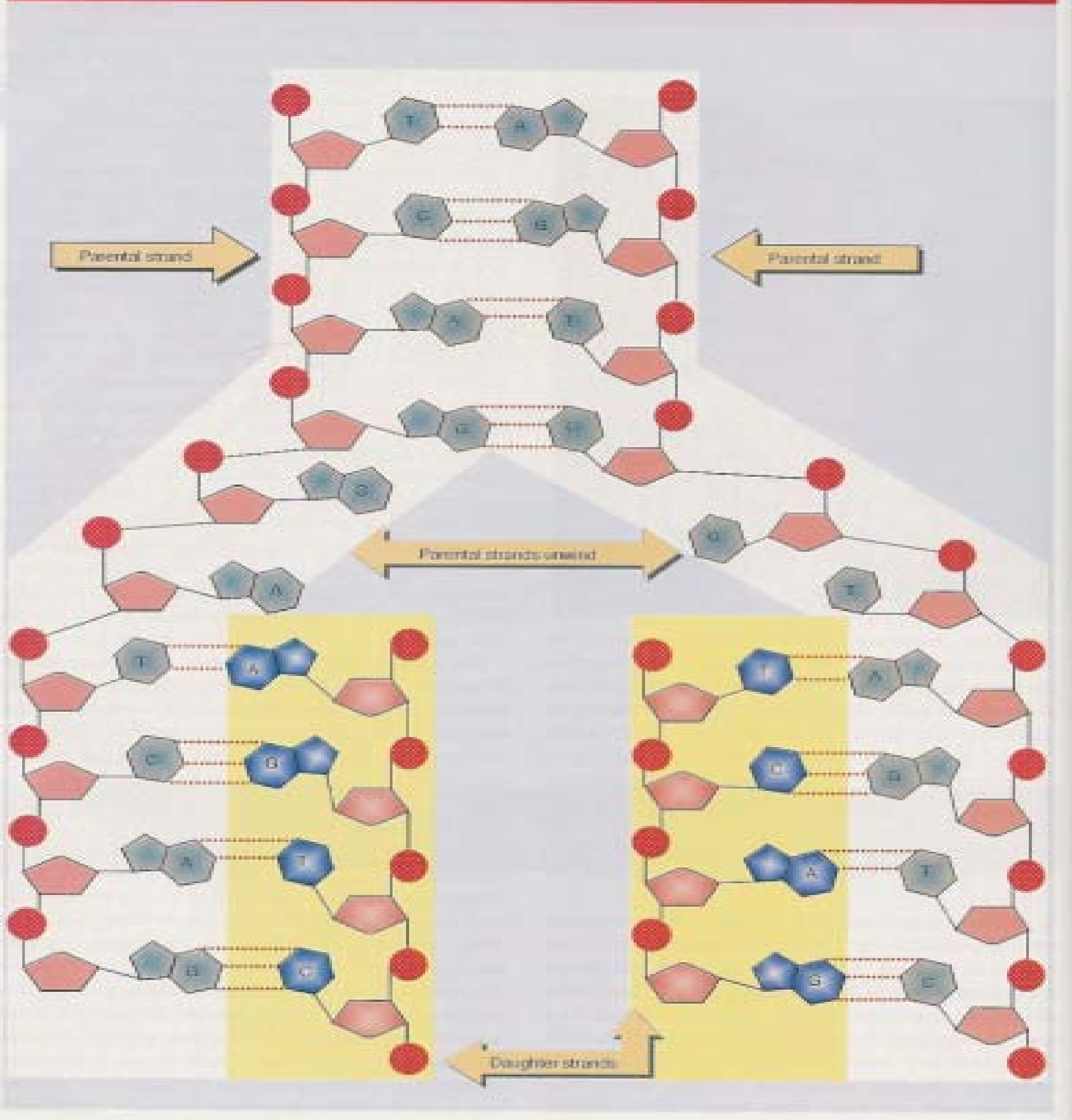


Figure 4.10 Nucleic acid synthesis occurs by adding the nucleoside-5'- monophosphate moiety of a nucleoside triphosphate to the 3'-OH end of the polynucleotide chain.



DNA Replication (carried out by DNA polymerase enzyme)

Figure 4.15 Base pairing provides the mechanism for faithfully replicating DNA.



Proteins

- Monomers are amino acids (20 naturally occurring varieties)
- Amino acids are linked by peptide bonds: primary sequences
- Primary sequences self assemble into secondary structures (alpha helices, beta sheets
- Secondary structures fold into tertiary structures (may be stabilized by covalent or non-covalent side chain interactions)
- Proteins may multimerize or associate with other heterologous proteins: quarternary structures

Figure 1.8 Amino acids are classified according to the nature of their side-groups. Each amino acid may be described by either a three letter abbreviation or a one letter code.

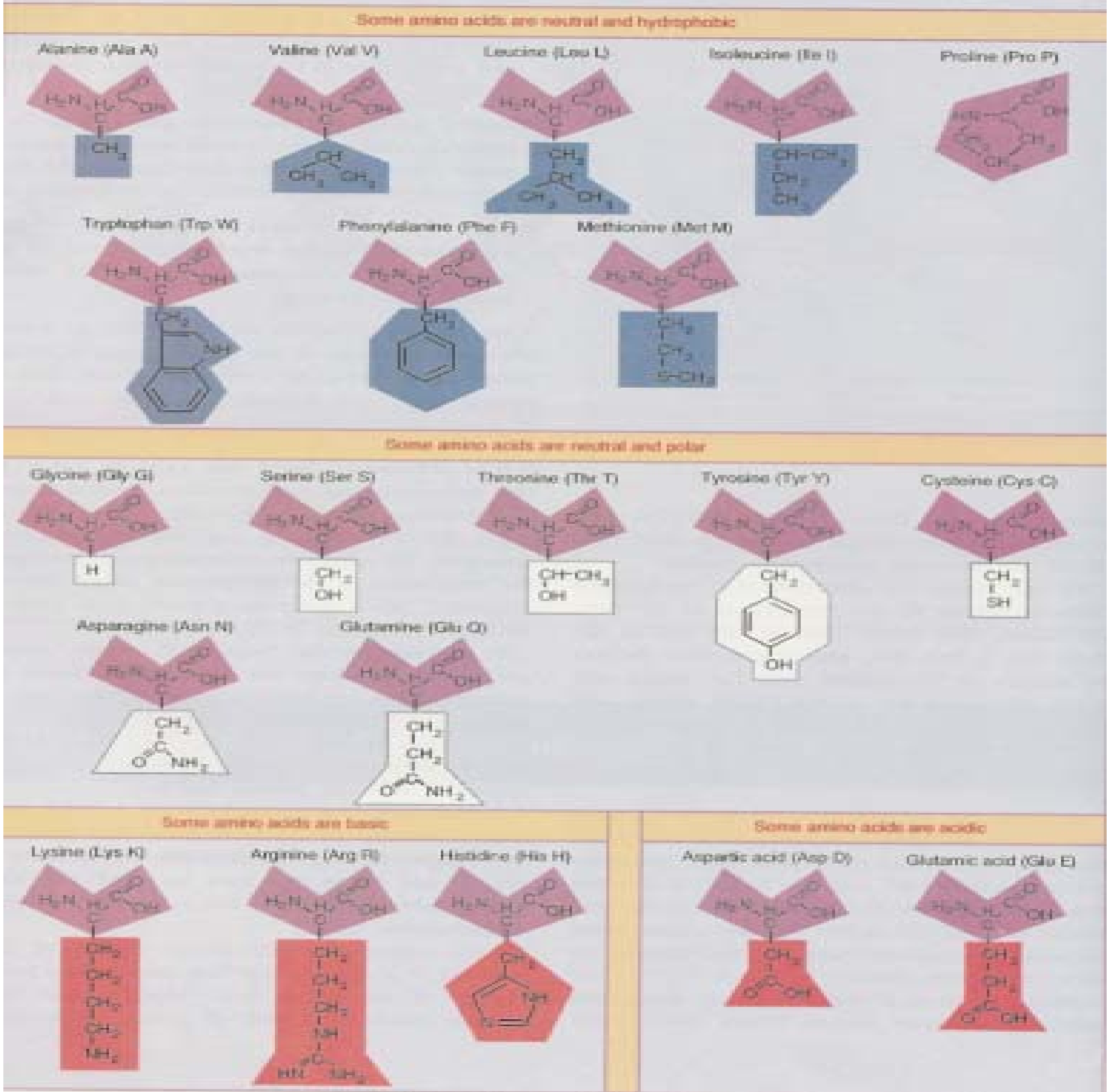


Figure 1.7 A peptide bond is formed by a condensation reaction in which a water molecule is lost.

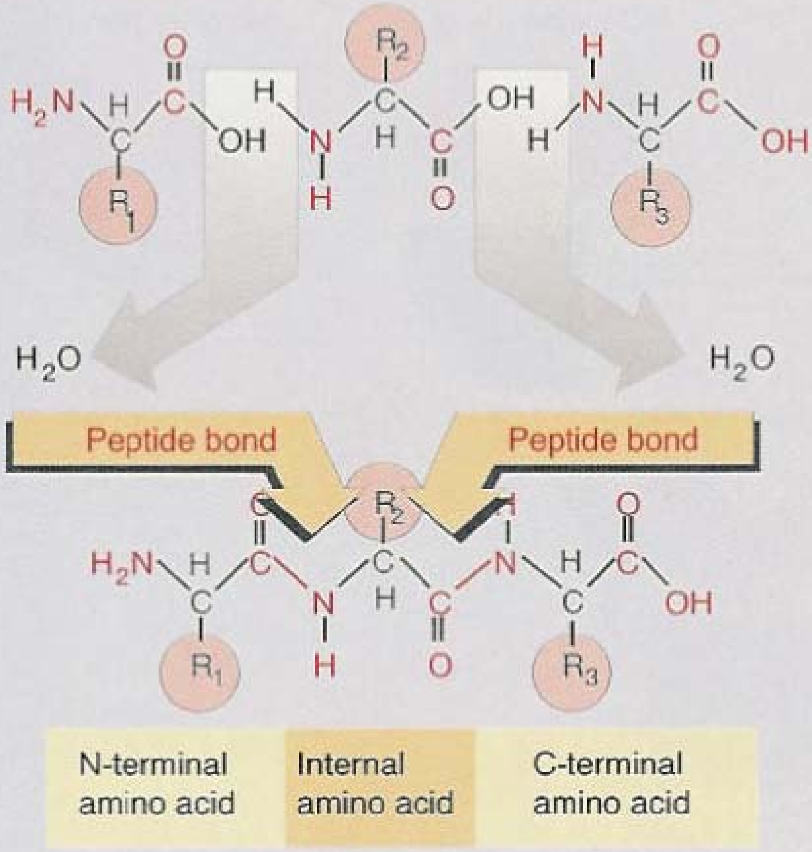


Figure 1.16 Basic and acidic amino acids may be attracted via ionic bonds, shielded by water in the aqueous environment of a protein.

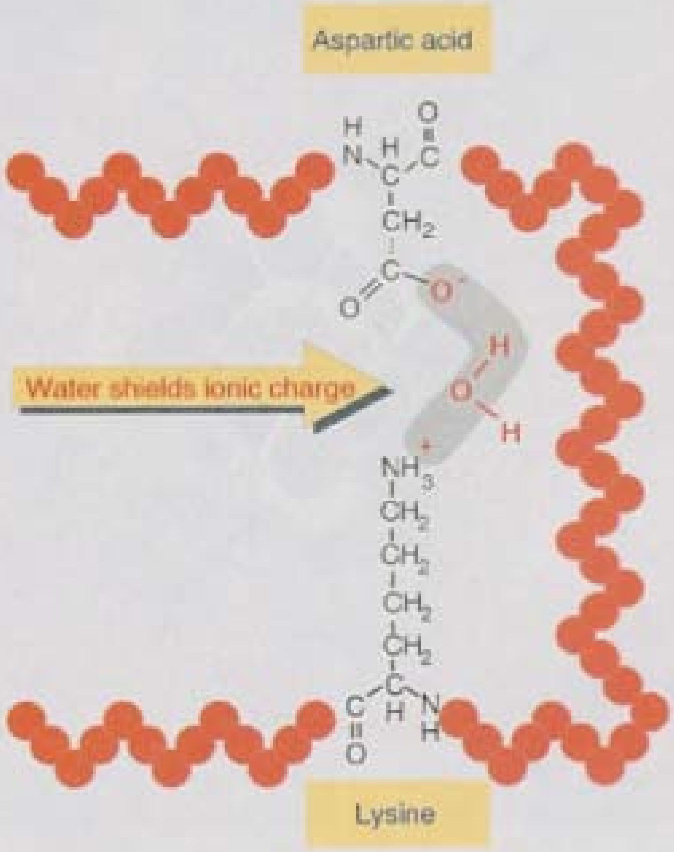


Figure 1.14 Hydrogen bonds may form between the side-groups of polar amino acids.

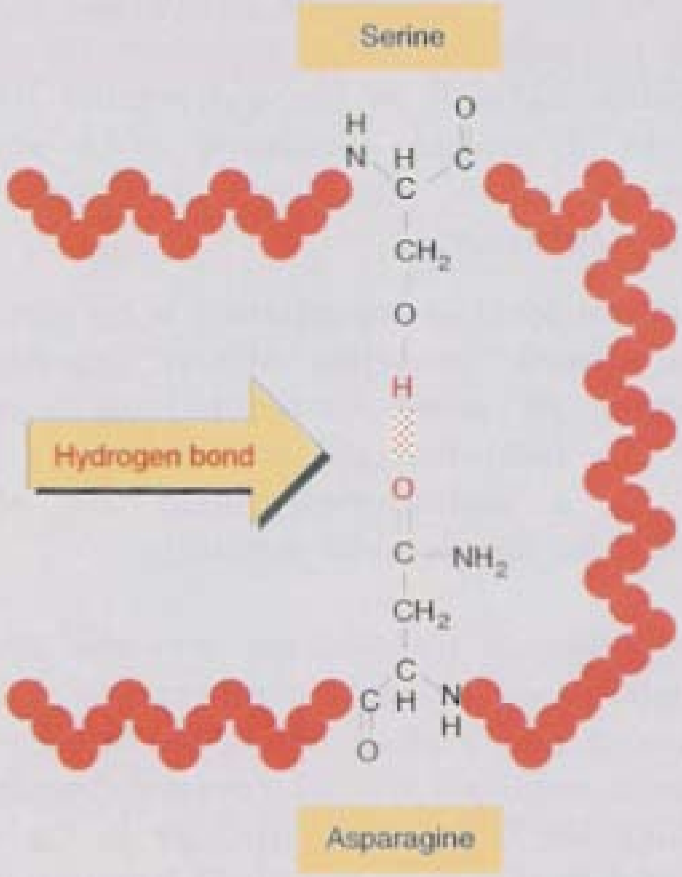


Figure 2.5 A transmembrane protein crosses the lipid bilayer. The hydrophobic transmembrane region resides within the lipid bilayer, and hydrophilic regions are exposed on either side of the membrane.

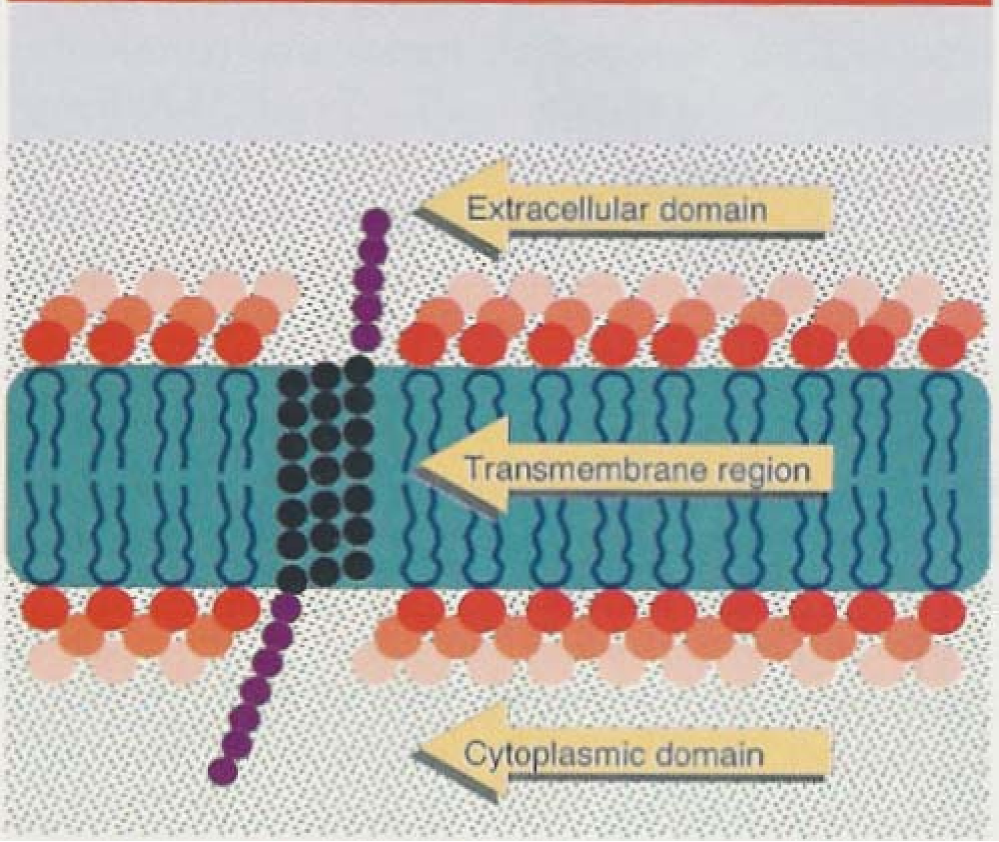
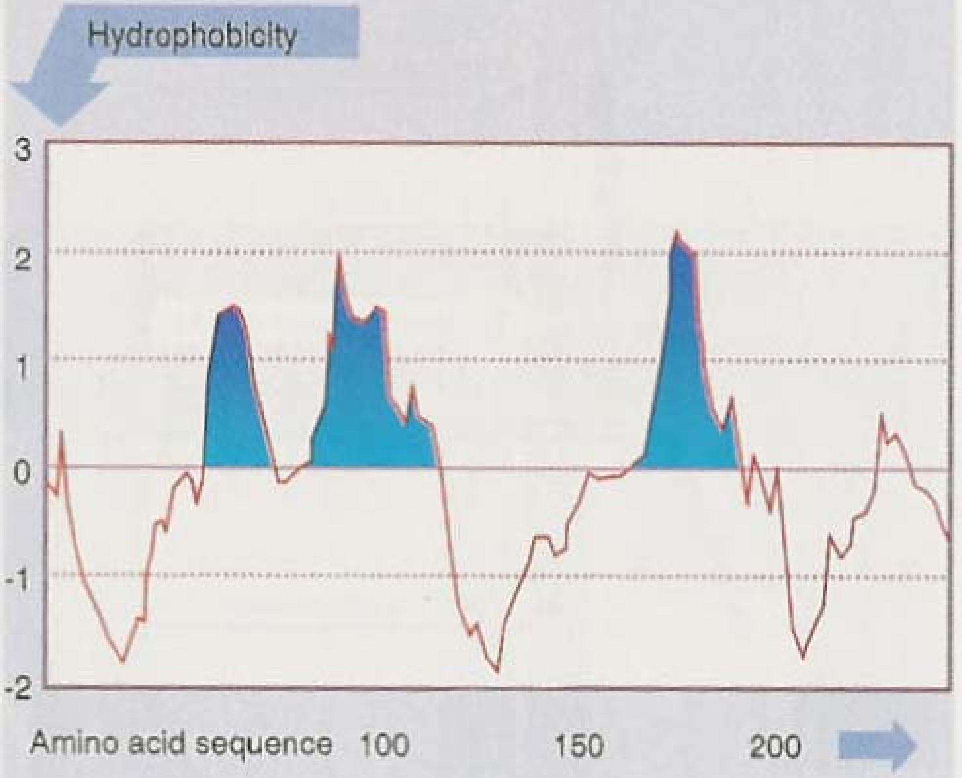


Figure 2.6 A hydropathy plot identifies potential membrane-spanning regions as the most hydrophobic sequences of the protein. Each potential membrane segment is identified by shading in color.



Watson, Crick, Brenner, etc:

The Central Dogma of Molecular Biology

Genome (usually)

Messenger RNA

Protein

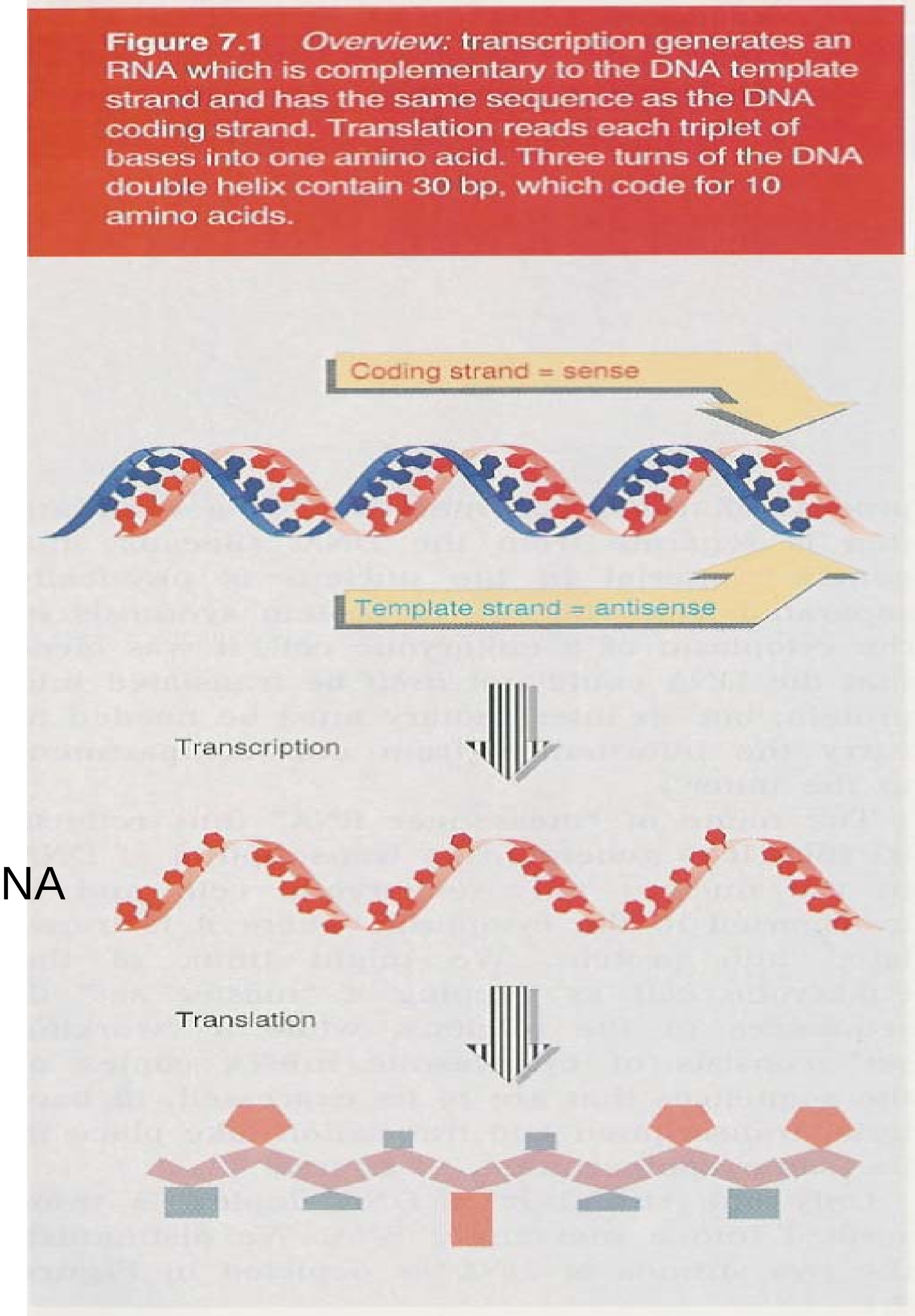


Figure 7.1 *Overview:* transcription generates an RNA which is complementary to the DNA template strand and has the same sequence as the DNA coding strand. Translation reads each triplet of bases into one amino acid. Three turns of the DNA double helix contain 30 bp, which code for 10 amino acids.

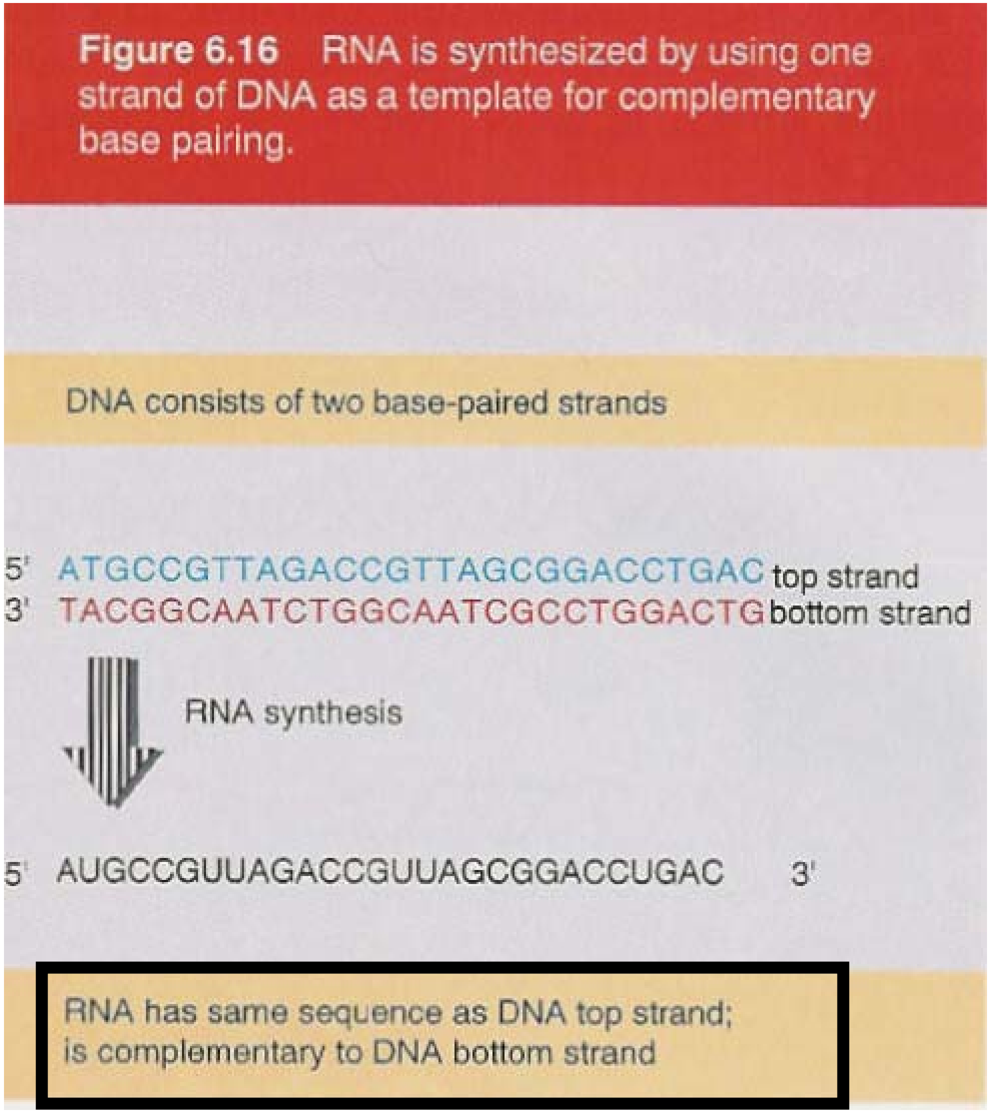


Figure 6.16 RNA is synthesized by using one strand of DNA as a template for complementary base pairing.

DNA consists of two base-paired strands

5' ATGCCGTTAGACCGTTAGCGGACCTGAC top strand
3' TACGGCAATCTGGCAATCGCCTGGACTG bottom strand

RNA synthesis

5' AUGCCGUUAGACCGUUAGCGGACCUGAC 3'

RNA has same sequence as DNA top strand;
is complementary to DNA bottom strand

Contradictions to the dogma:

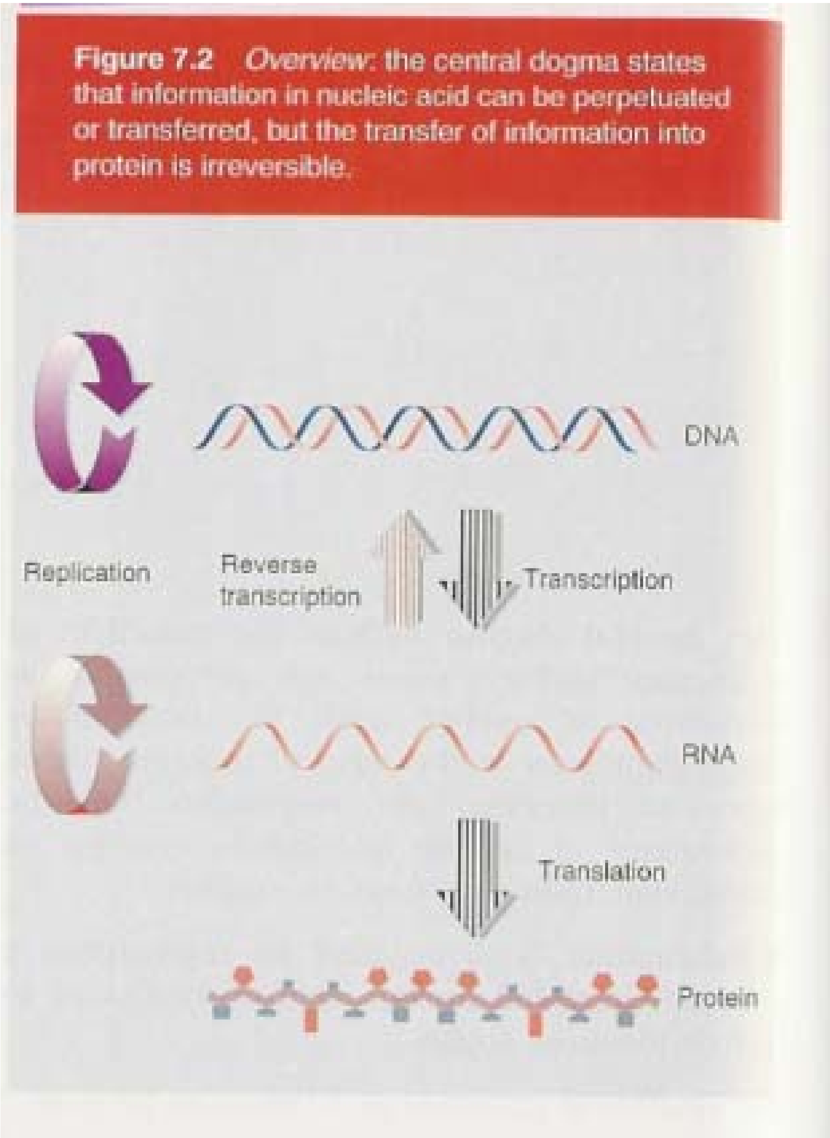


Figure 7.2 *Overview:* the central dogma states that information in nucleic acid can be perpetuated or transferred, but the transfer of information into protein is irreversible.

Formed and propagated by this mechanism:

- Retroviruses
- Processed Pseudogenes
- Retrotransposons

Information transfer from RNA to protein

- Genetic code is comprised of non-overlapping triplets of nucleotides [codons]
- The code is redundant – more than one codon per amino acid
- T(ransfer)RNA is the intermediate molecule that carries both nucleic acid codon and amino acid.
- In ribosome, tRNA anticodon is read and amino acid is delivered to growing peptide

Figure 7.3 A tRNA has the dual properties of an adaptor that recognizes both the amino acid and codon. The 3' adenosine is covalently linked to an amino acid. The anticodon base pairs with the codon on mRNA.

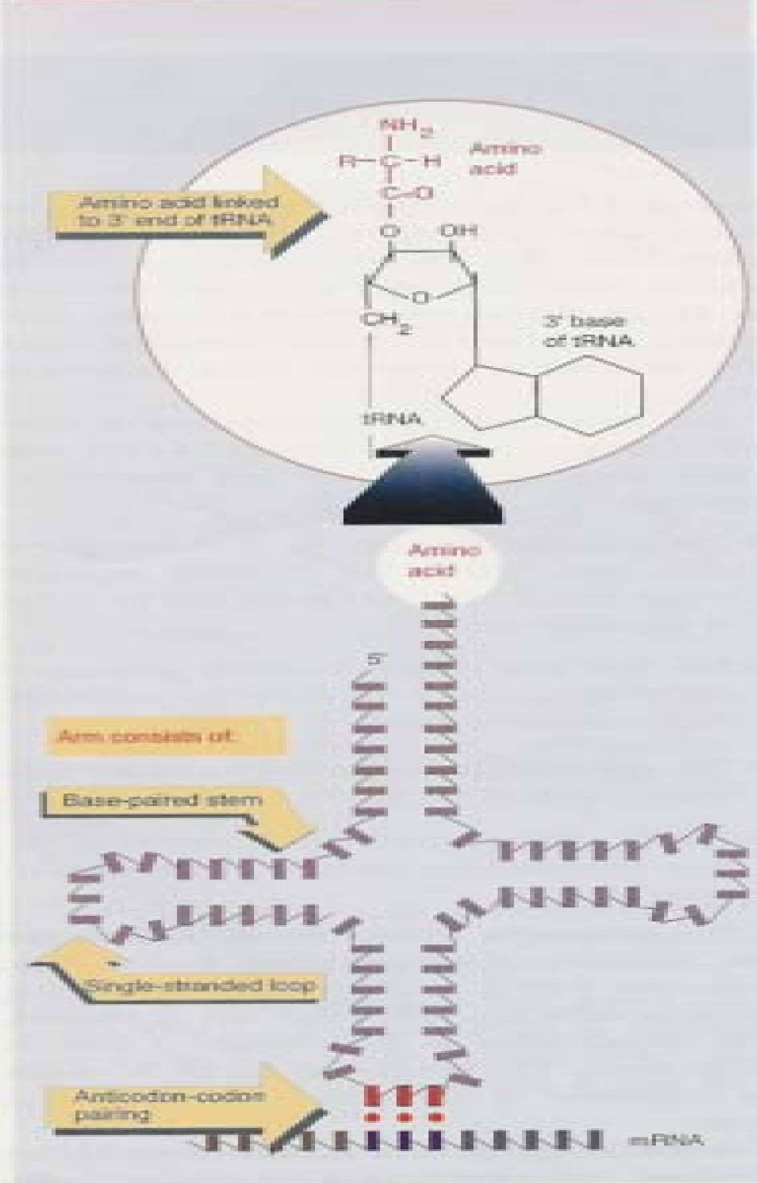
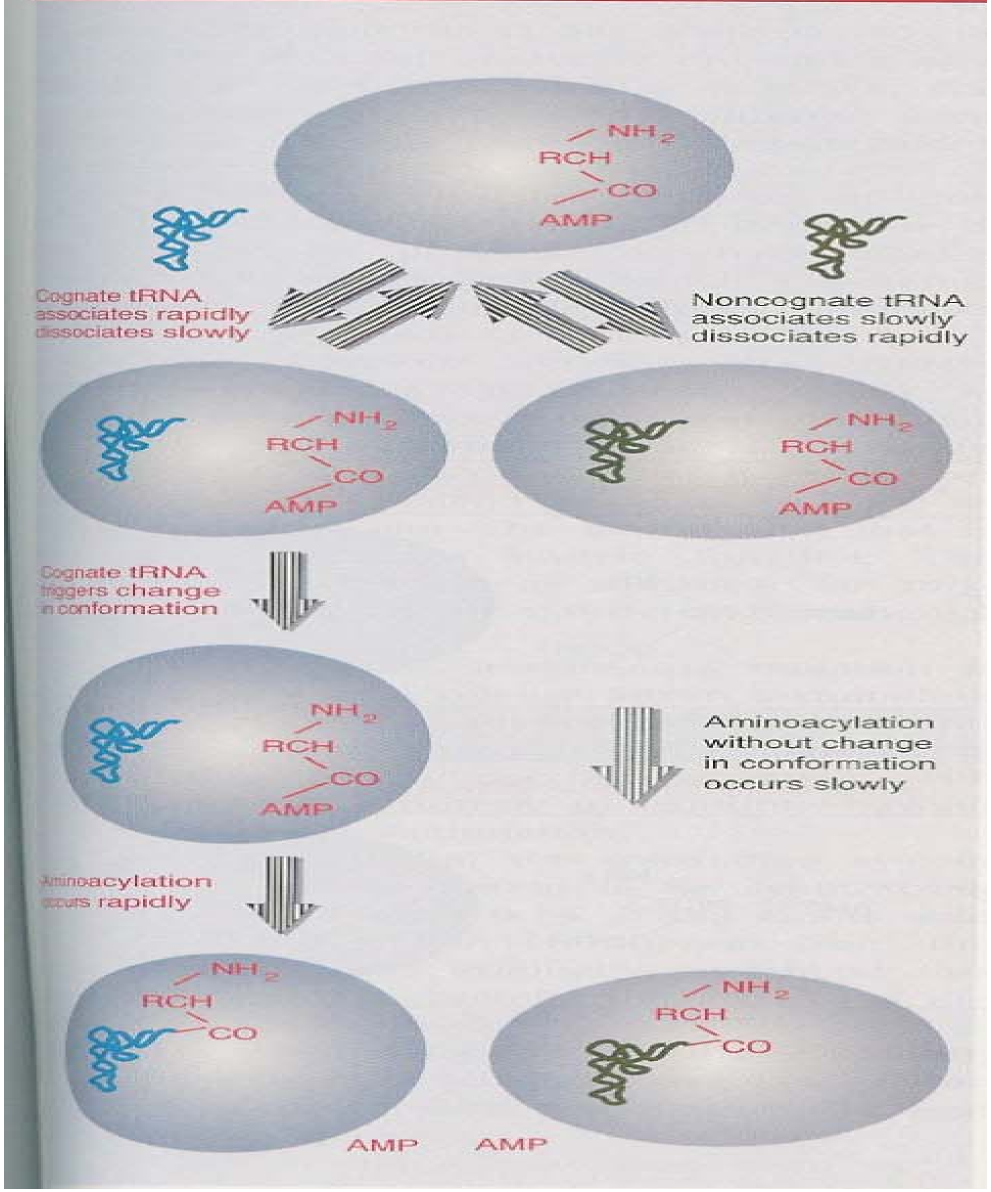


Figure 9.12 Recognition of the correct tRNA by synthetase is controlled at two steps. *First,* the enzyme has a greater affinity for its cognate tRNA. *Second,* the aminoacylation of the incorrect tRNA is very slow.



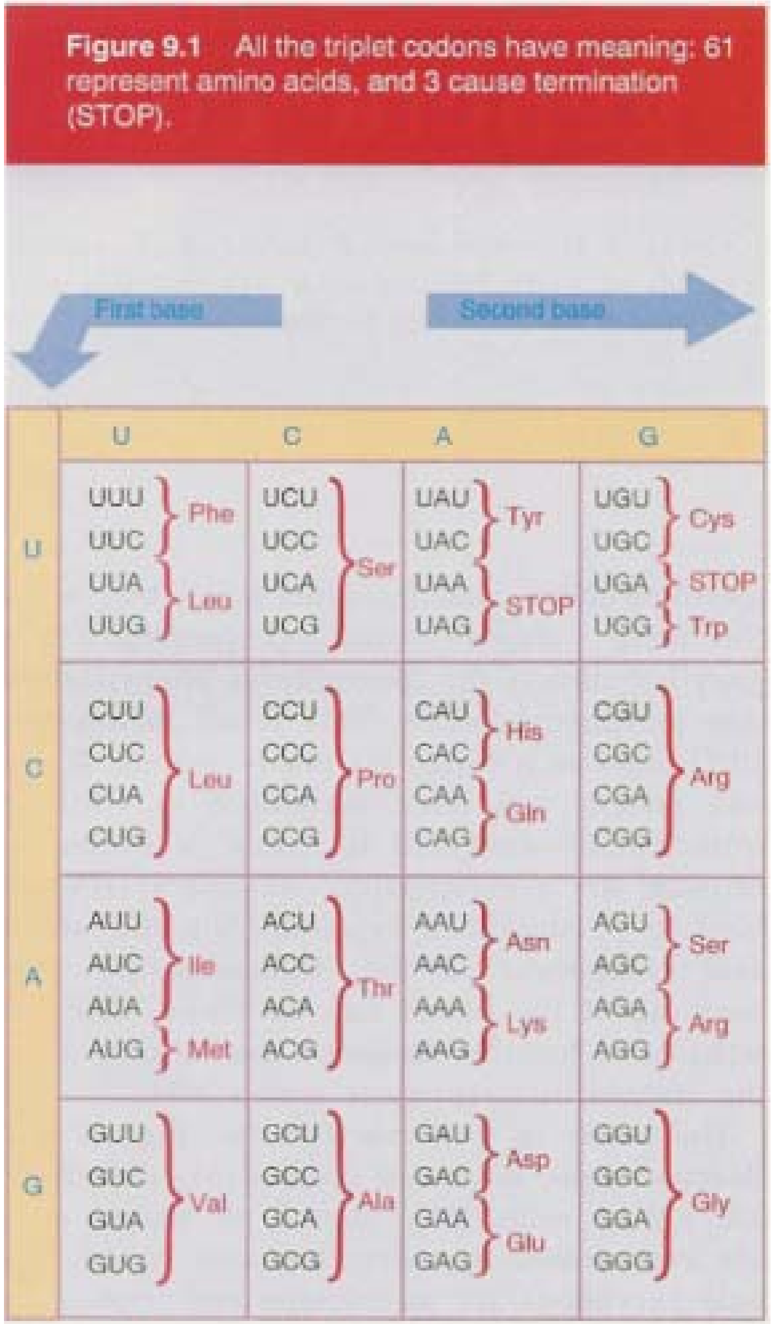


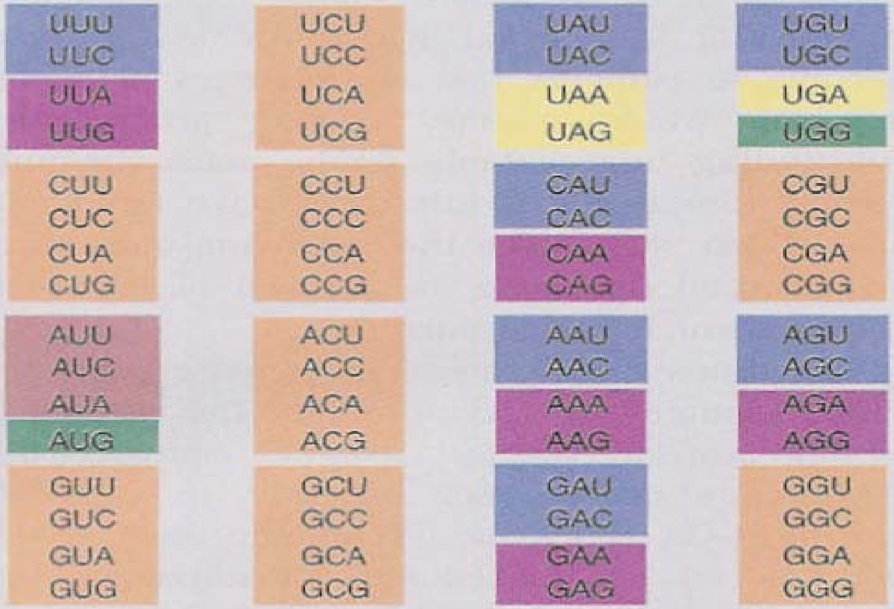
Figure 9.1 All the triplet codons have meaning: 61 represent amino acids, and 3 cause termination (STOP).

First base

Second base

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

Figure 9.3 Third bases have the least influence on codon meanings. Boxes indicate groups of codons within which third-base degeneracy ensures that the meaning is the same.



UUU	UCU	UAU	UGU
UUC	UCC	UAC	UGC
UUA	UCA	UAA	UGA
UUG	UCG	UAG	UGG
CUU	CCU	CAU	CGU
CUC	CCC	CAC	CGC
CUA	CCA	CAA	CGA
CUG	CCG	CAG	CGG
AUU	ACU	AAU	AGU
AUC	ACC	AAC	AGC
AUA	ACA	AAA	AGA
AUG	ACG	AAG	AGG
GUU	GCU	GAU	GGU
GUC	GCC	GAC	GGC
GUA	GCA	GAA	GGA
GUG	GCG	GAG	GGG

Third base relationship	Third bases with same meaning	Number of codons
🟧 third base irrelevant	U, C, A, G	32
🟦 purines differ from pyrimidines	U or C	14
🟪	A or G	10
▨ unique definitions	U, C, A	3
🟩	G only	2

Figure 7.16 Initiation may occur independently at each cistron in a polycistronic mRNA. When the intercistronic region is longer than the span of the ribosome, dissociation at the termination site is followed by independent reinitiation at the next cistron.

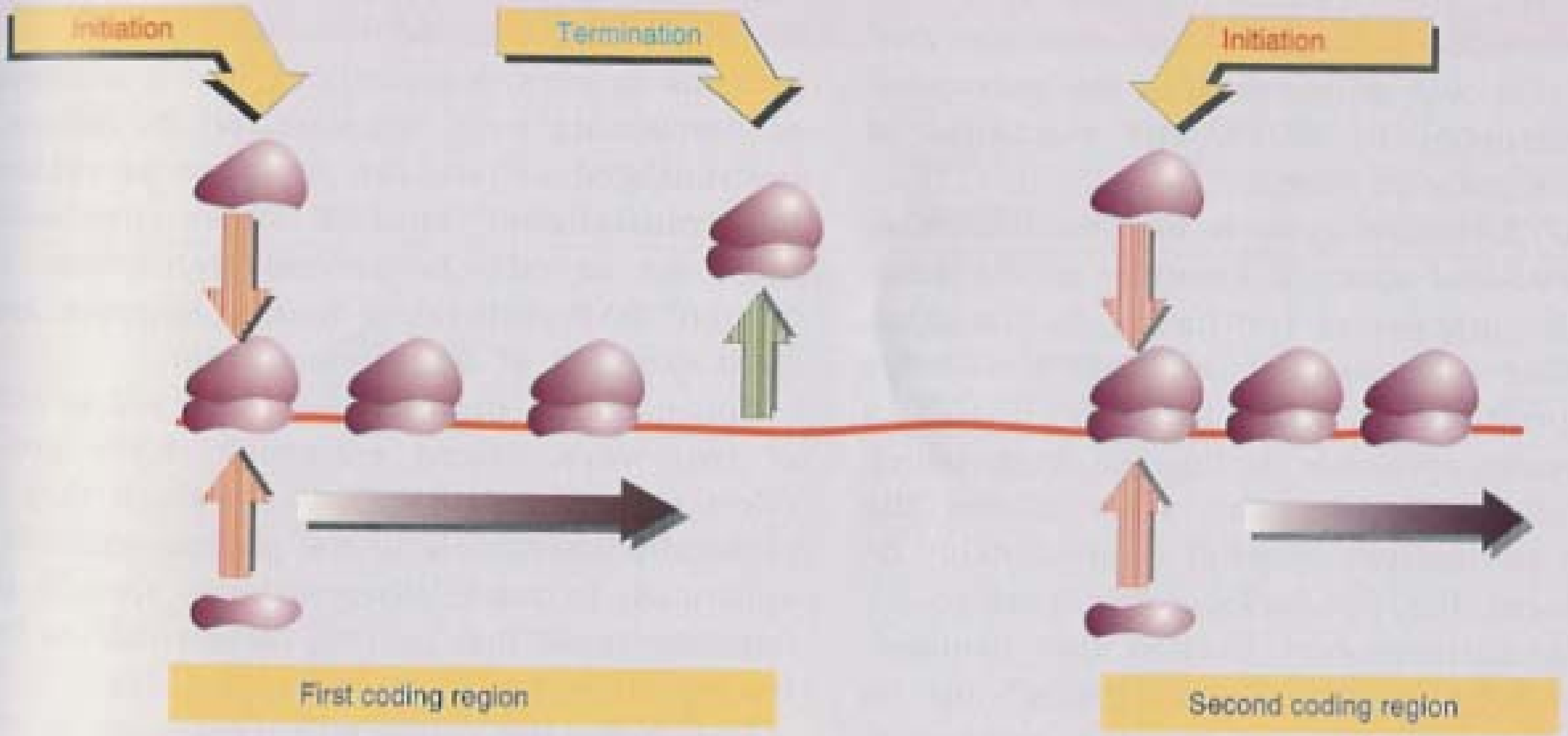
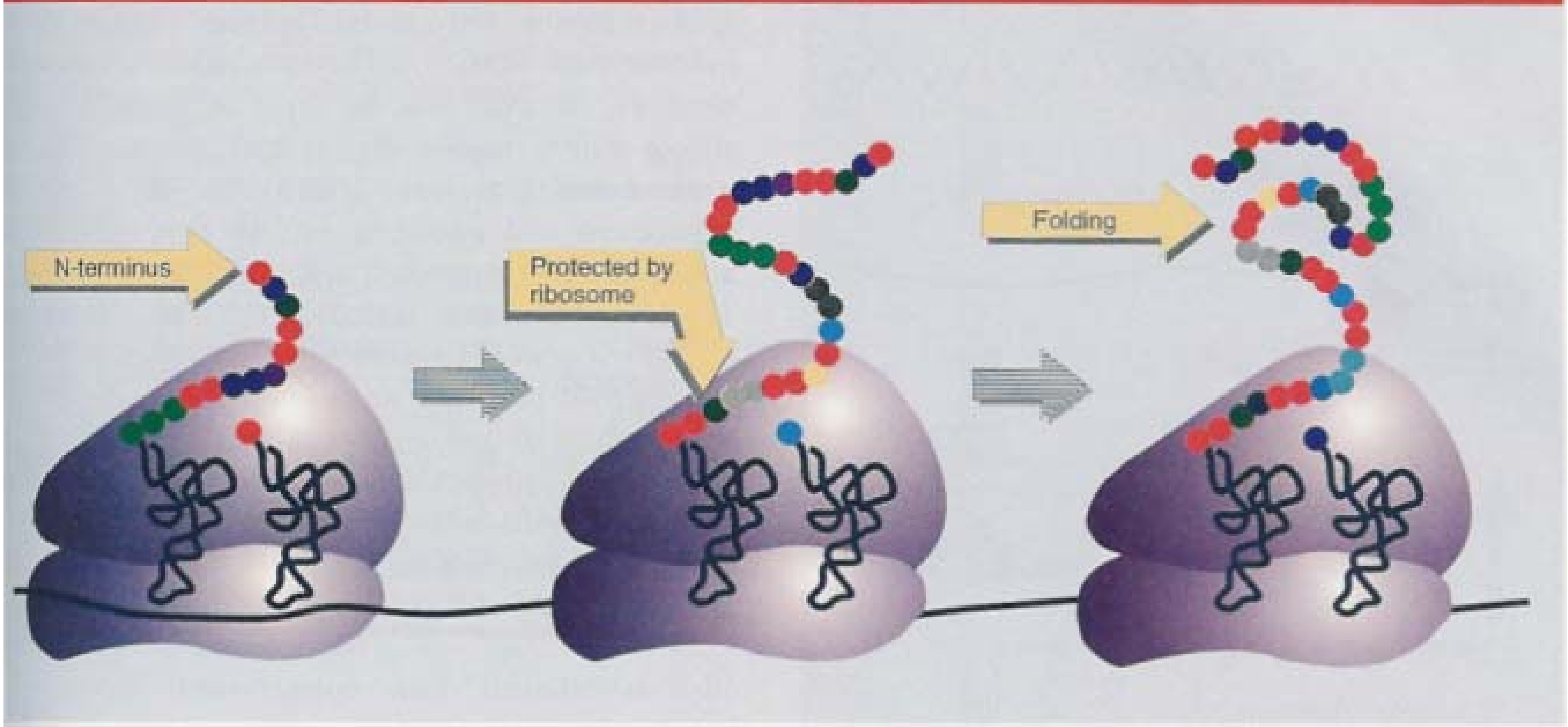
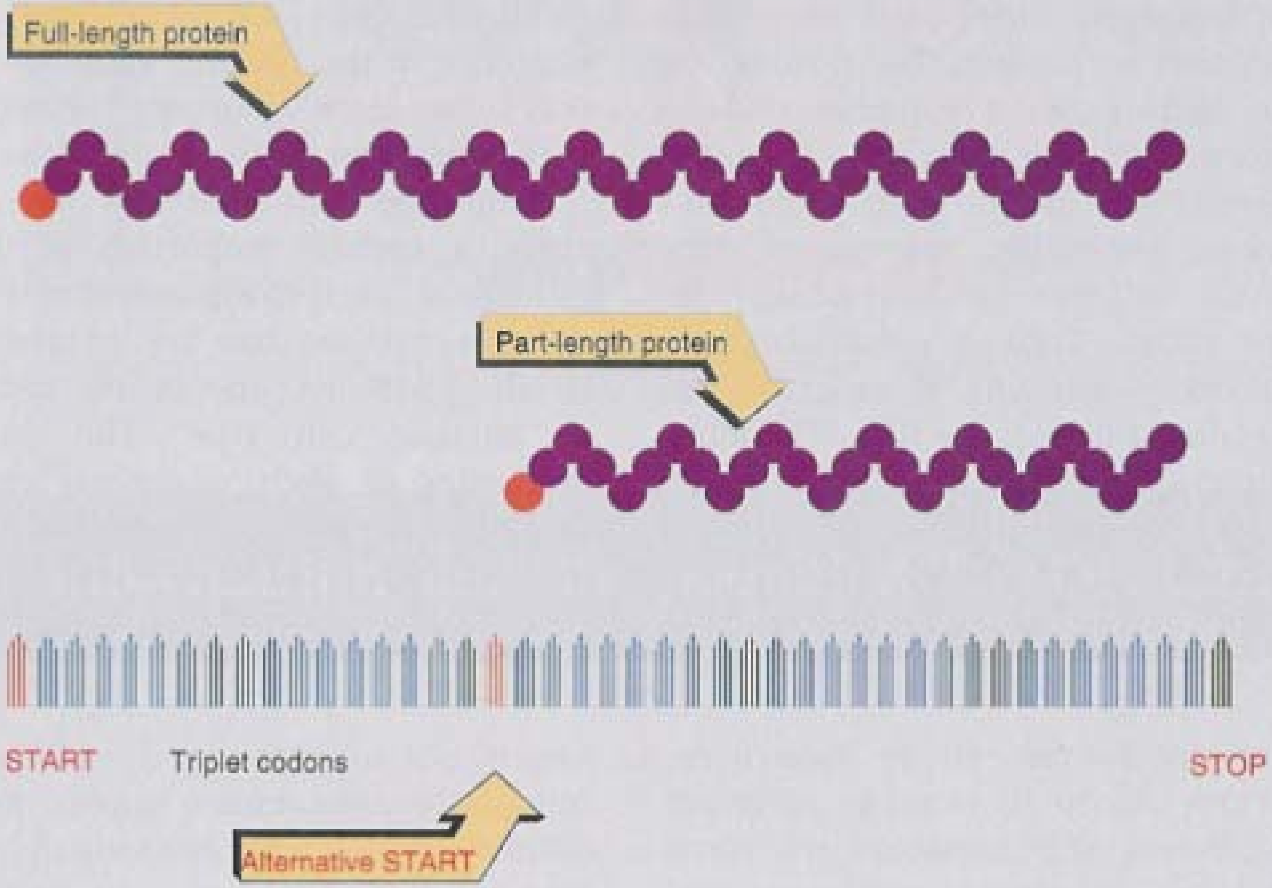


Figure 7.9 A polyribosome consists of an mRNA being translated simultaneously by several ribosomes moving in the direction from 5' to 3'. Each ribosome has two tRNA molecules, the first carrying the last amino acid added to the chain (connected to the nascent protein, i.e., the polypeptide chain synthesized so far), the second carrying the next amino acid to be added.



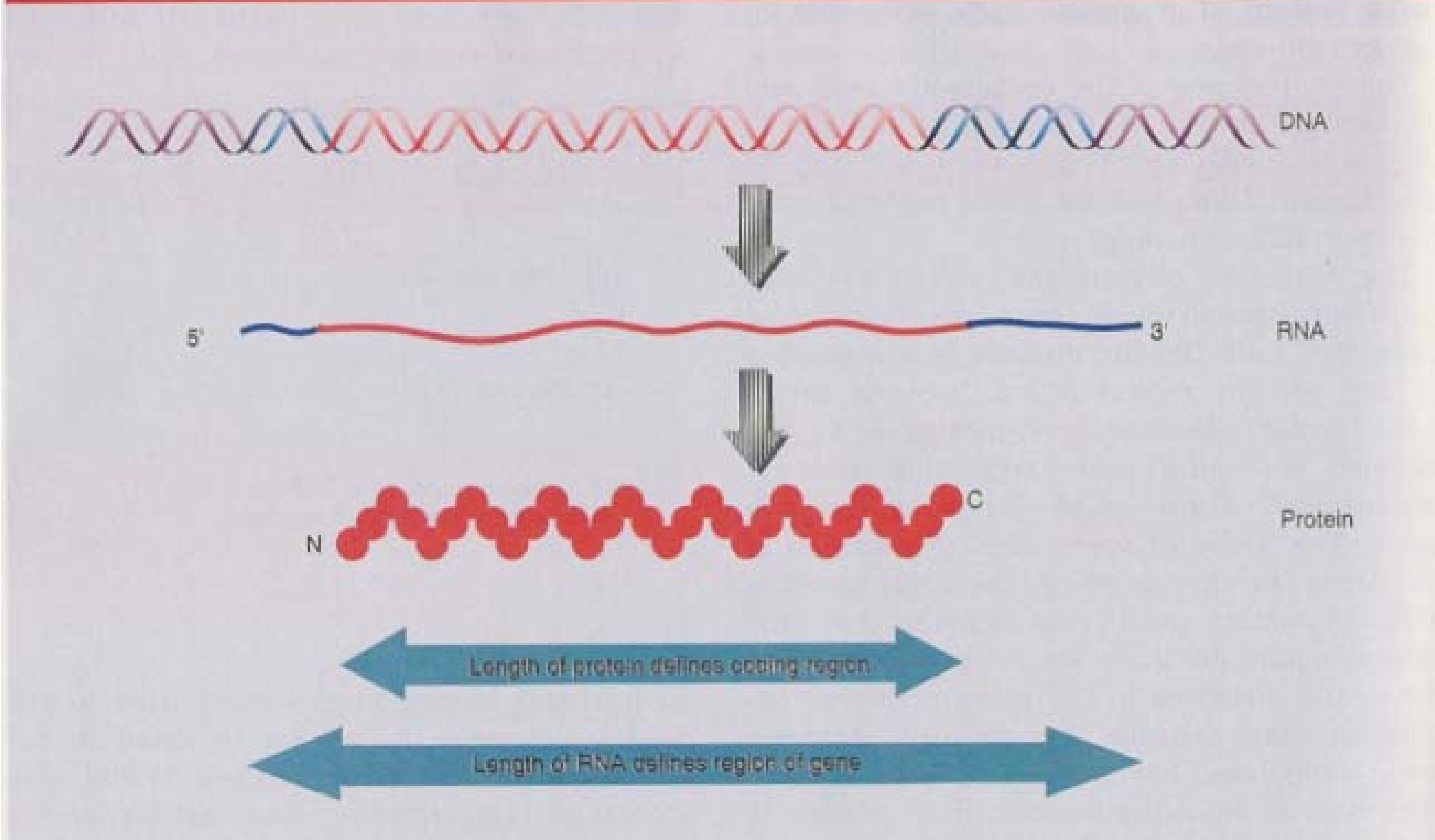
mRNAs with premature termination codons are degraded by non-mediated decay.

Figure 6.22 Two proteins can be generated from a single gene by starting (or terminating) expression at different points.



Gene structure

Figure 6.17 The gene may be longer than the sequence coding for protein.



Information processing in eukaryotes*

*cells with nuclei (containing the genome)

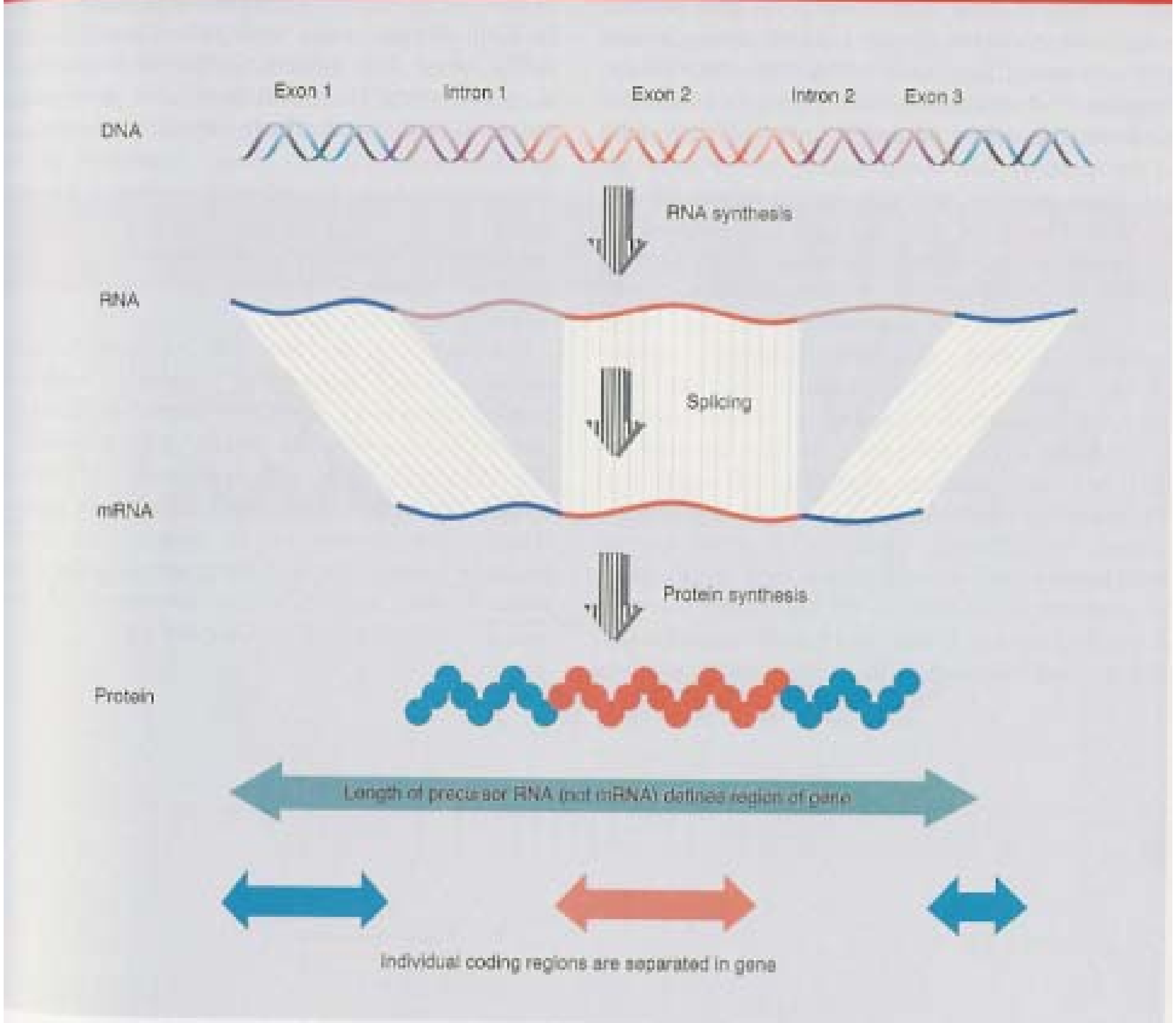
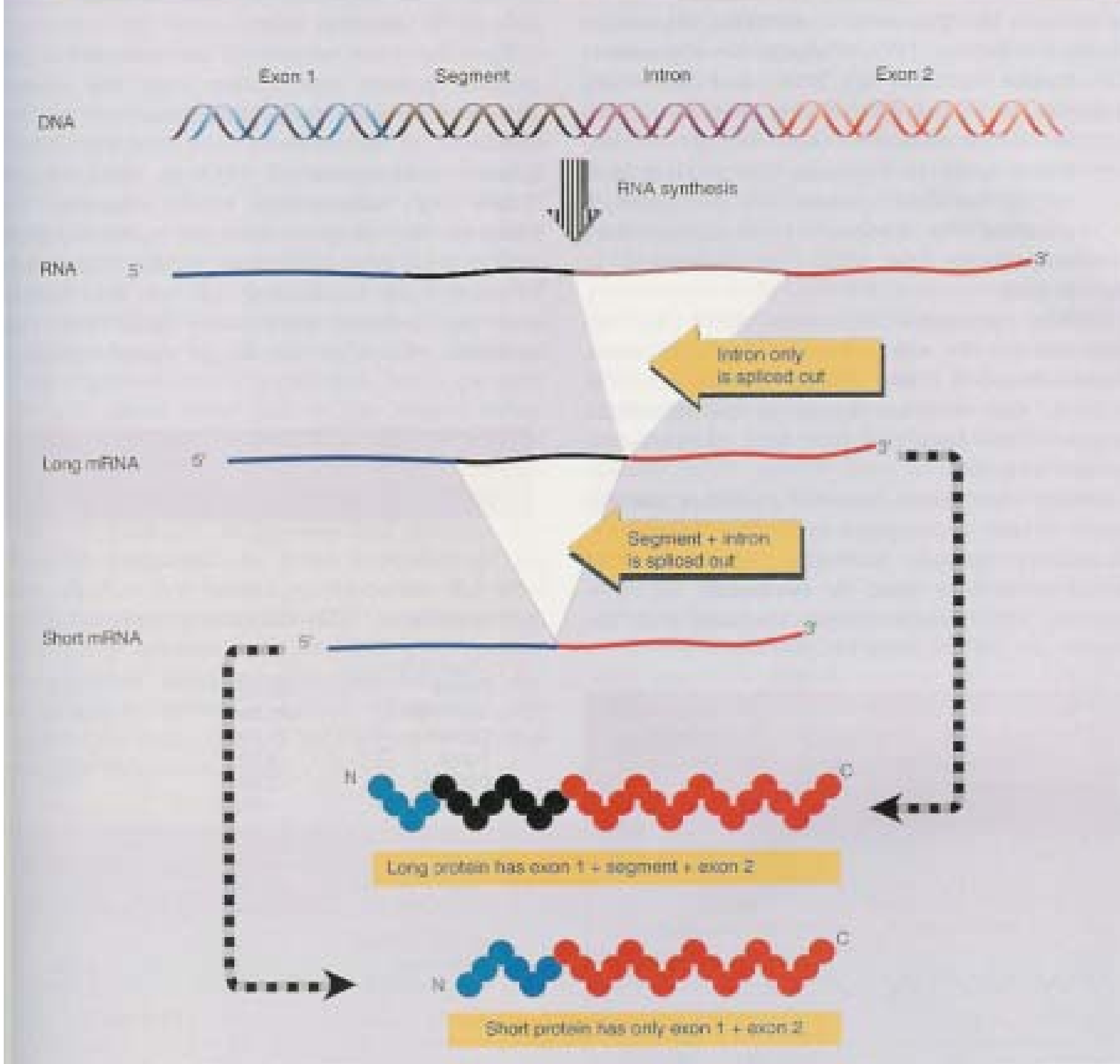


Figure 6.21 Interrupted genes are expressed via a precursor RNA, from which the introns are removed when the exons are spliced together to give the messenger RNA.

Figure 6.24 Genes may be difficult to define when there are alternative pathways for expression.



Mutations

- Types of direct mutations
- Example of process leading to mutation
- Mutation spectral distribution is sequence and mutagen dependent
- Suppressor mutations
- Regulatory sequence mutations

Figure 4.19 Frameshift mutations show that the genetic code is read in triplets from a fixed starting point.

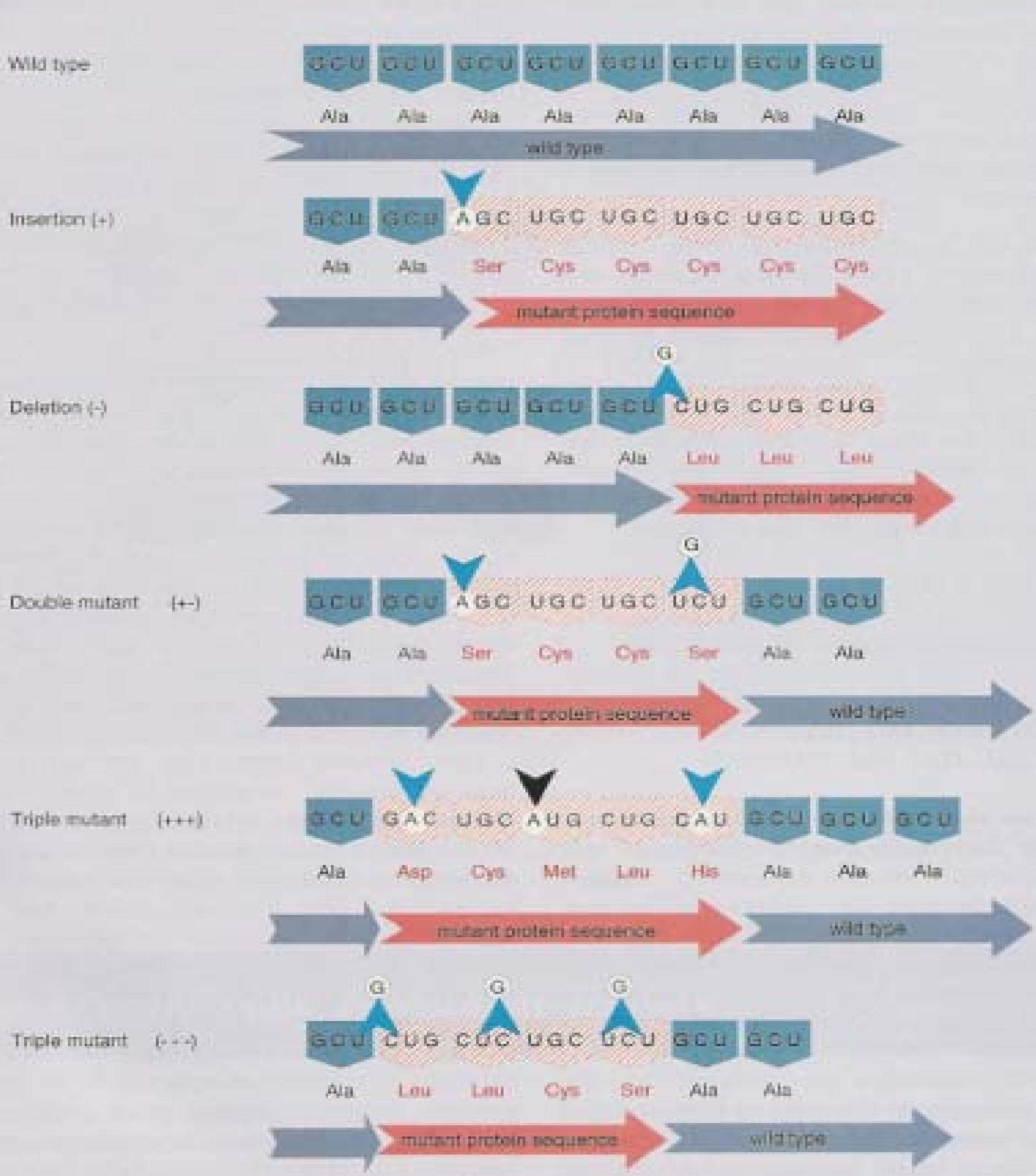
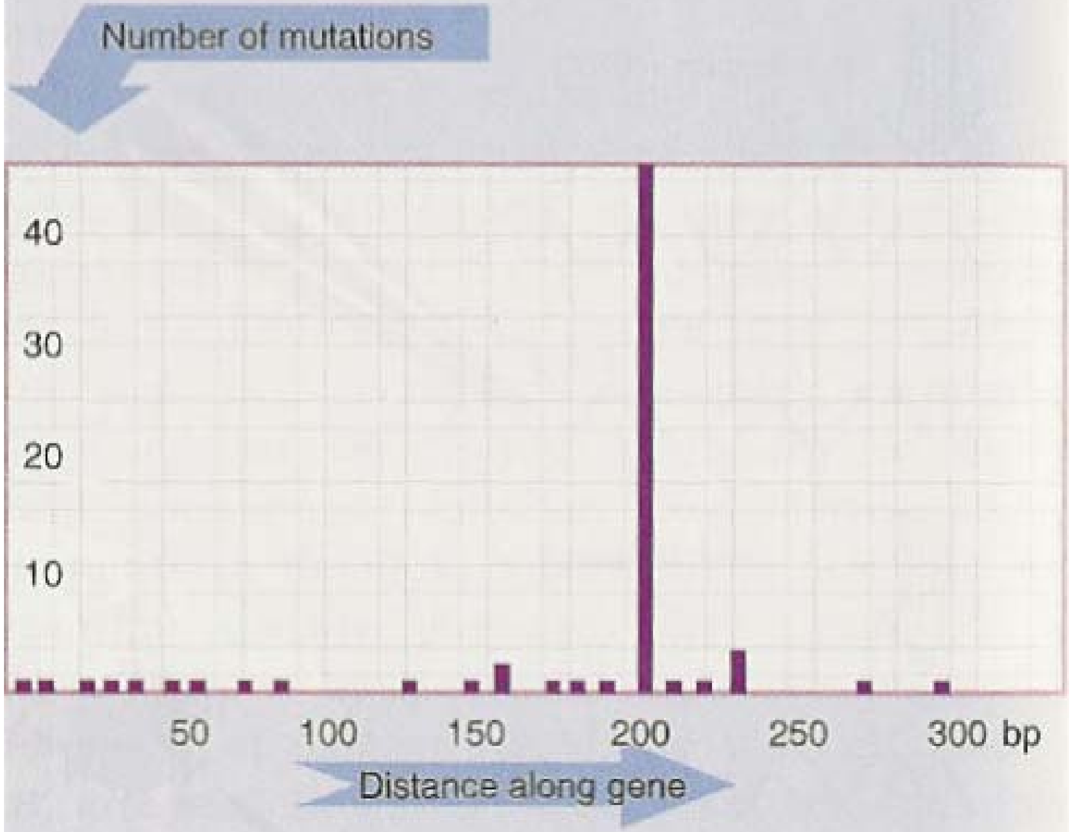
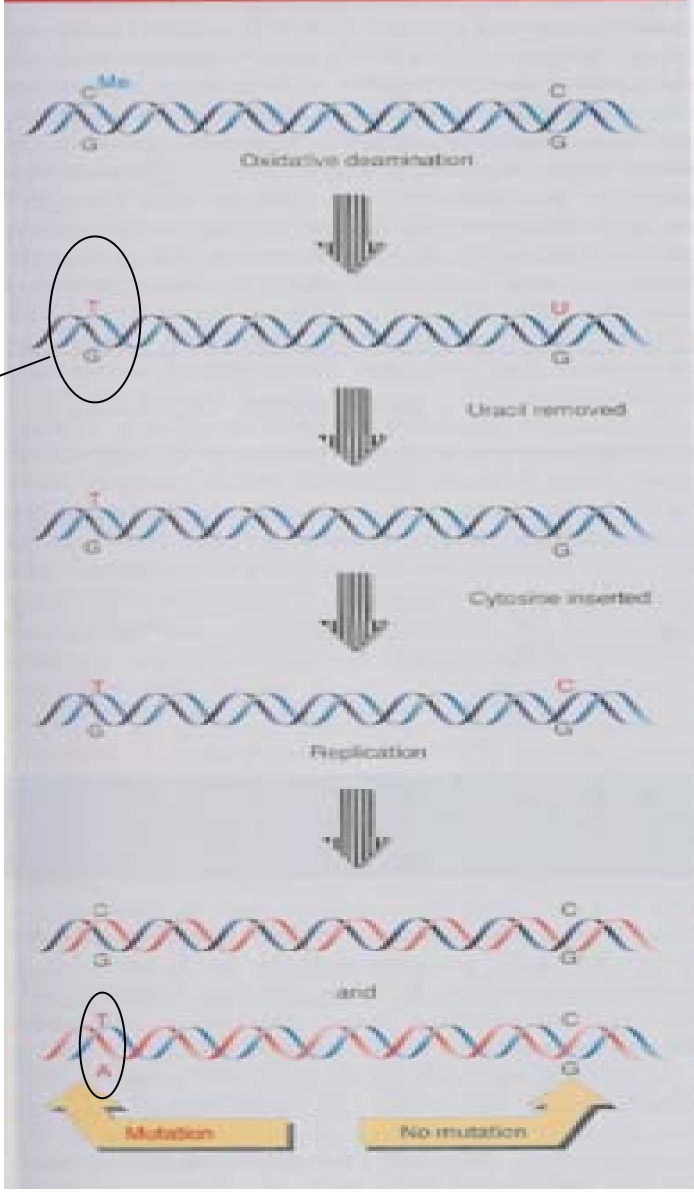


Figure 4.22 Spontaneous mutations occur throughout the *lacI* gene of *E. coli*, but are concentrated at a hotspot.



Deamination Of 5-methyl cytosine

Figure 4.23 The deamination of 5-methylcytosine produces thymine (causing C-G to T-A transitions), while the deamination of cytosine produces uracil (which usually is removed and then replaced by cytosine).



mismatch

!!!

Suppressor tRNA mutations

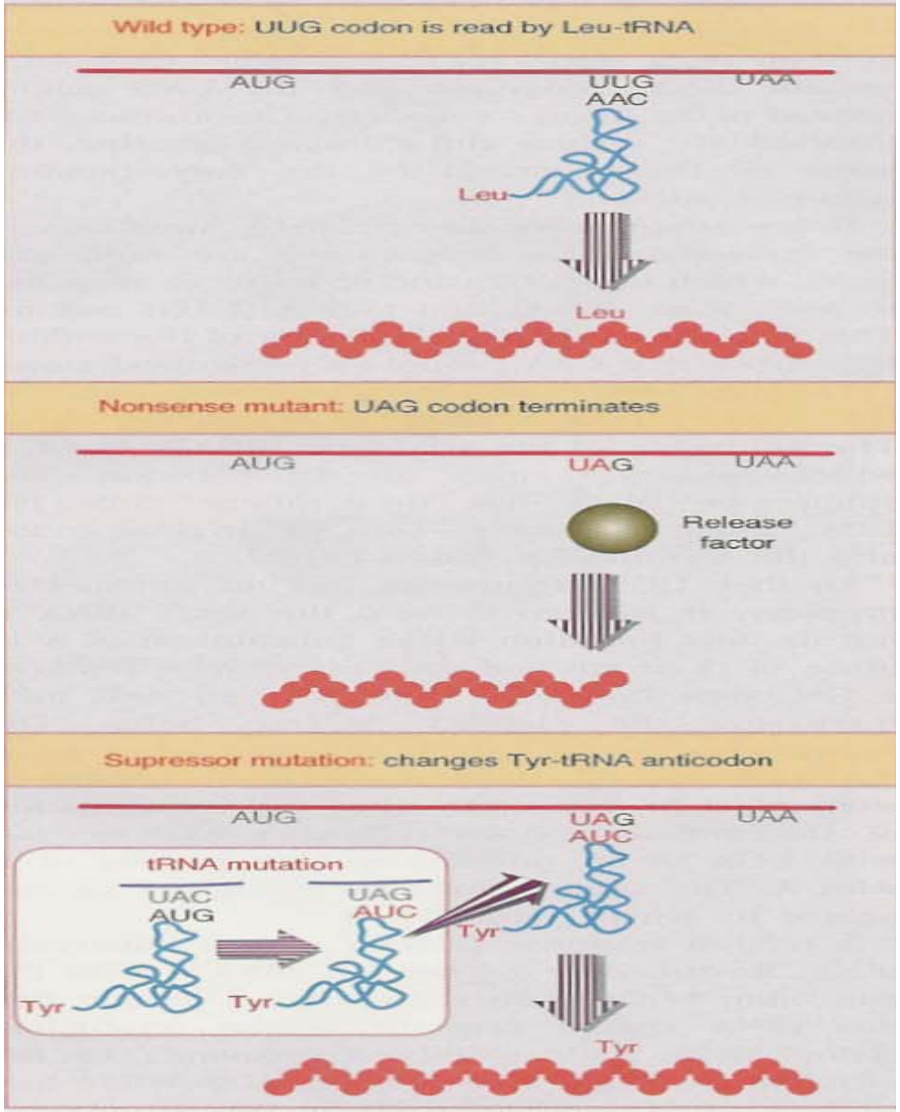
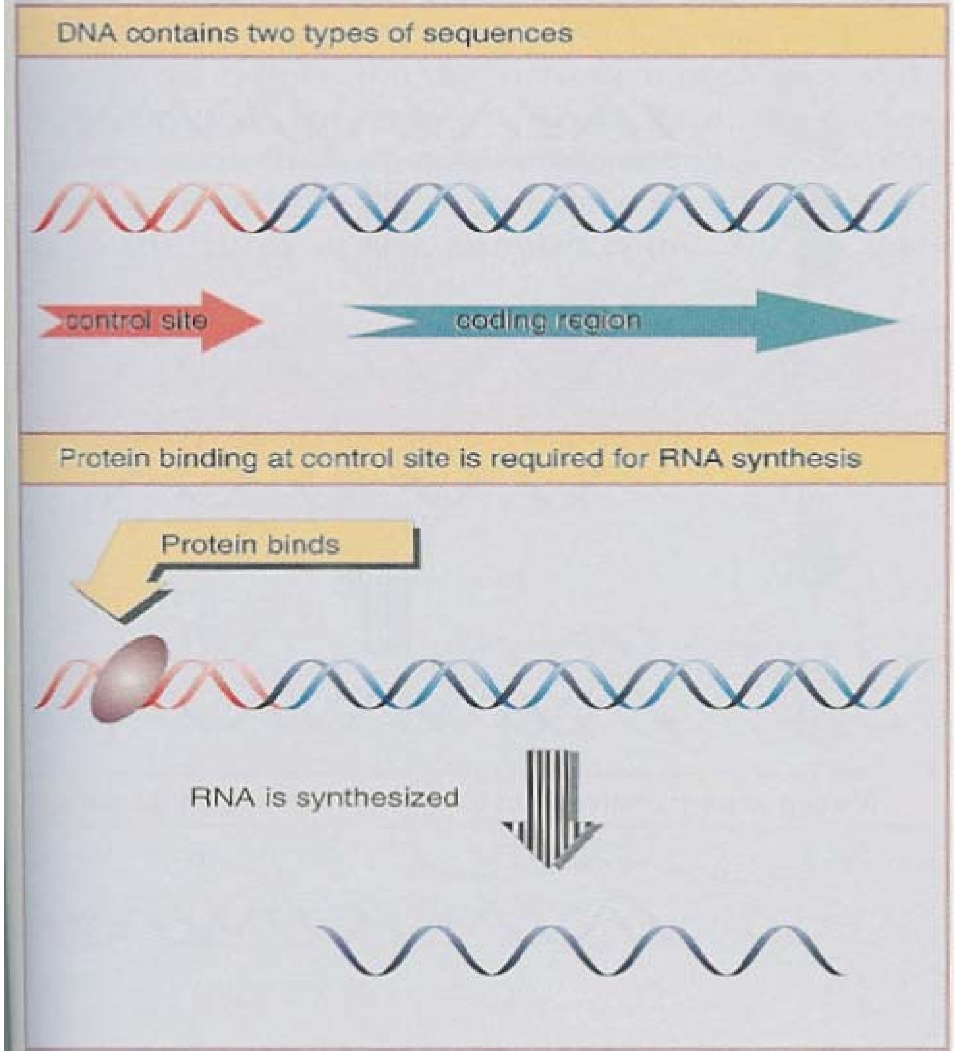


Figure 9.14 Nonsense mutations can be suppressed by a tRNA with a mutant anticodon, which inserts an amino acid at the mutant codon, producing a full length protein in which the original Leu residue has been replaced by Tyr.

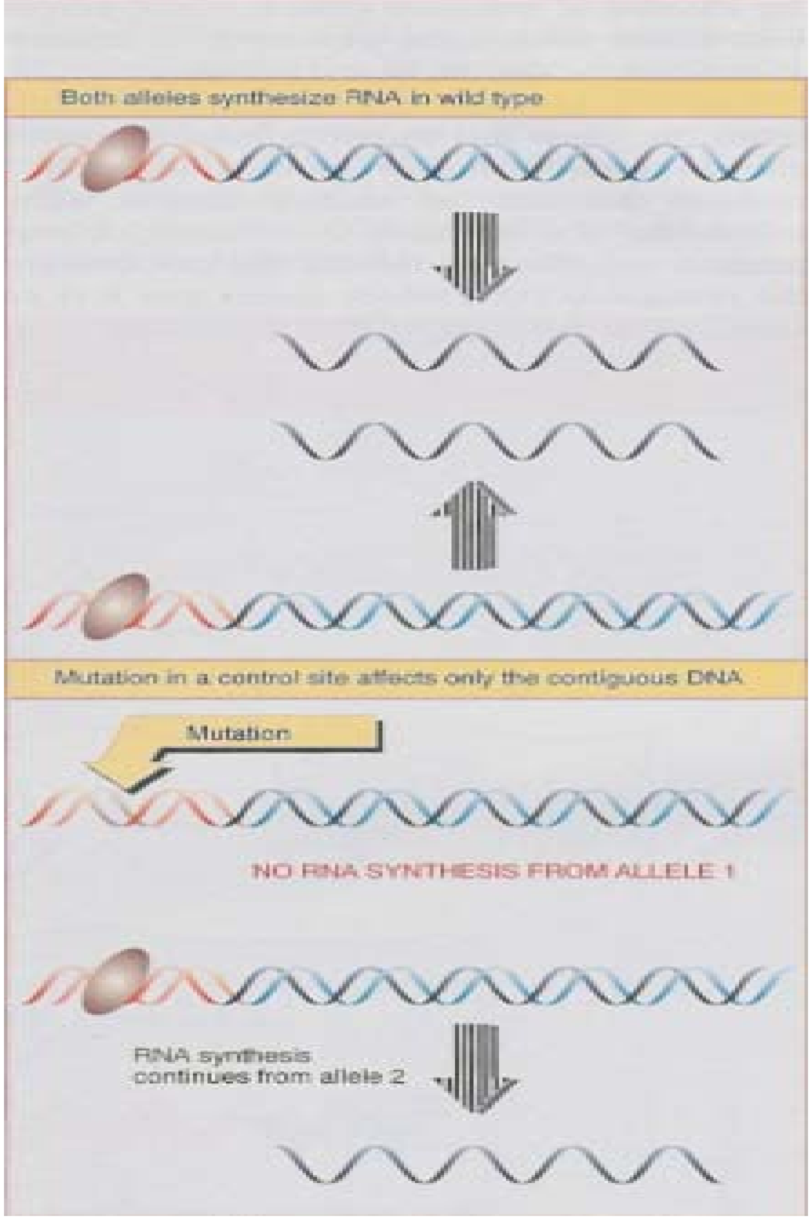
Figure 6.18 Control sites in DNA provide binding sites for proteins; coding regions are expressed via the synthesis of RNA.



DNA contains two types of sequences

Protein binding at control site is required for RNA synthesis

Figure 6.19 A *cis*-acting site controls the adjacent DNA but does not influence the other allele.



Summary

- Universal composition of nucleic acids
- Structure of DNA suggests replication mechanism
- Triplet genetic code redundantly defines amino acids
- tRNA is the information transducer
- Proteins are characterized by primary, secondary & tertiary structures
- Gene structures and regulation of gene expression differ in prokaryotes and eukaryotes
- Mutations come in many flavors