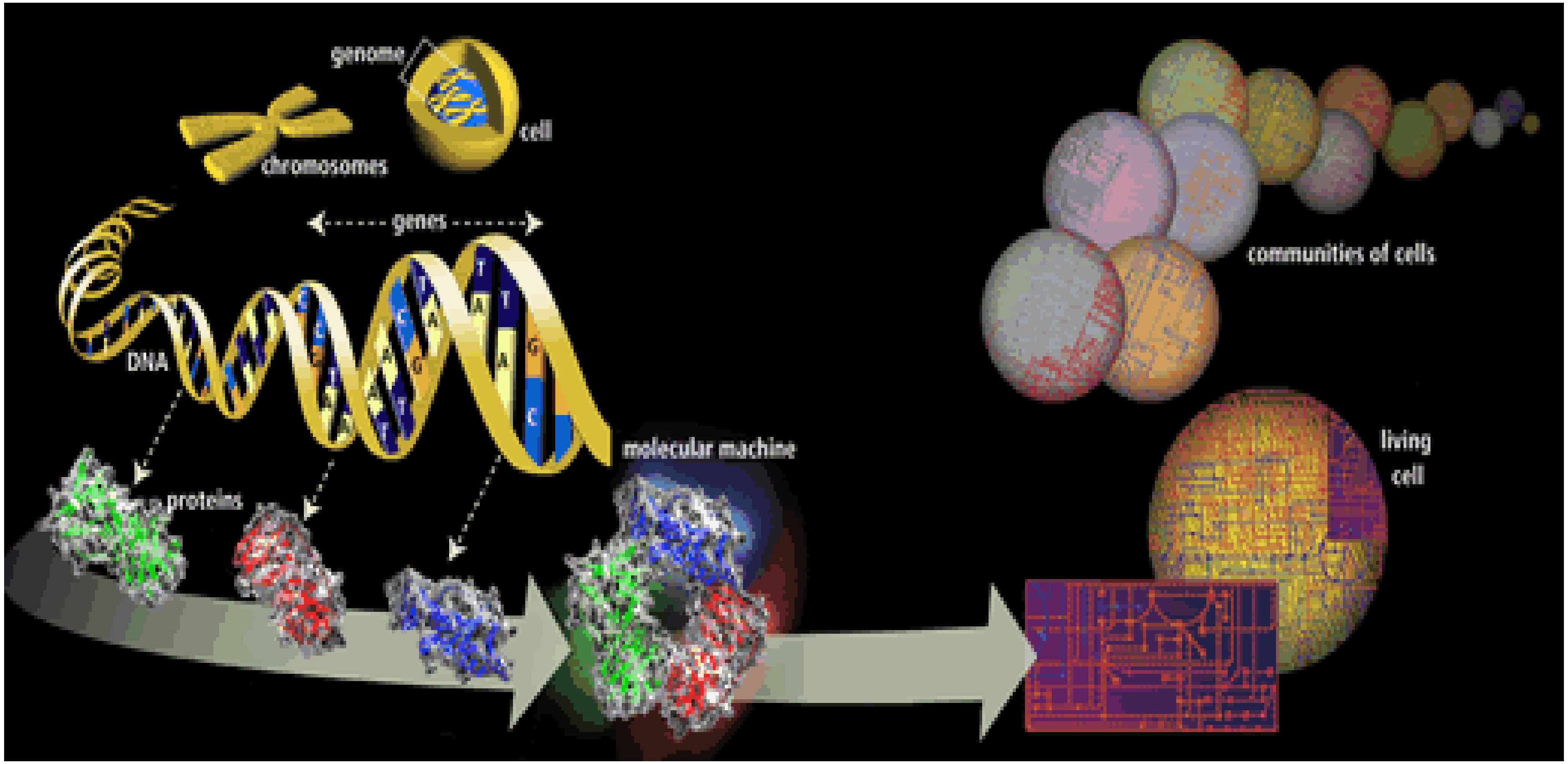
Chapter 0. DNA and the flow of genetic information

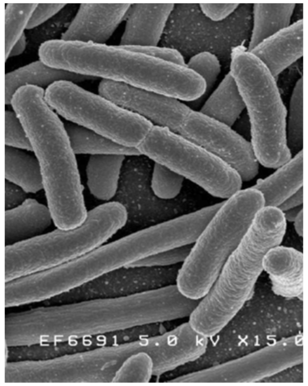


Contents of the chapter

- Cells: the fundamental units of life
  - Unity and diversity of the cells
  - Different cellular organization
  - Model organisms
- Properties of a living cell
  - The use of energy
  - Informative molecules : proteins and nucleic acids
- Structure of proteins and DNA
- From DNA to protein: how cells read the genome
- Overview of gene expression

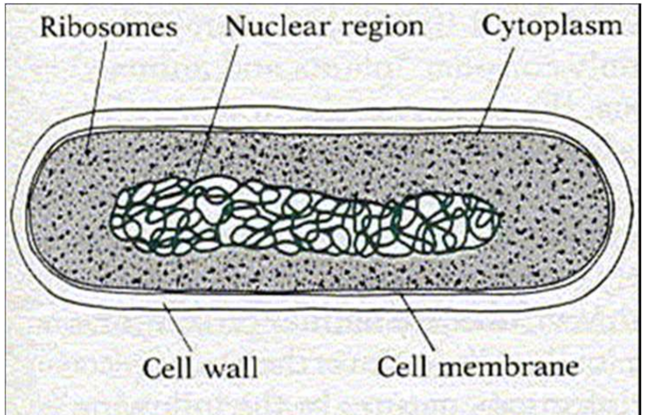
Cells: the fundamental units of life

- All living organisms are constructed from cells

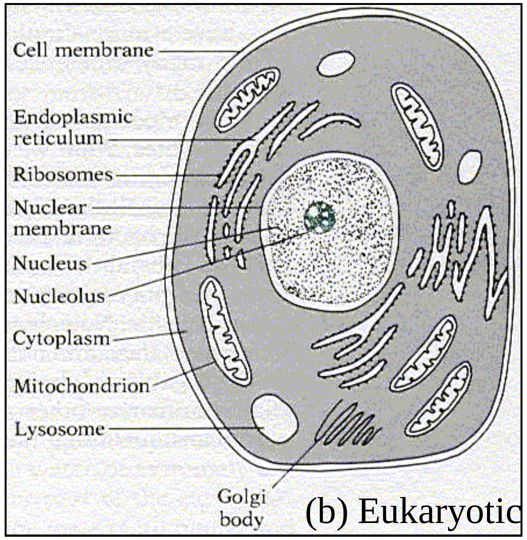



A colony of bacteria, a carnivorous plant and a panda are all made of cells that have a fundamentally similar chemistry and operate according to the same basic principles
... even though they come in a variety of shapes and sizes

- All cells are enclosed by a plasma membrane, which separates the inside of the cell from its environment



(a) Prokaryotic



(b) Eukaryotic

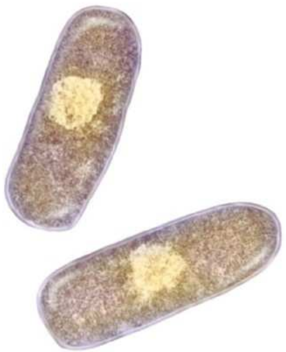
- All cells contain DNA as a store of genetic information and use it to guide the synthesis of RNA and proteins
- Eukaryotic cells posses a nucleus containing genomic DNA and other organelles not found in prokaryotes

- Model organisms

Eukaryote

Prokaryote

(a) ***Escherichia coli*** **(bacterium)**



Circular chromosome

Genome size :4.64 x 10^6 bp

Gene number: 4289

Unicellular

(b) ***Saccharomyces cerevisiae*** **(levure)**

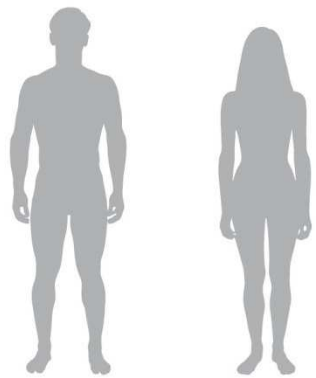
16 pairs of linear chromosomes

Genome size: 12.1 x 10^6 bp

Gene number: 6100

Multicellular

(c) ***Homo sapiens*** **(homme)**



24 pairs of linear chromosomes

Genome size: 3.4 x 10^9 bp

Gene number : 32000

- Animals, plants and some fungi consist of diverse eukaryotic cell types having specialized functions
- They turn on different sets of genes according to developmental history and environmental signals

Properties of a living cell

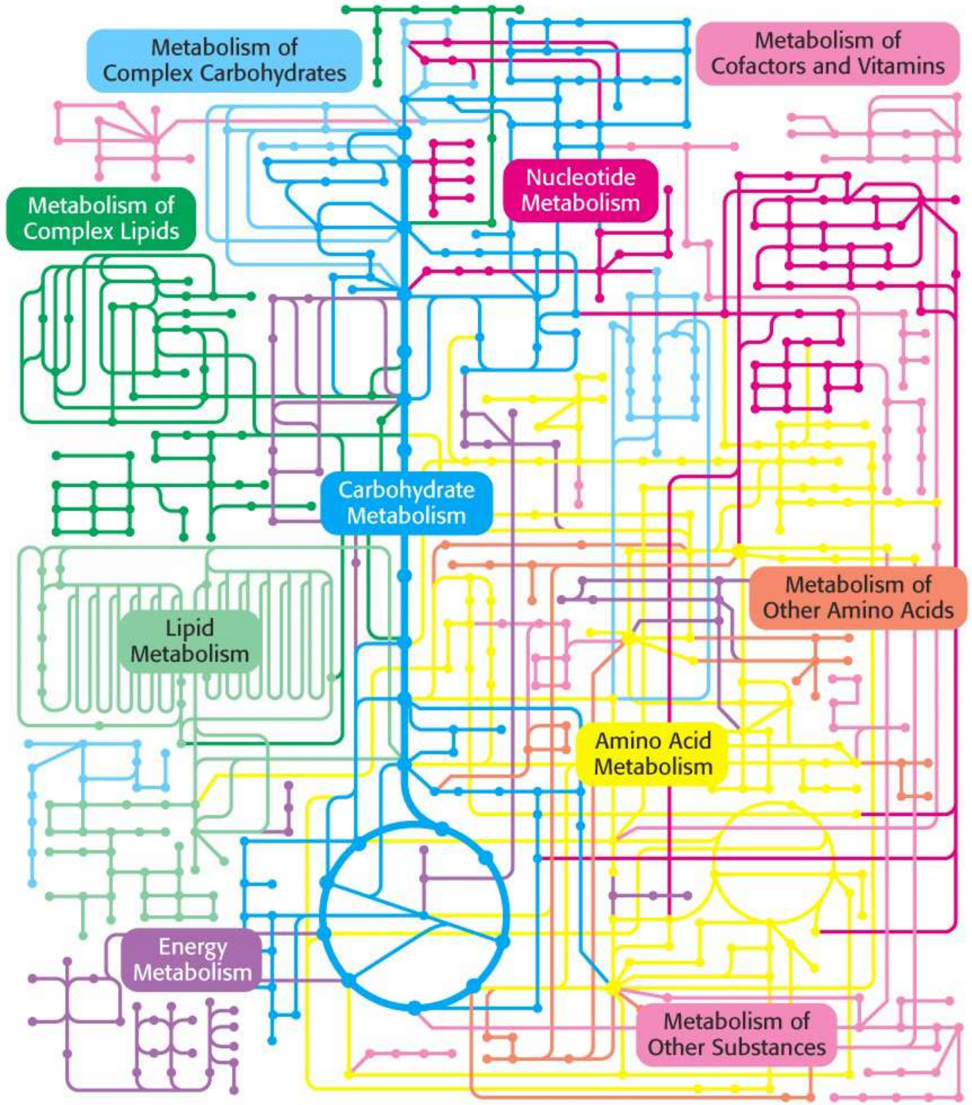
- Living organisms are composed of carbon-based (organic) molecules:
  - Sugars are a primary source of chemical energy and can also be joined together to form polysaccharides (storage or plant wall component)
  - Fatty acids are an even richer energy source than sugars; their most essential function is to form lipid molecules that assemble into cell membranes
  - Proteins , the most diverse and versatile class : enzymes, transporters, signal transducers, structural proteins, receptors, motors,…
  - Nucleic acids (RNA and DNA) consiting of 4 types of nucleotides

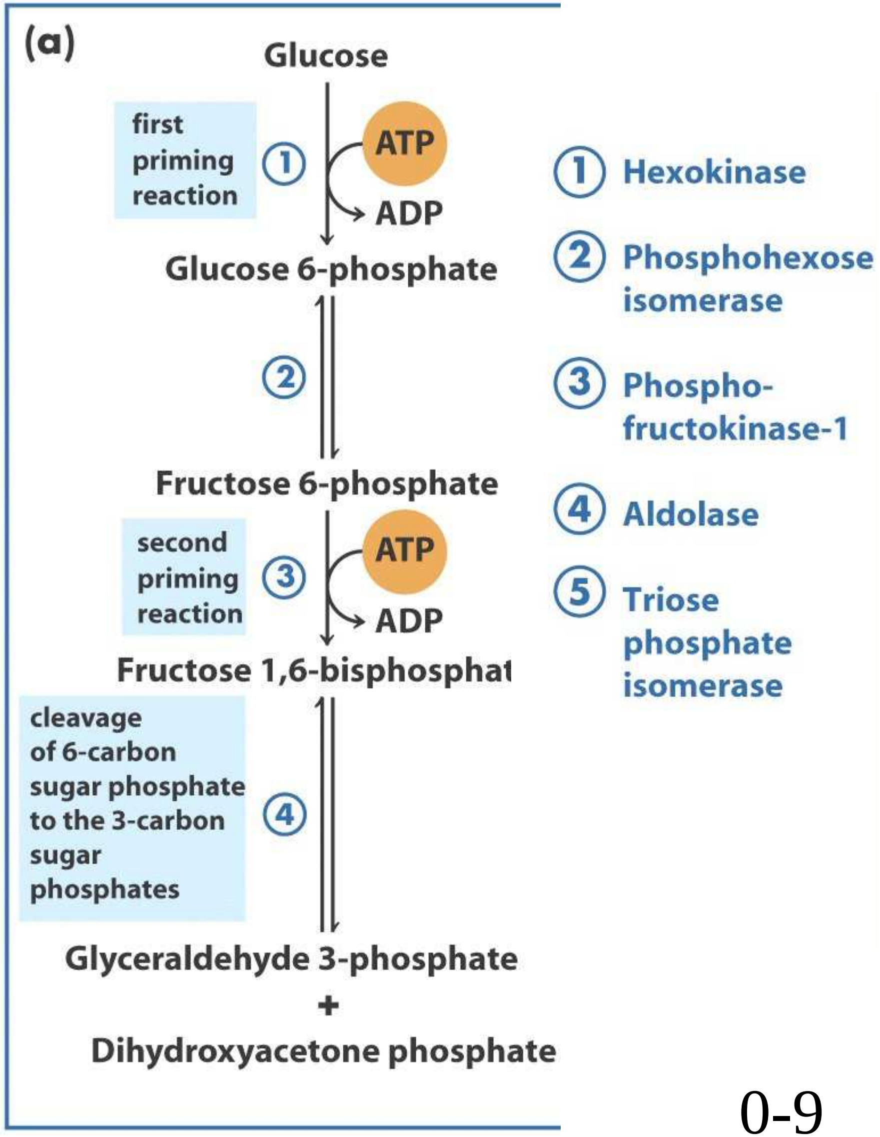
- Living organisms are composed of lifeless molecules that conform to all the physical and chemical laws described for inanimate matter
- Yet living organisms possess attributes not exibited by any random collection of molecules
- A living cell is a self contained, self-assembling, self-adjusting, self perpetuating system

The use of energy by cells

- The cell extracts free energy and raw material from its environment
- The chemical transformations are organized into a network of reaction pathways (metabolism) catalyzed by enzymes
  - The coupling of an energetically favorable reaction to an energetically unfavorable one allow otherwise impossible chemical transformation to occur
- The energy maintains the cell in a dynamic steady state, far from equilibrium
- The ultimate source of energy is the sun

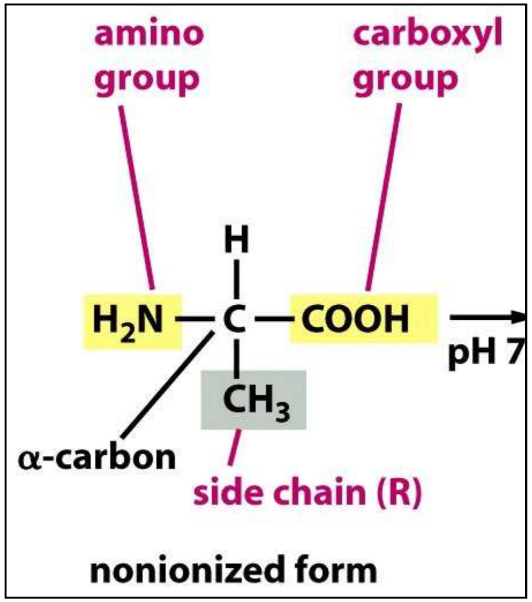
Chemical transformations are organized into metabolic pathways, including regulatory loops

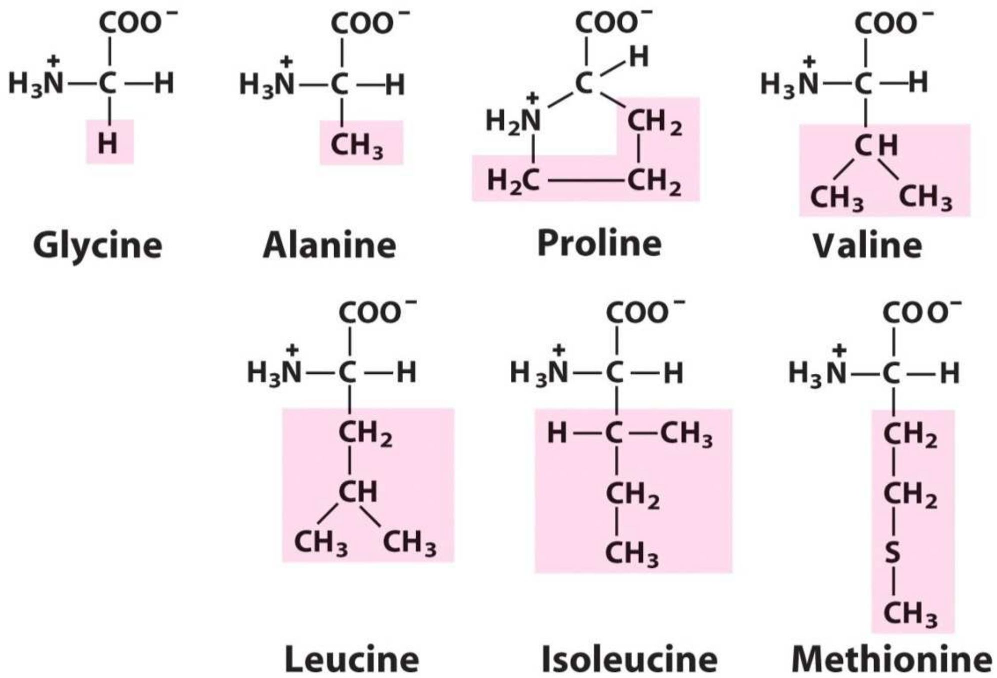




Structure of proteins

- Living cells contain an enormously diverse set of proteins, each made as a linear chain of amino acids linked together by covalent peptide bonds

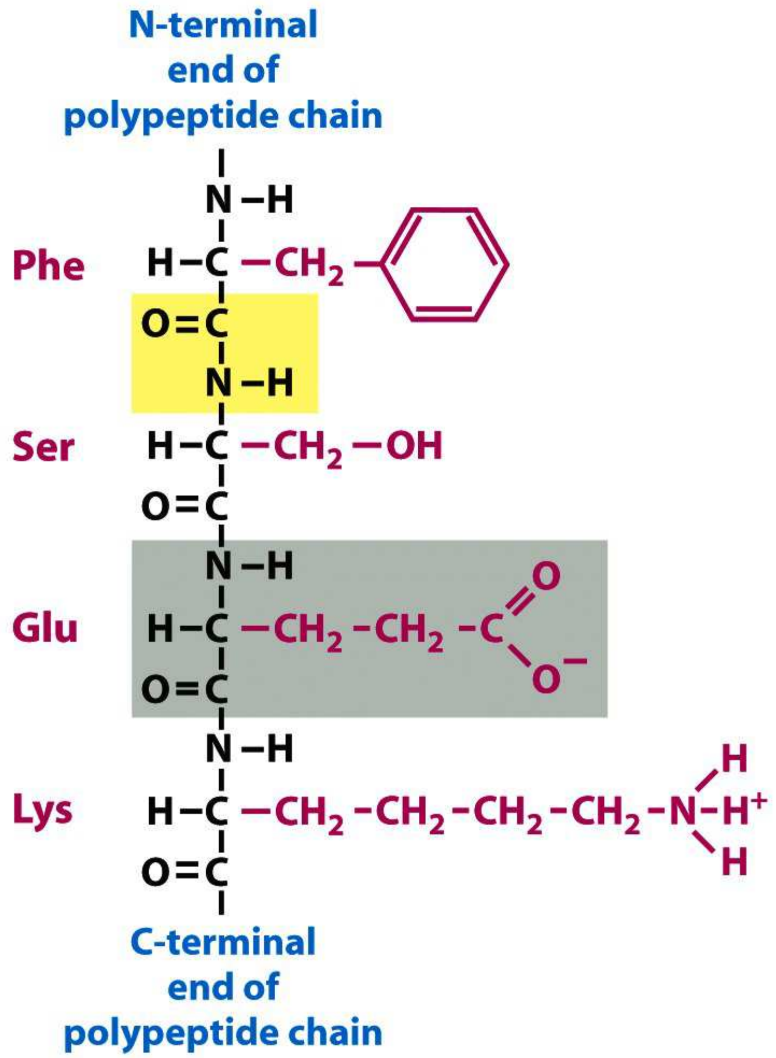




Structure of a L-amino acid (monomer)

20 amino-acids differing by the nature of the R group are used to make proteins

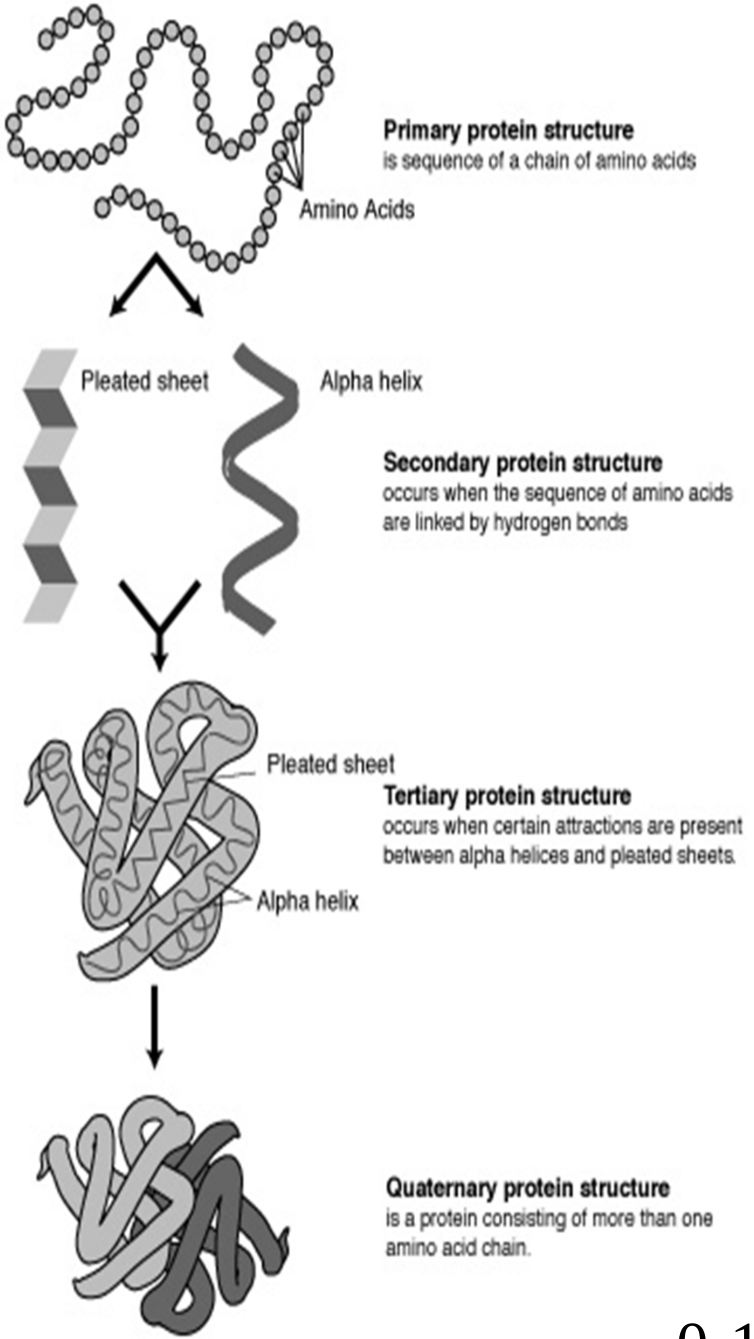
- Each type of protein has a unique amino acid sequence, which determines both its 3D shape and its biological activity



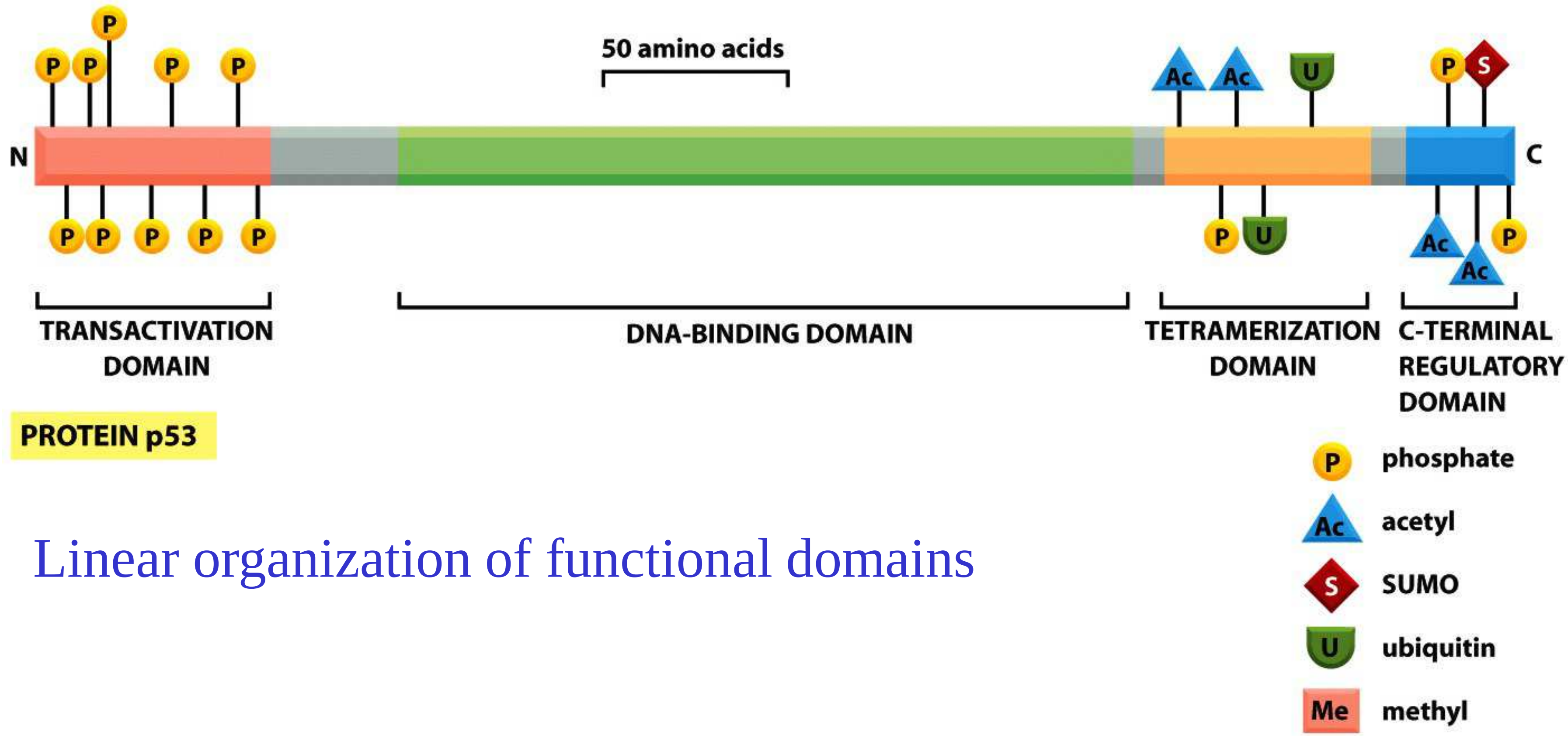
Primary structure of a polypeptide chain

Secondary and tertiary protein structures

- The structure that a protein adopts is vital to its chemistry
- Its structure determines which of its amino acids are exposed and carry out the protein's function
- Its 3-D structure also determines what substrates/ligands it can react with



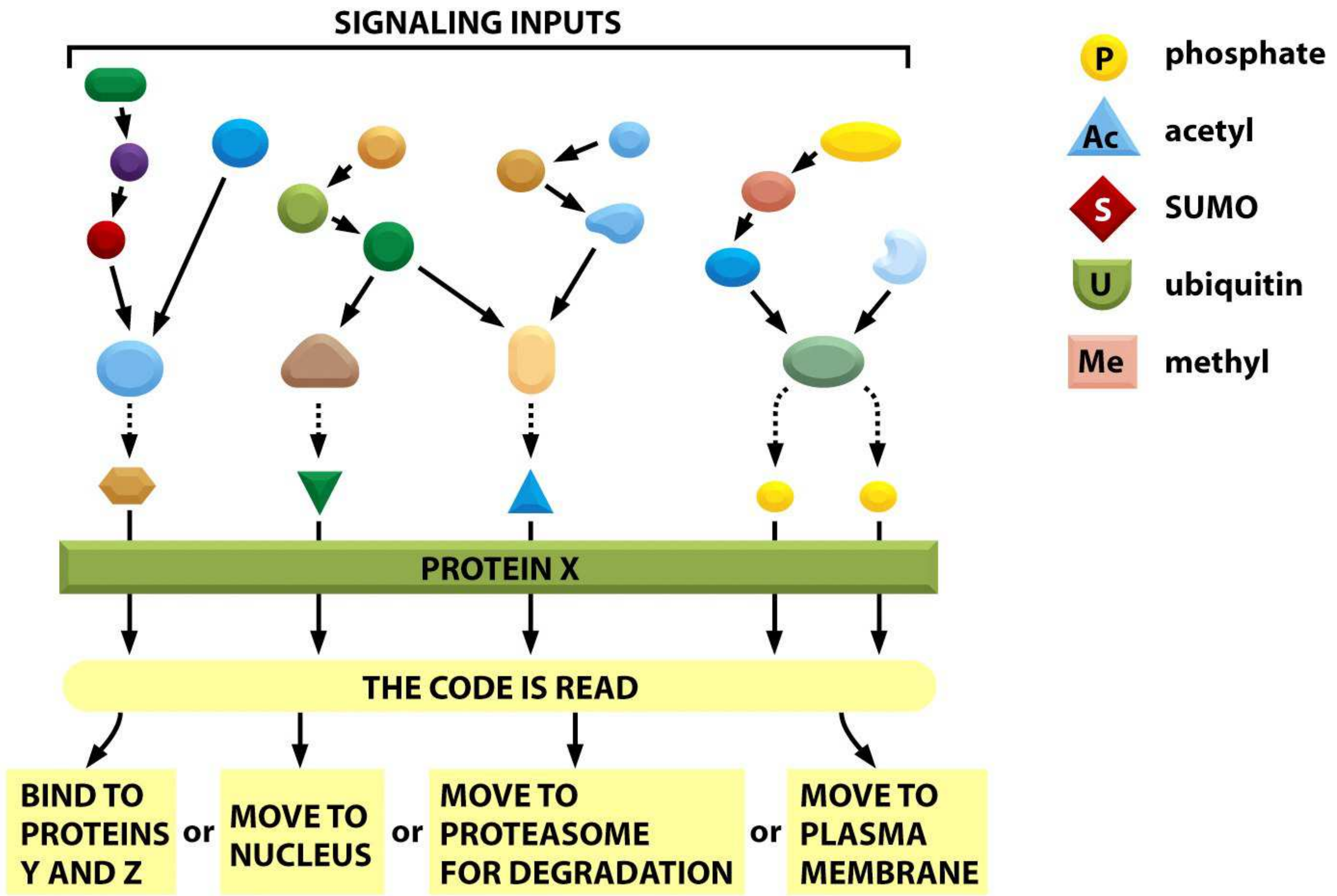
- The structure of many proteins can be subdivided into smaller regions having independent function or 3-D structure = domains



Linear organization of functional domains

Figure 3-81a *Molecular Biology of the Cell* (© Garland Science 2008)

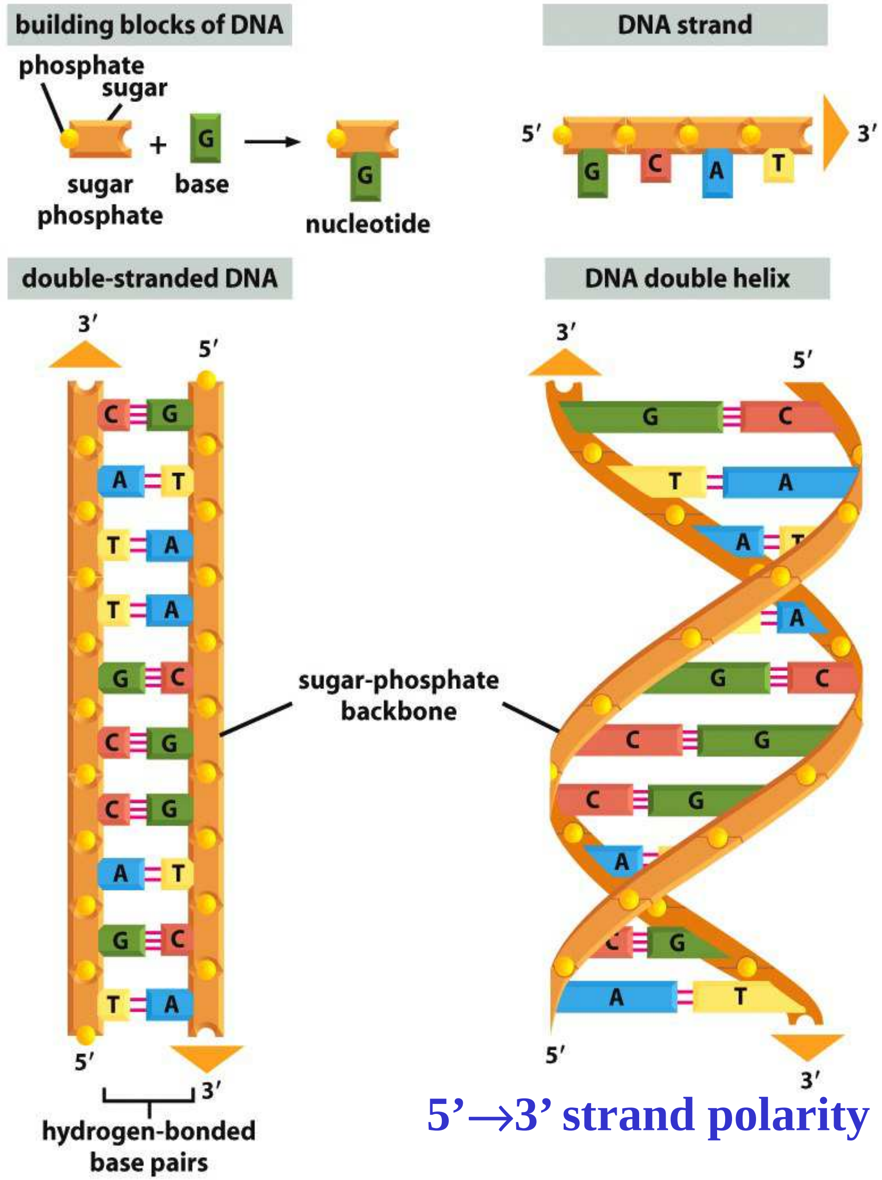
- Covalent modifications added to a protein's amino acid side chains can control the location and function of the protein and can serve as docking sites for other proteins



Life depends on the stable storage and inheritance of genetic information

- Genetic information is carried out by DNA and is encoded in the linear sequence of 4 nucleotides :A,T,G and C

- A DNA molecule consists of two complementary chains of nucleotides
- The chains/strands run in opposite directions to each other and are therefore anti-parallel



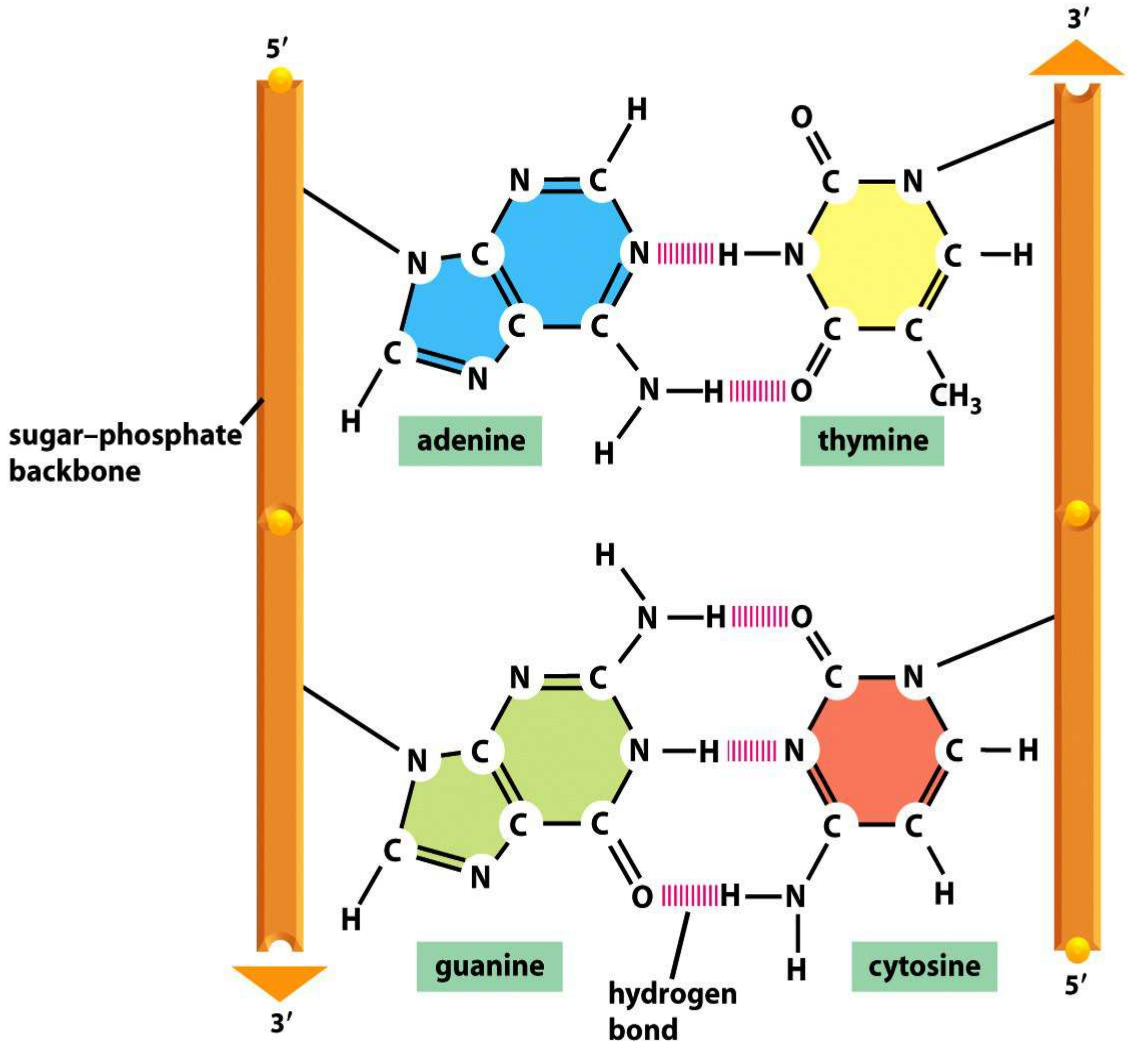
5’→3’ strand polarity

• Complementary DNA strands are held together by hydrogen bonds between G-C and A-T base pairs

- Adenine can be paired with thymine **A = T**

- and guanine with cytosine **G ≡ C**

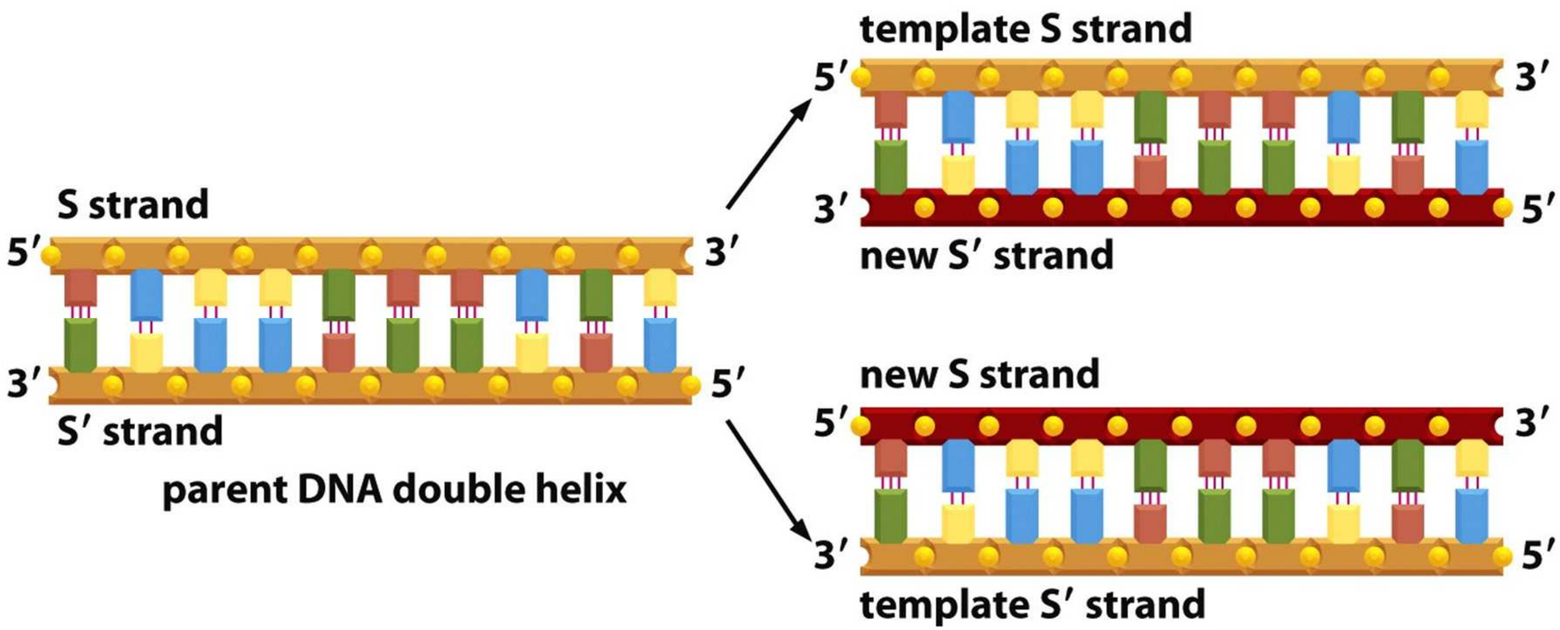
Note that base pairs have the same shape (Watson et Crick)



Base pairing enables DNA replication

- DNA acts as a template for its own duplication

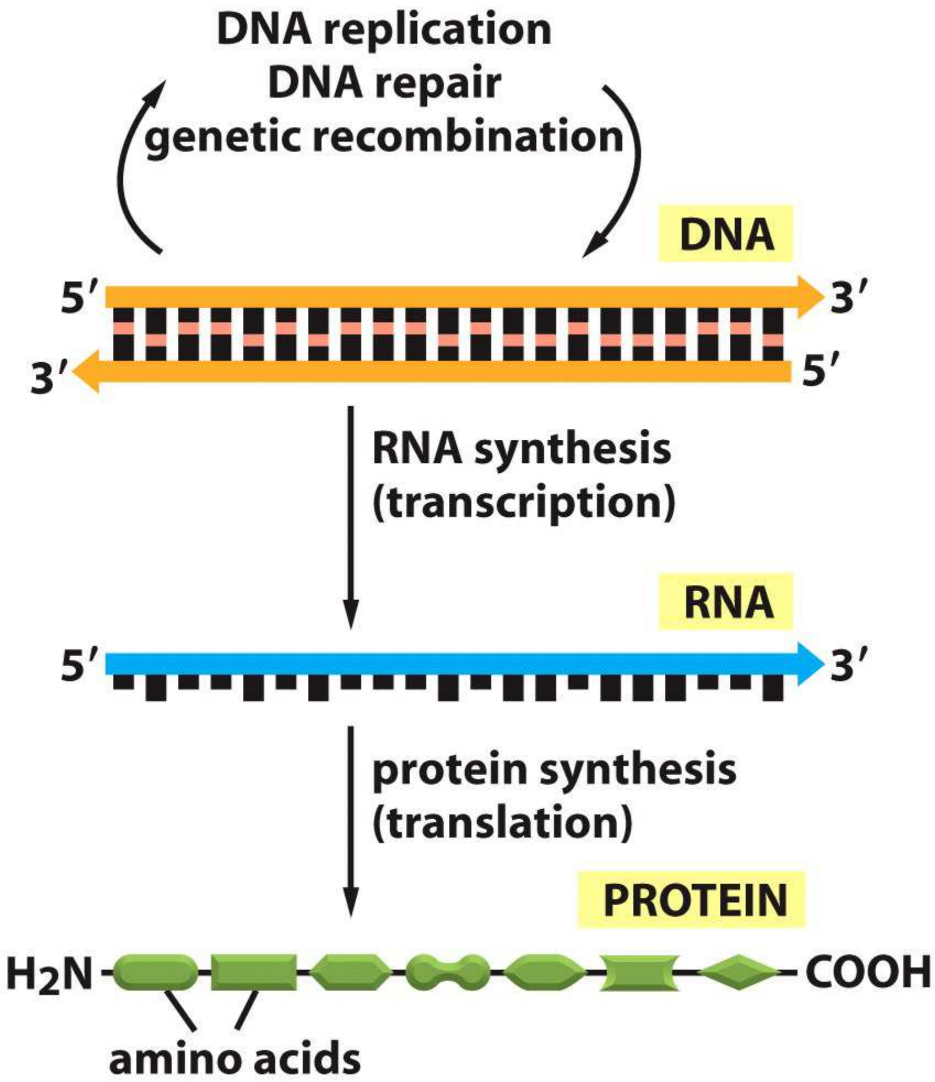
Because the nucleotide A will pair only with T, and G with C, each strand S and S′ can serve as a template to specify the sequence of nucleotides in its new complementary strand



- DNA synthesis begins at replication origins (ORIs)

From DNA to protein: How cells read the genomes

- Genetic information directs the synthesis of protein



Gene expression

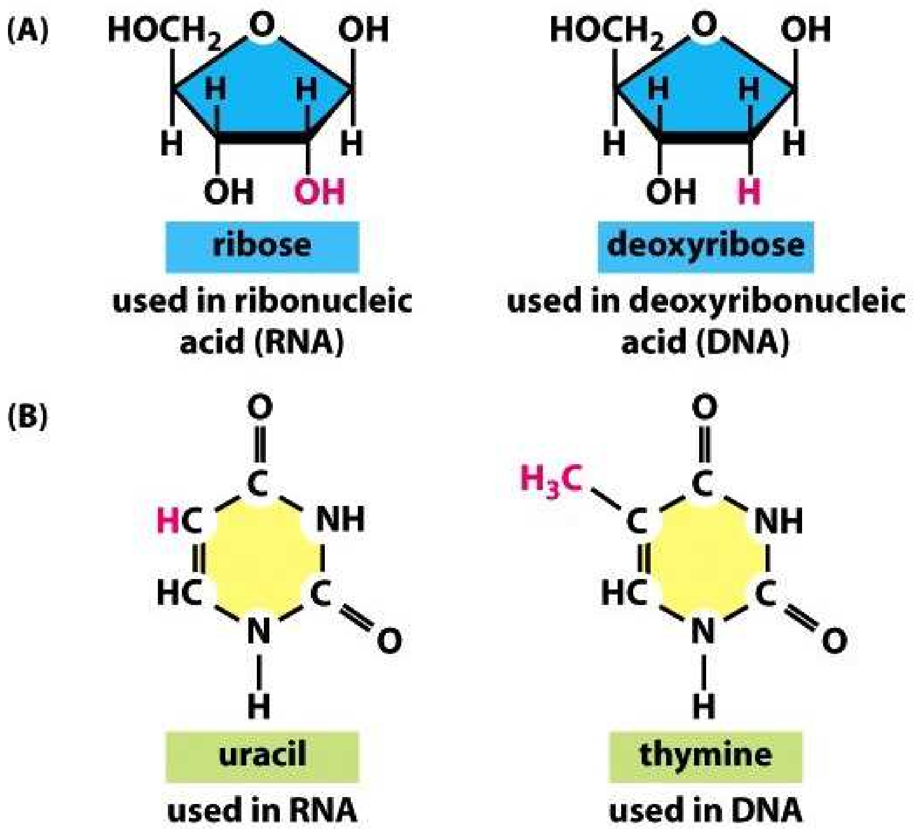
Information coded in the sequence of base pairs in DNA is passed to molecules of RNA

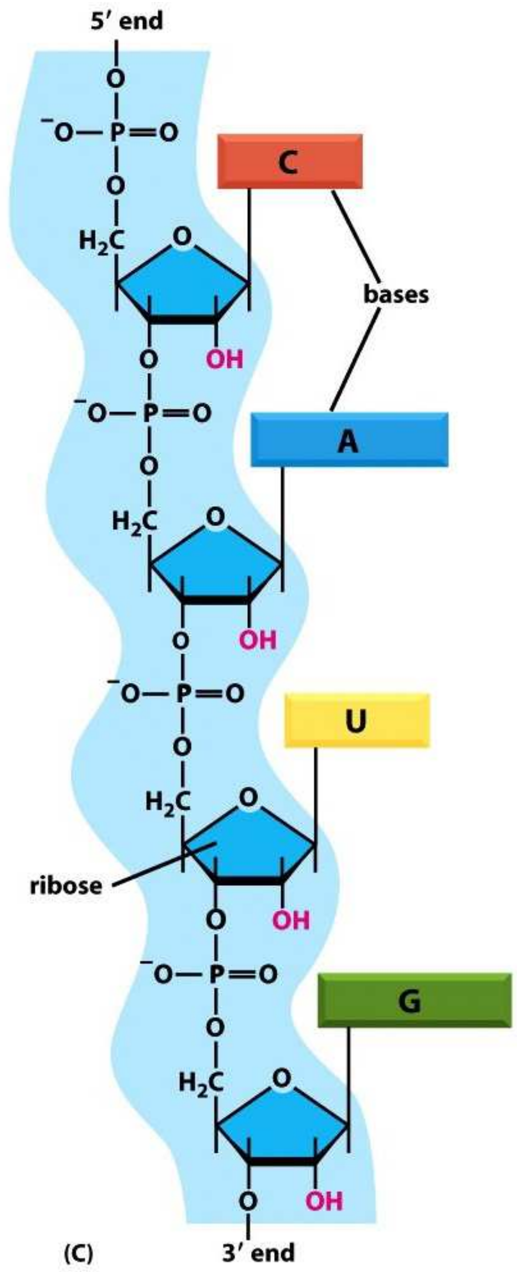
Information in RNA is passed to proteins

- RNA structure

RNA differs in several respects from DNA :

- ribose instead of deoxyribose
- Base uracil (U) instead of thymine (T)
- Single-stranded molecules





- To begin transcription, RNA polymerase binds to specific DNA sites called promoters that lie immediately upstream of genes

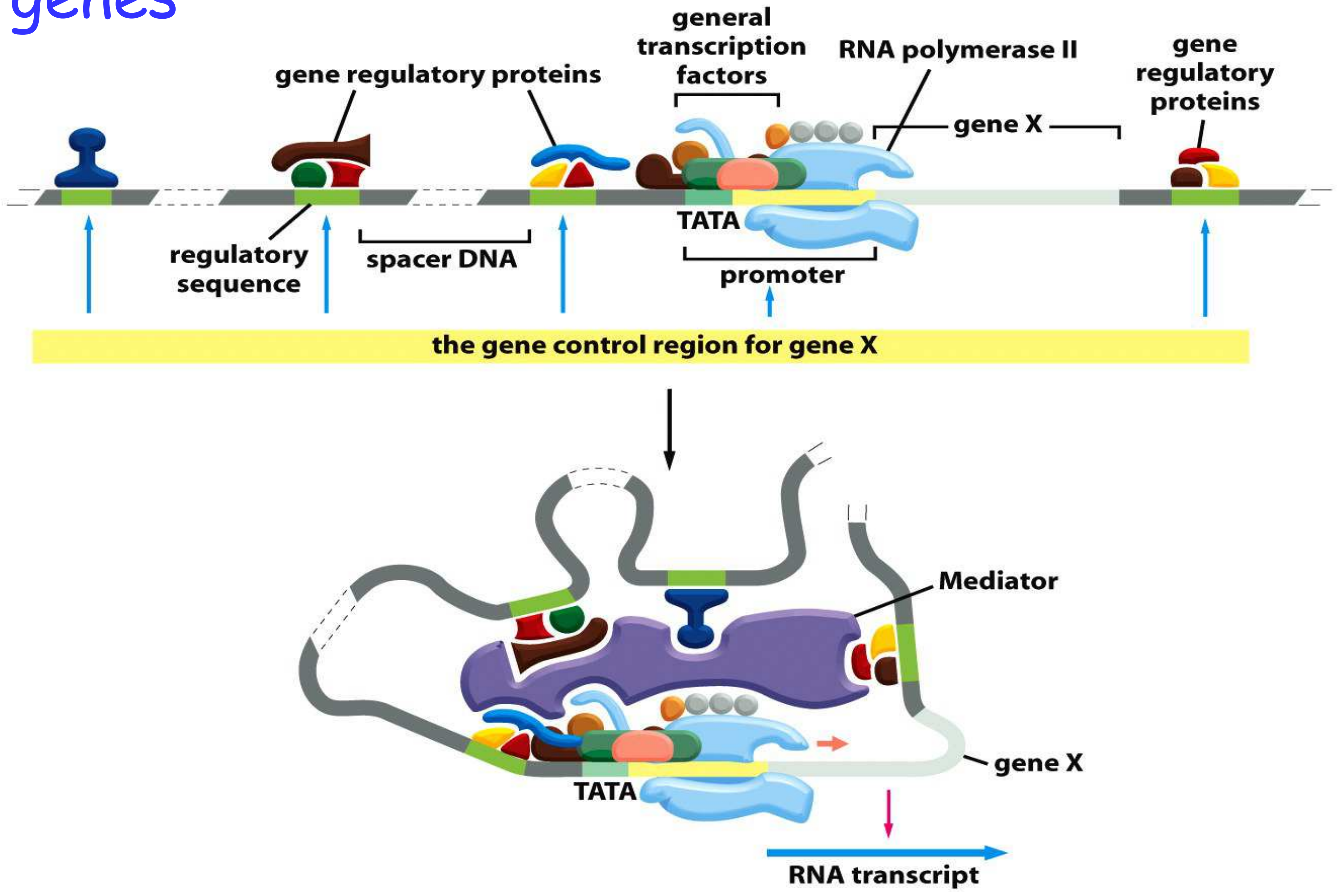


Figure 7-44 Molecular Biology of the Cell 5/e (© Garland Science 2008)

- Gene expression is tightly regulated by proteins that bind to regulatory sites and influence the capability of the transcription machinery to initiate RNA synthesis

- Most protein-coding genes in eukaryotic cells are composed of a number of coding regions (exons) interspersed with noncoding regions (introns)

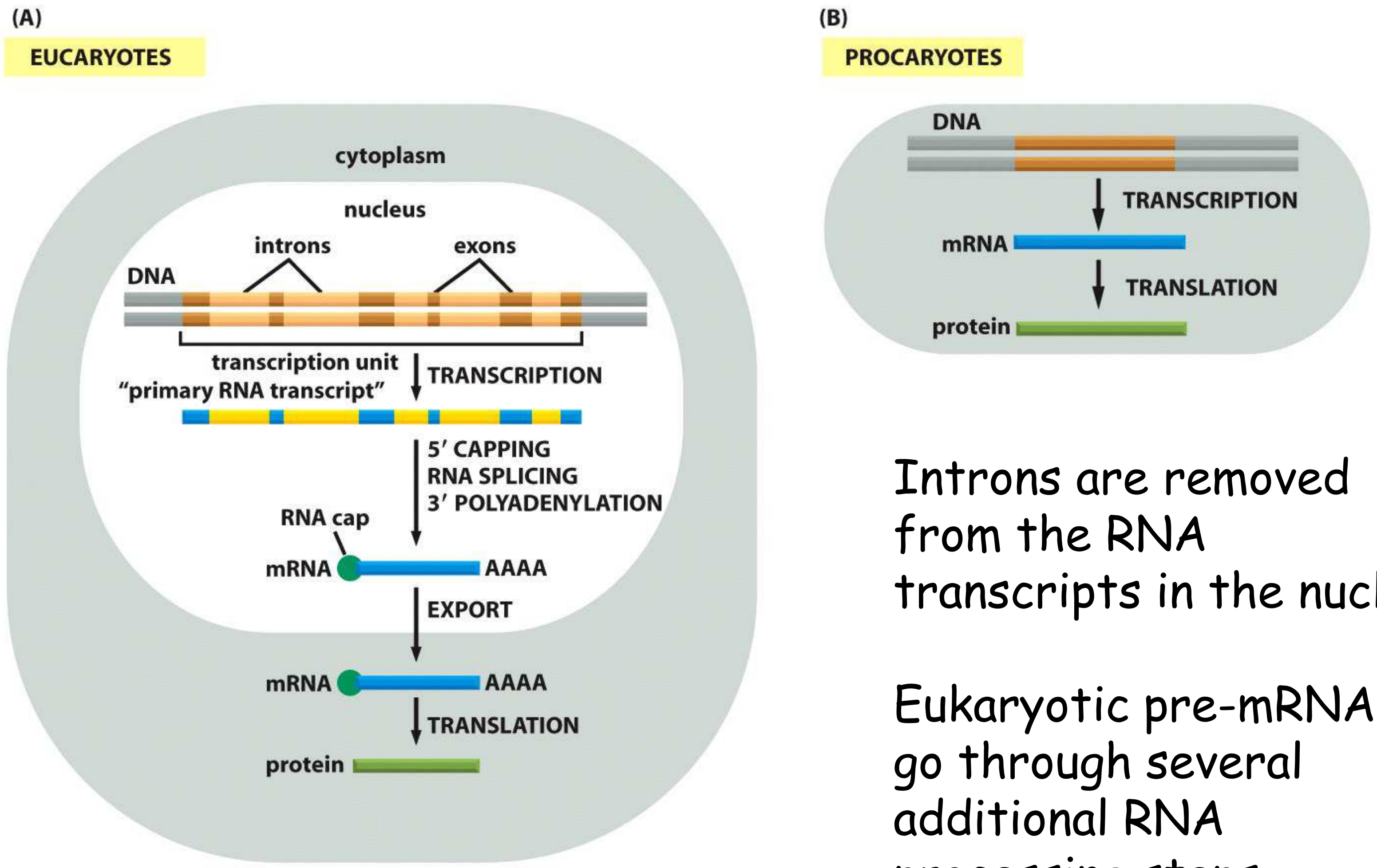


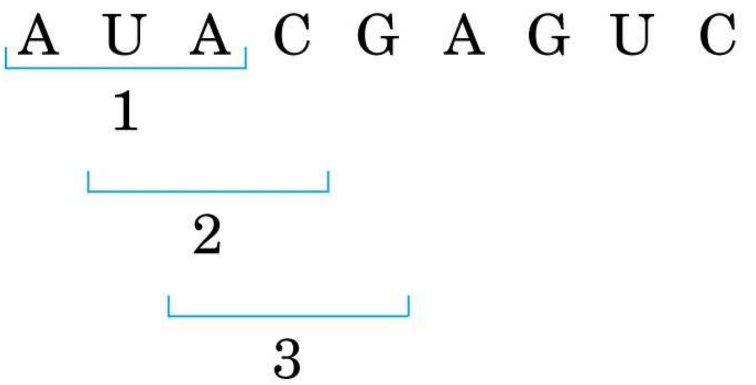
Figure 6-21 Molecular Biology of the Cell 5/e (© Garland Science 2008)

Introns are removed from the RNA transcripts in the nucleus

Eukaryotic pre-mRNAs go through several additional RNA processing steps

- The nucleotide sequence in mRNA is read in sets of three nucleotides called codons
- Each codon corresponds to one amino acid (nonoverlapping code)

In a nonoverlapping code codons do not share nucleotides
AUA–CGA–GUC



In an overlapping code, some nucleotides are shared by different codons
AUA – UAC – ACG – CGA

Note :

– mRNA is translated in 5’→3’direction
– No punctuation

- The correspondence between amino acids and codons is specified by the genetic code

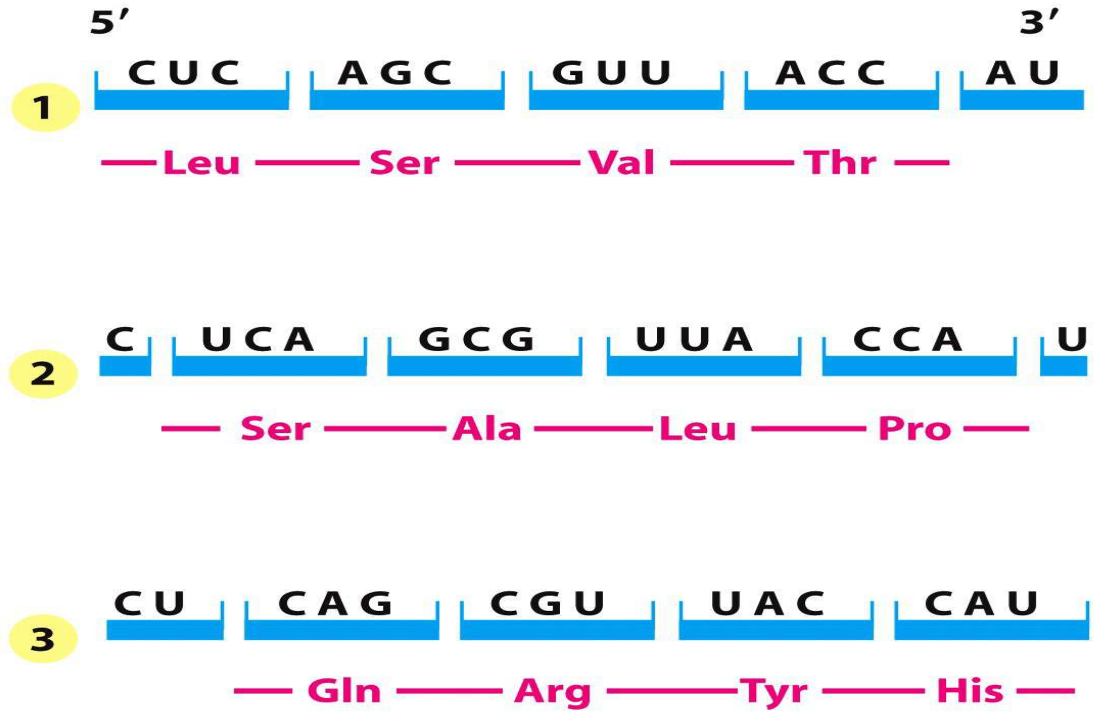
There are 64 different codons, including 3 stop codons

Most amino acids are specified by more than one codon

Correspondence is determined by tRNA molecules bearing a complementary anticodon and synthetases that link amino acids to their appropriate tRNAs

First position (5′ end)	Second position				Third position (3′ end)
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stop	Stop	A
	Leu	Ser	Stop	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

- In principle, an mRNA molecule can be translated in three possible reading frames



- But protein synthesis begins when a ribosome assembles at an initiation codon (AUG); the first one relative to the 5′ end of the mRNA (eukaryotic cells) or the closest one to the ribosome binding site (RBS in prokaryotic cells)
- Synthesis stops at a stop codon (UAA, UAG, UGA)

- The sequence of nucleotides in a DNA molecule is read from the 5′ to the 3′ end in sequential sets of 3 nucleotides on the possible reading frames (1-3).
- Putative ORFs are bordered by stop codons at both sides
- The longest ORF is assumed to encode the actual message and is translated

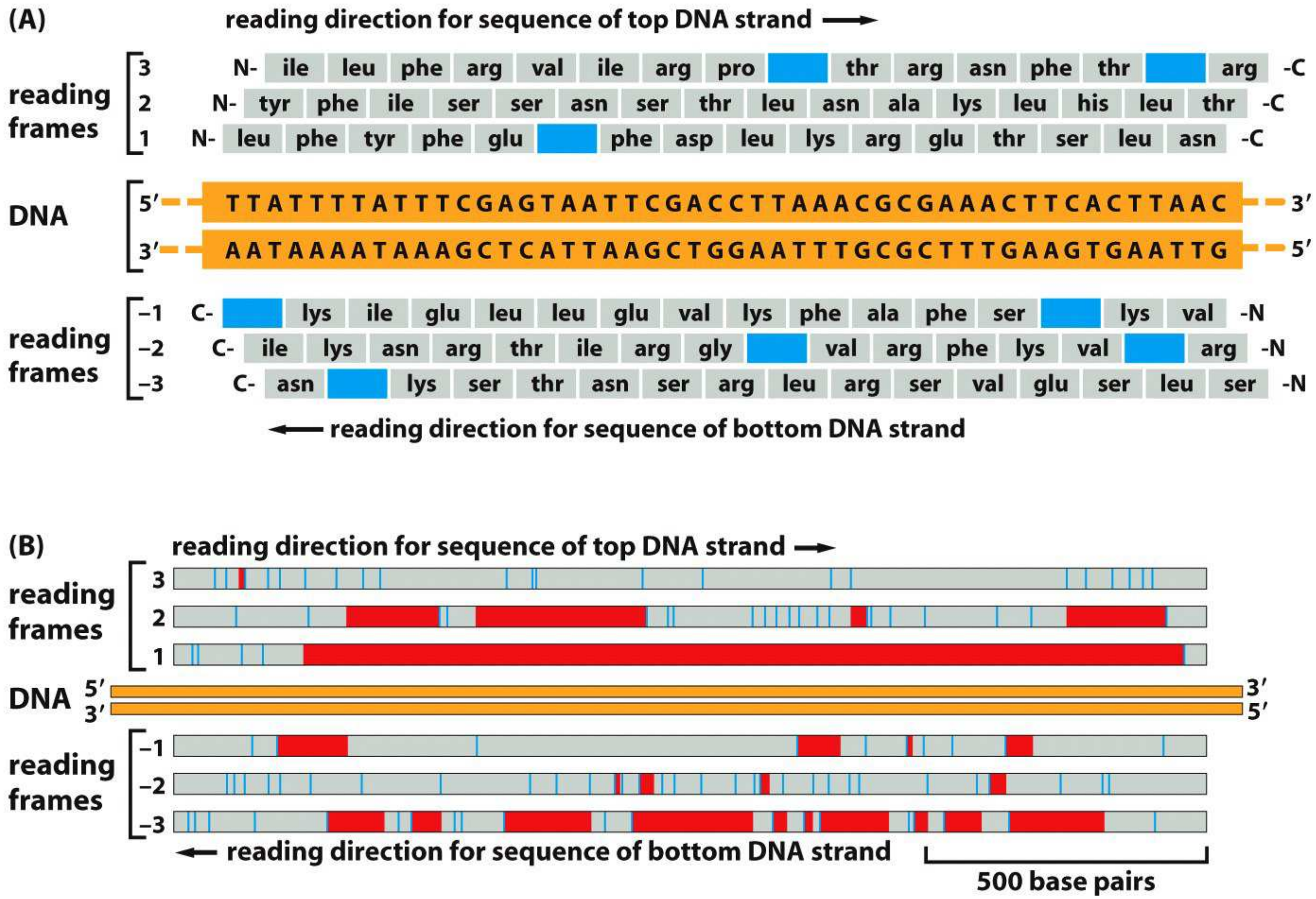


Figure 8-52 Molecular Biology of the Cell 5/e (© Garland Science 2008)

DNA sequencing methodology

(A) Use of dideoxyribonucleosides

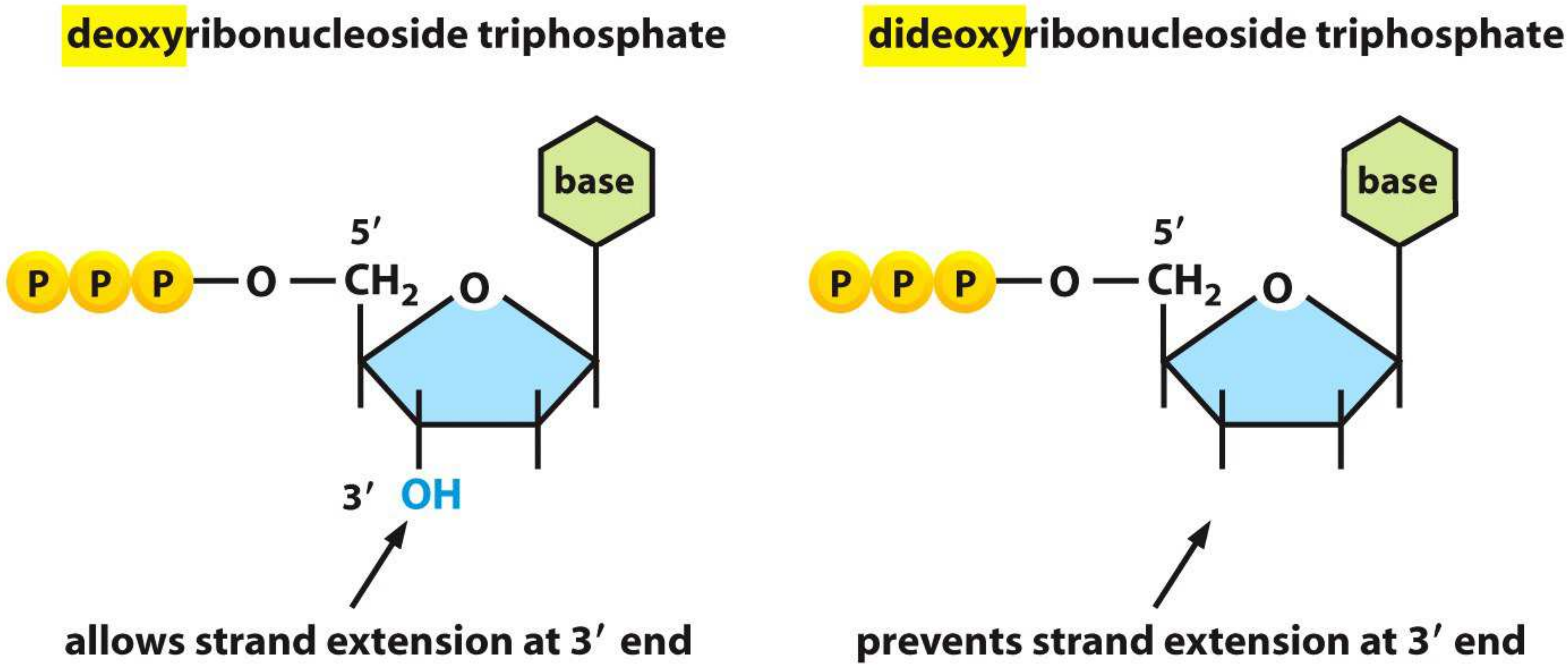


Figure 8-50a Molecular Biology of the Cell 5/e (© Garland Science 2008)

(B) *In vitro* DNA synthesis

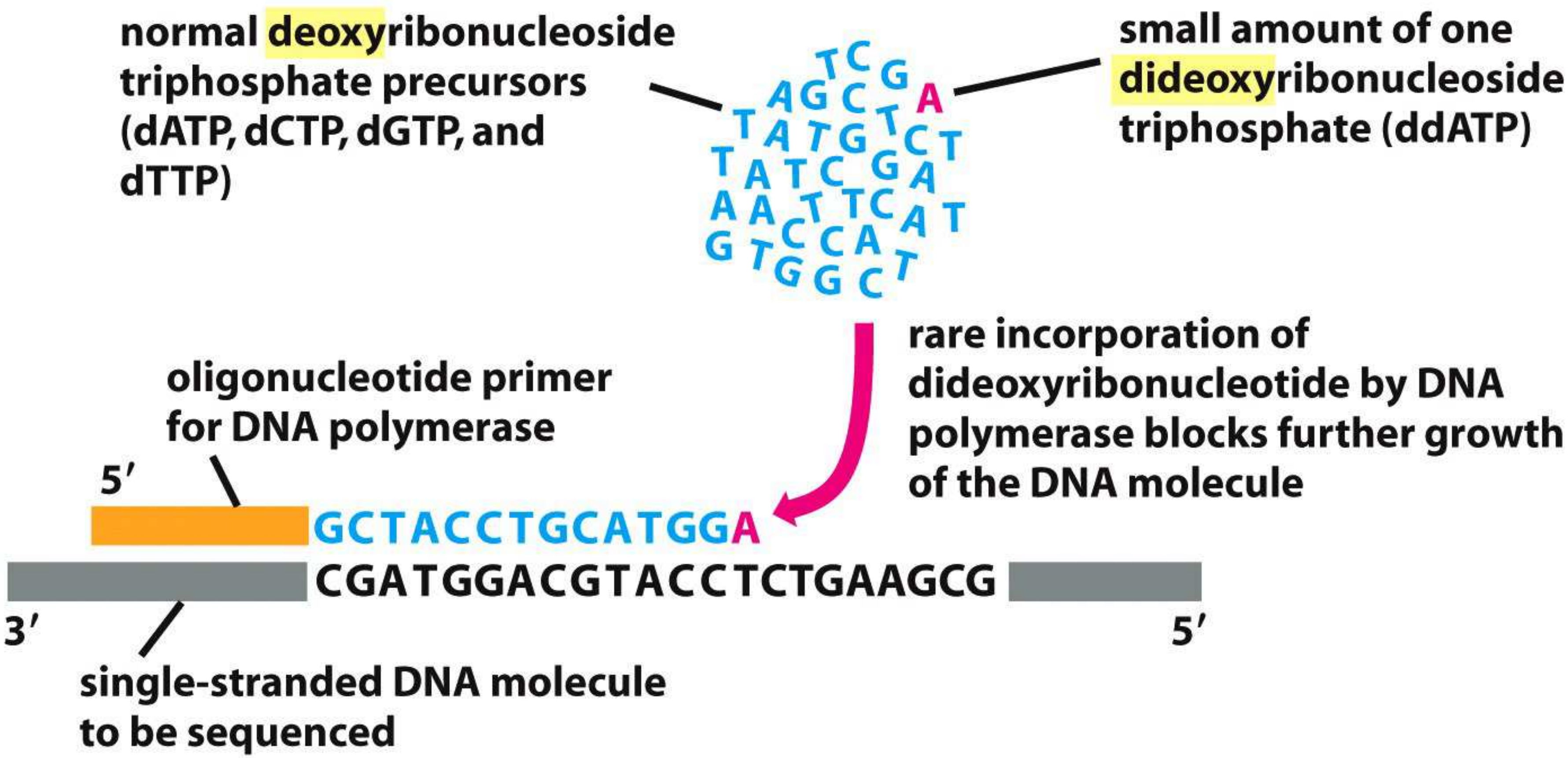


Figure 8-50b Molecular Biology of the Cell 5/e (© Garland Science 2008)

c) Experimental scheme

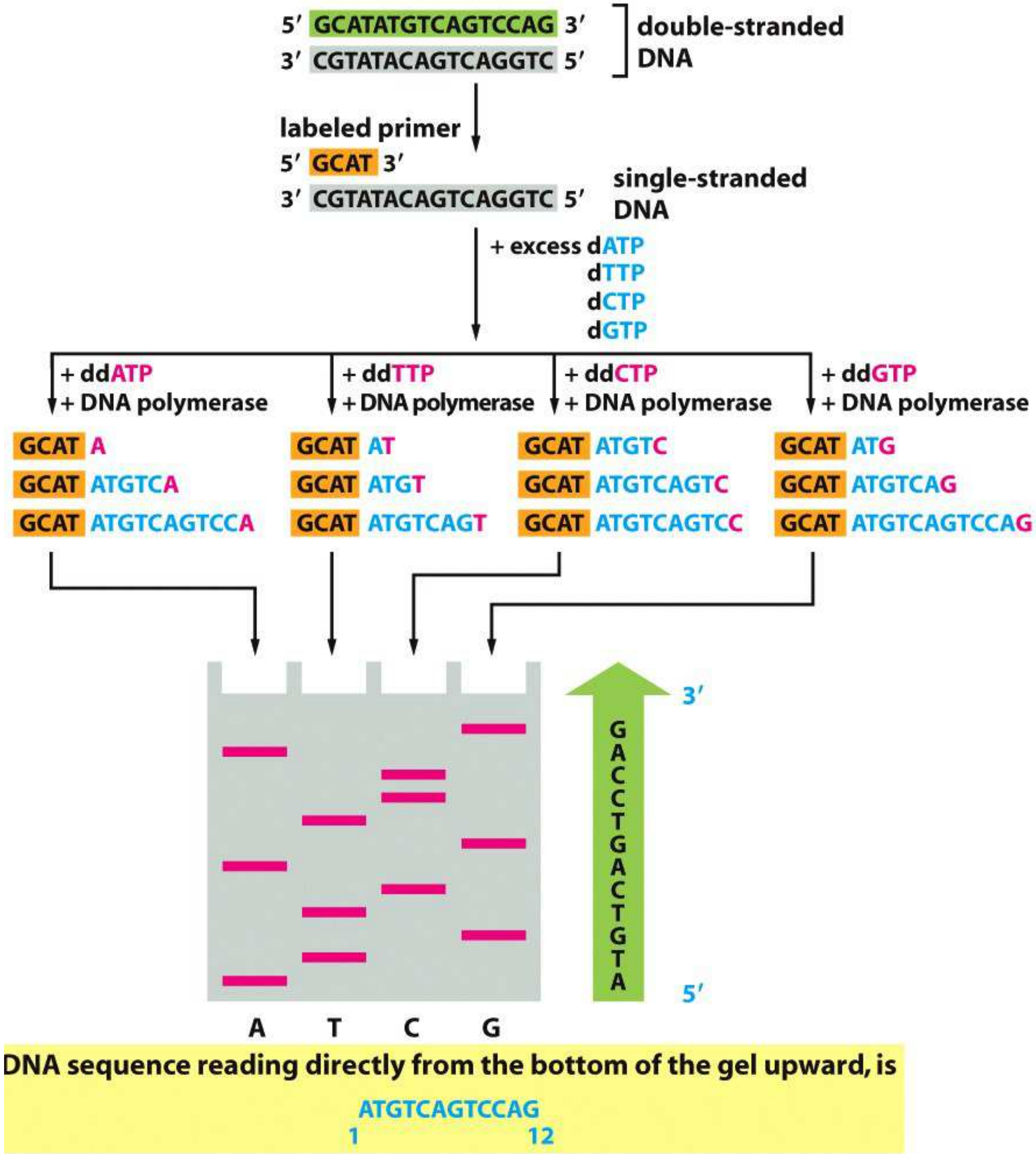


Figure 8-50c Molecular Biology of the Cell 5/e (© Garland Science 2008)

- Improvement using dyes and capillary gel electrophoresis