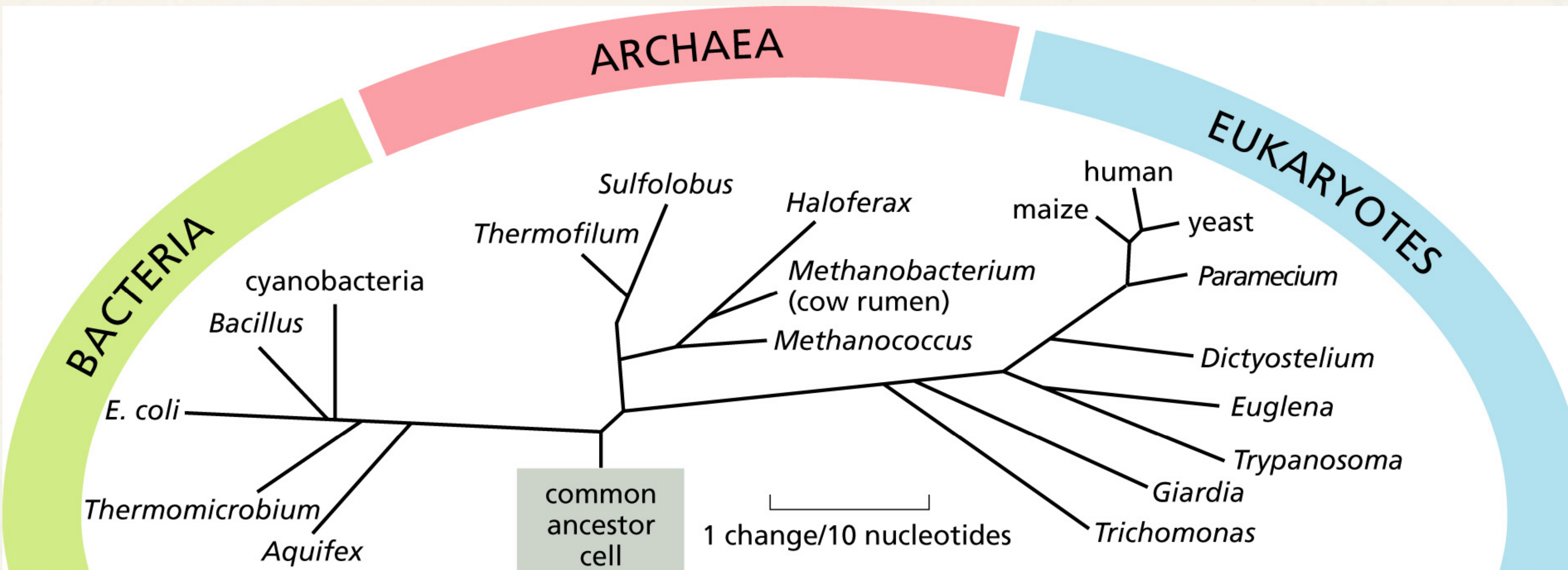
Bioinformatics: Network Analysis

Molecular Cell Biology: A Brief Review

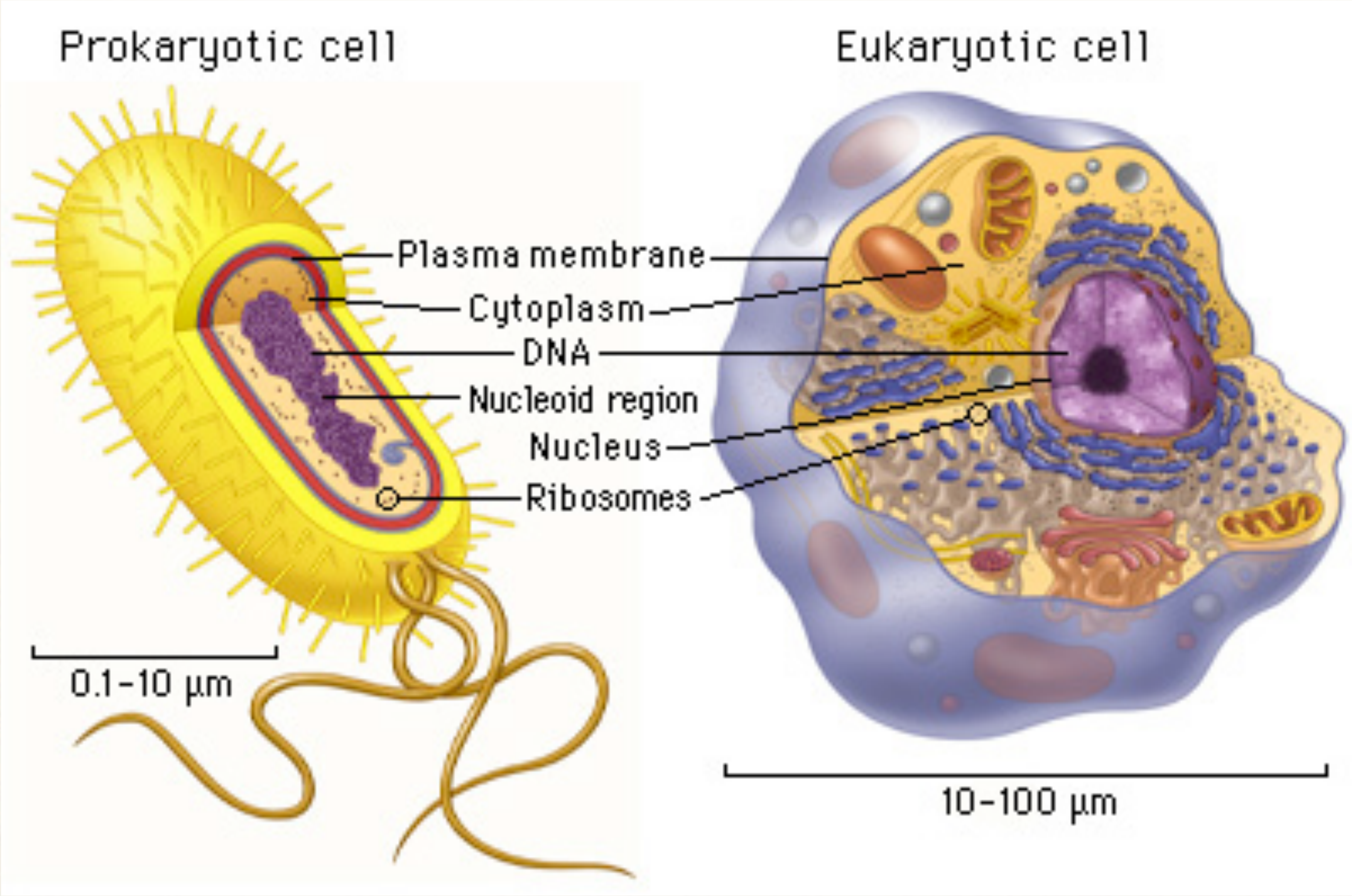
COMP 572 (BIOS 572 / BIOE 564) - Fall 2013

Luay Nakhleh, Rice University

The Tree of Life



Prokaryotic vs. Eukaryotic Cell Structure



Source: Pearson Education, Inc. The Biology Place

Prokaryotic vs. Eukaryotic Cells

	Prokaryotes	Eukaryotes
Size	1-10 Mm in length	10-100 Mm in length
Nucleus	does not exist	exists, and separated from the cytoplasm
Intracellular organization	no compartments	compartments (nucleus, cytosol, mitochondria, etc.)
Gene structure	no introns	introns and exons
Cell division	simple cell division	mitosis or meiosis
Ribosome	consists of a large 50S subunit and a small 30S subunit	consists of a large 60S subunit and a small 40S subunit
Reproduction	parasexual recombination	sexual recombination
Organization	mostly single cellular	mostly multicellular, and with cell differentiation

Source: Systems Biology in Practice, Klipp et al.

The Nucleic Acid World

- The full diversity of life on this planet—from the simplest bacterium to the largest mammal—is captured in a linear code inside all living cells.

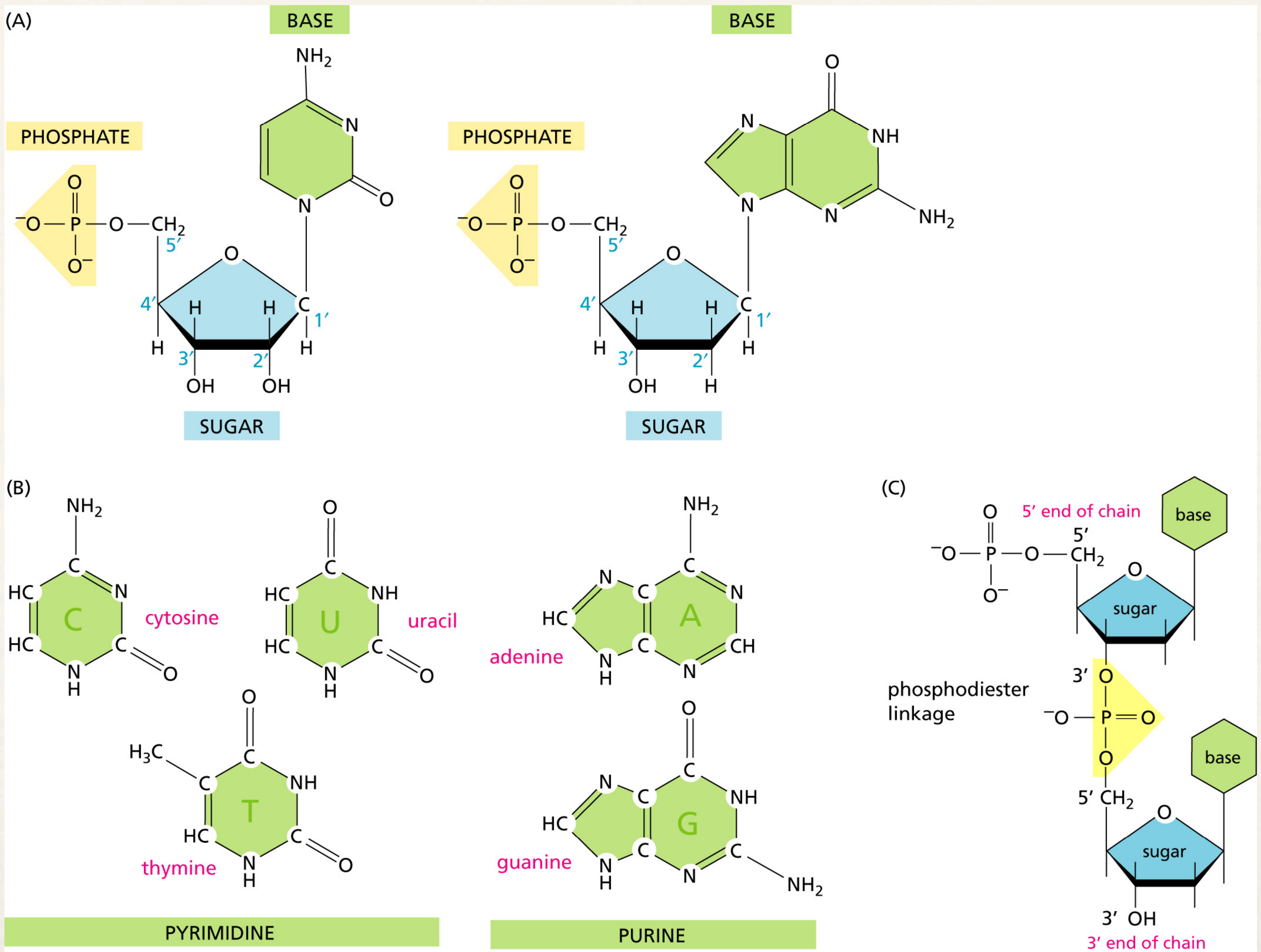
DNA

- Deoxyribonucleic Acid
- DNA molecules are linear polymers of just four different nucleotide building blocks.
- Genomic DNA molecules are immensely long, containing millions of bases each, and it is the order of these bases, the nucleotide sequence or base sequence of DNA, which encodes the information for making proteins.

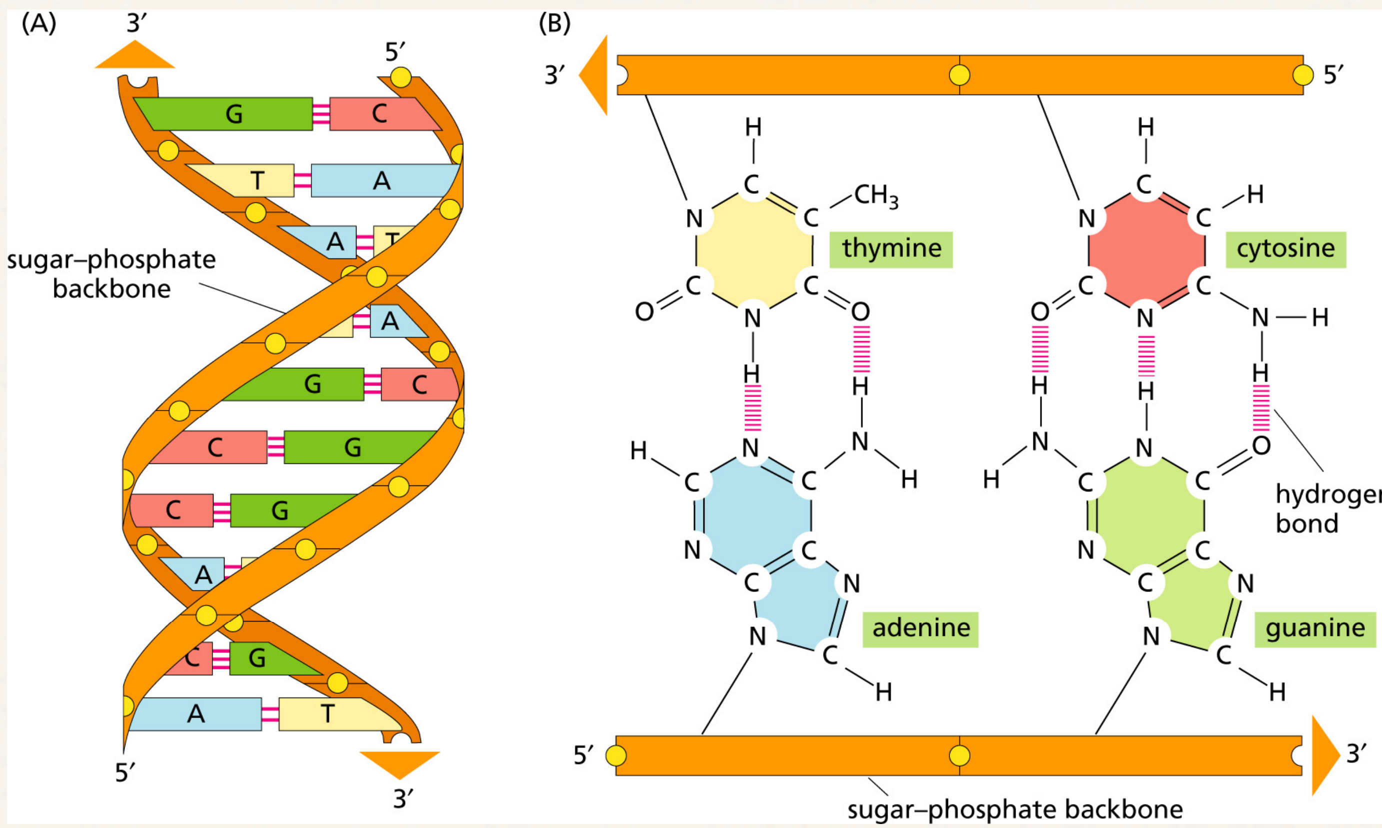
RNA

- Ribonucleic Acid
- RNA molecules are also linear polymers, but are much smaller than genomic DNA.
- Most RNA molecules also contain just four different base types.
- Several classes of RNA molecules are known, some of which have a small proportion of other bases.

The Building Blocks of DNA and RNA



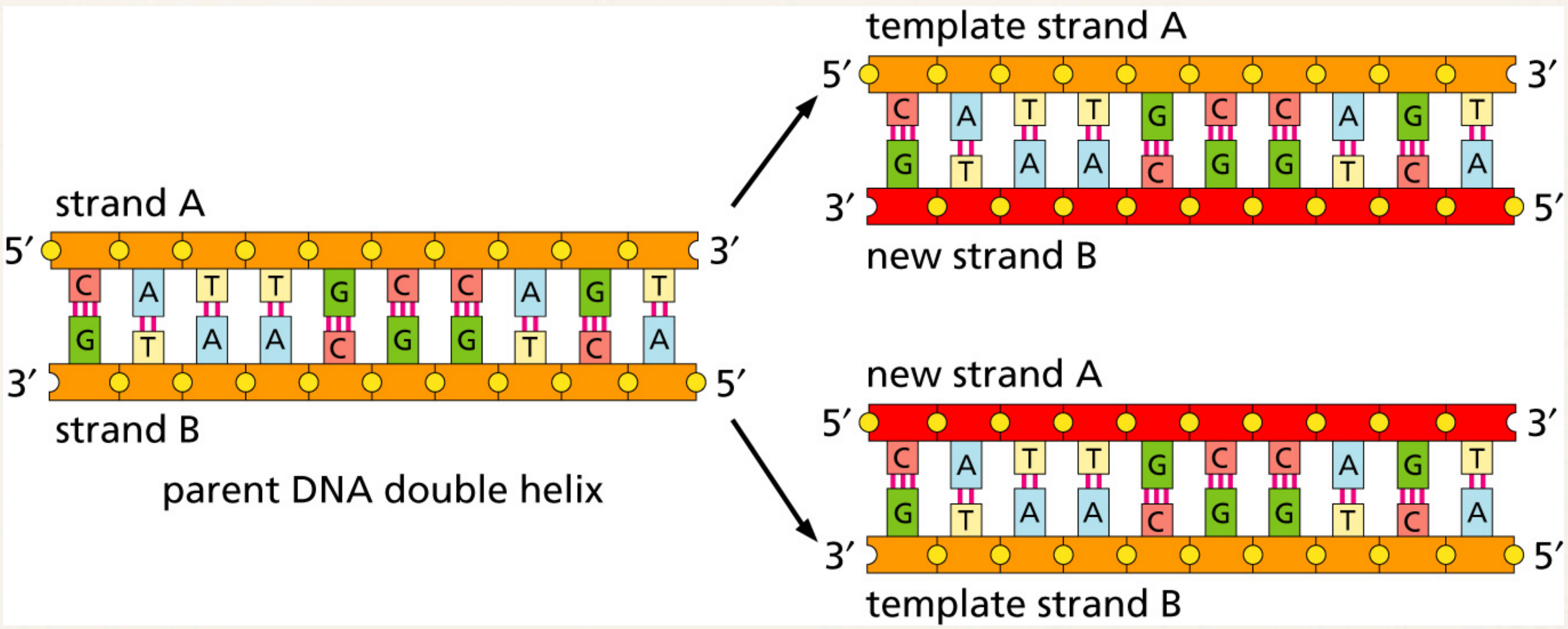
The Double Helix (DNA)



Watson-Crick base-pairing: A—T, C—G

Each strand of a DNA double helix has a base sequence that Is complementary to the base sequence of its partner strand.

DNA Replication

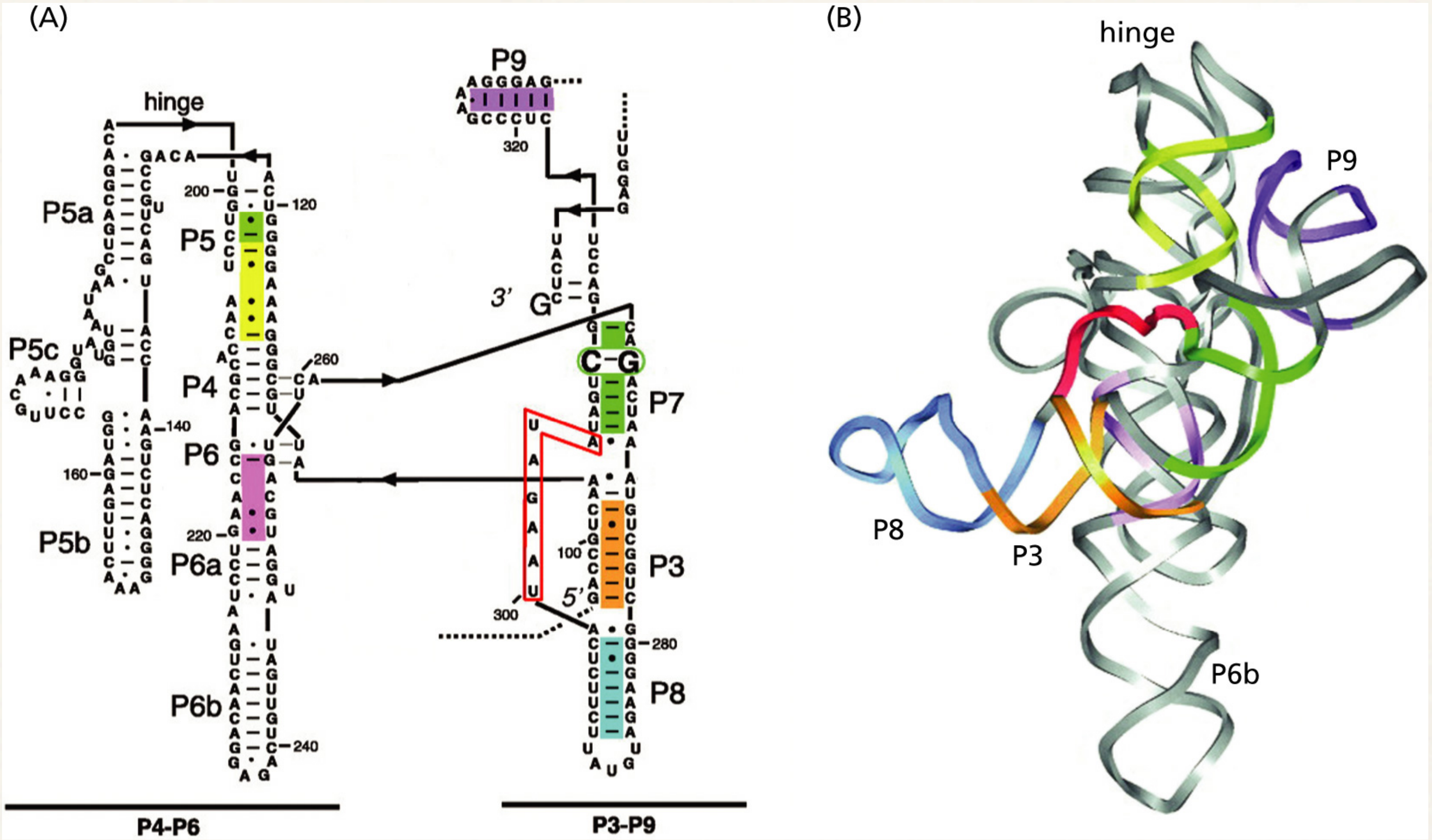


* Hydrogen bonds are noncovalent bonds: the two DNA strands can be easily separated.
* There are a number of processes in which strand separation is required.
* One such process is DNA replication, which is a necessary prelude to cell division.

RNA Structure

- Almost all RNA molecules in living systems are single stranded.
- As a result, RNA has much more structural flexibility than DNA, and some RNAs can even act as enzymes, catalyzing a particular chemical reaction.

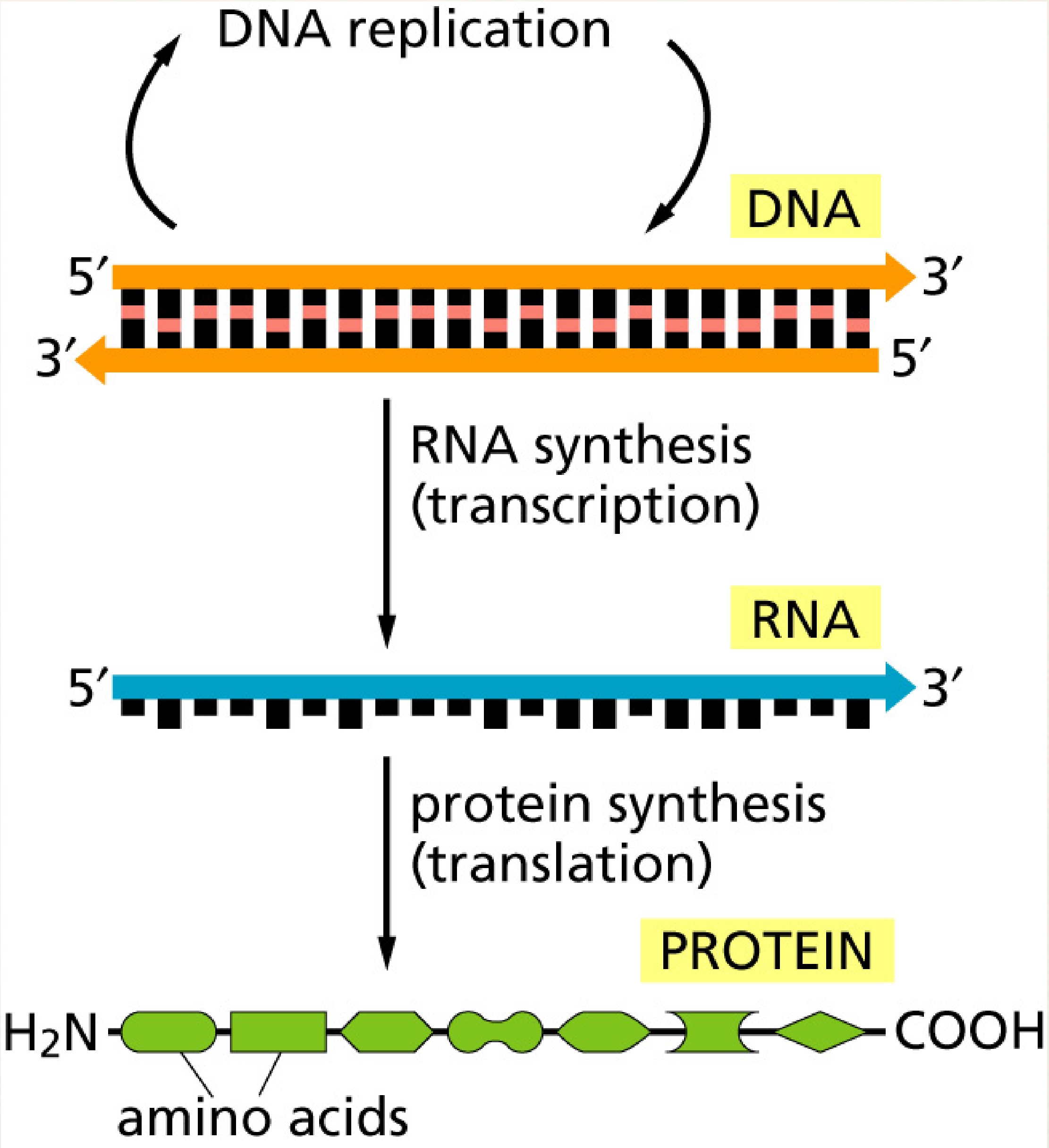
Secondary and Tertiary Structures of RNA



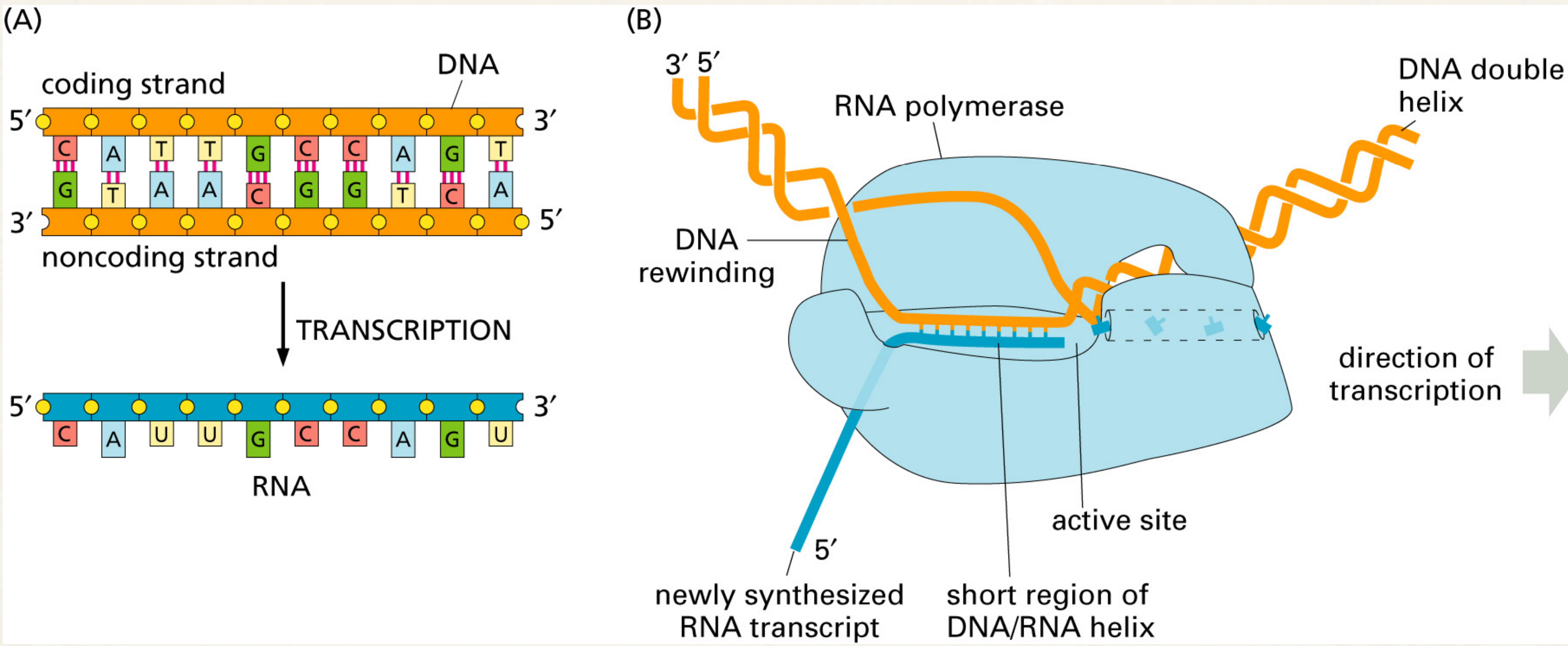
The Tetrahymena ribozyme

The Central Dogma

- A single direction of flow of genetic information from the DNA (information store), through RNA, to proteins
- This scheme holds for all known forms of life, with variations in the details of the processes involved in different organisms
- Not all genetic information in the DNA encodes proteins
- RNA can also be the end product, and other regions of the genome have as yet no known function of product
- The genomic DNA encodes all molecules necessary for life, whether they are proteins or RNA or ...



Transcription



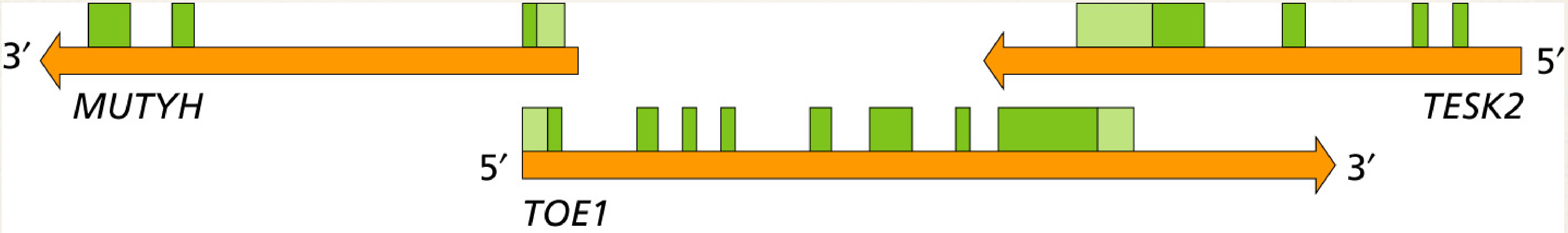
(A) One strand of the DNA is involved in the synthesis of an RNA strand complementary to the strand of the DNA

(B) The enzyme RNA polymerase reads the DNA and recruits the correct building blocks of RNA to string them together based on the DNA code

Terminology

- RNA transcribed from a protein-coding gene is called messenger RNA (mRNA)
- When a gene is being transcribed into RNA, the gene is said to be expressed

Overlapping Genes



Although only one segment of the DNA strand is transcribed for any given gene, it is also possible for genes to overlap so that one or both strands at the same location (locus) encode parts of different proteins.

This most commonly occurs in viruses as a means of packing as much information as possible into their very small genomes but it could also occur in mammals (the above figure shows overlapping genes in the human genome)

Regulated Gene Expression

- The genomic DNA sequence contains more information that just the protein sequences. The transcriptional apparatus has to locate the sites where gene transcription should begin, and when to transcribe a given gene. At any one time, a cell is only expressing a few thousand of the genes in its genome. To accomplish this regulated gene expression, the DNA contains control sequences in addition to coding regions (More on this in a few slides).

Translation

- mRNA is translated into protein according to the genetic code, which is the set of rules governing the correspondence of the base sequences in DNA or RNA to the amino acid sequence of a protein.

- Each amino acid is encoded by a set of three consecutive bases (codon)

The Standard Genetic Code

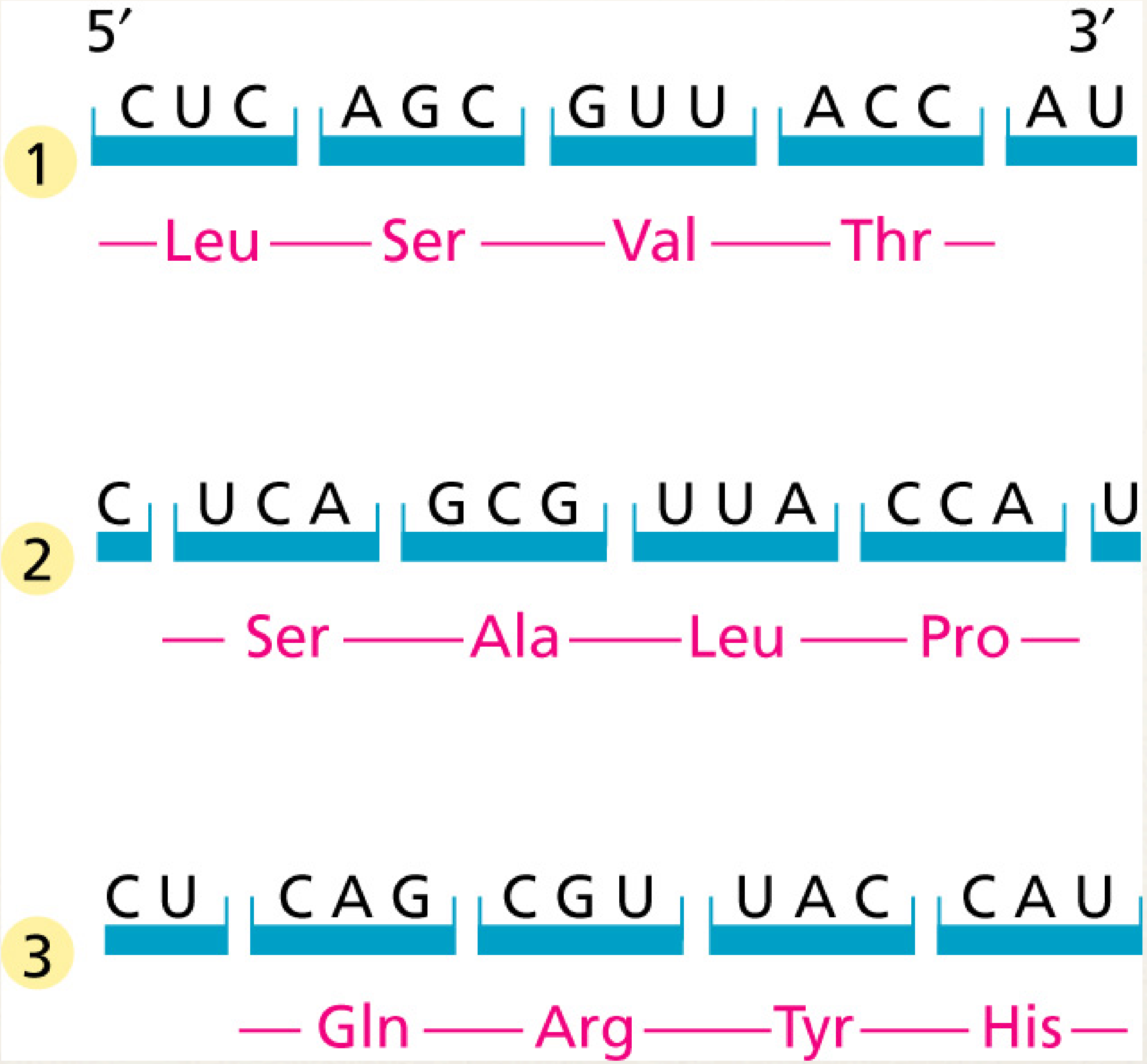
Second letter

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

Reading Frames

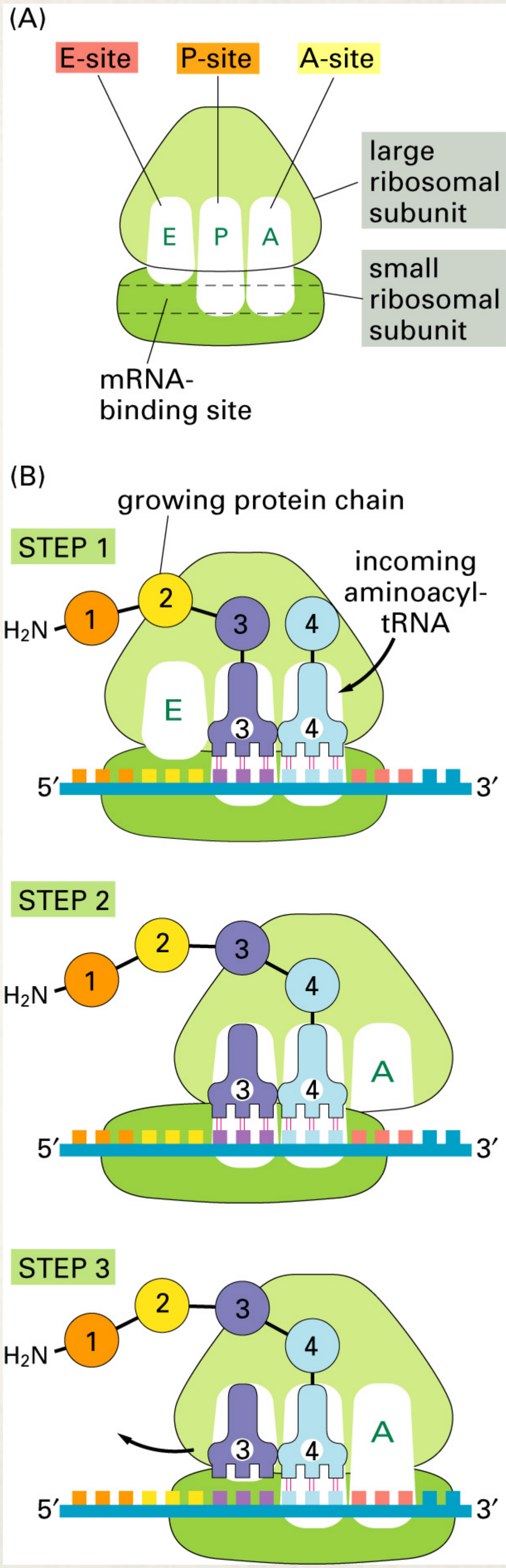
- Translation occurs in nonoverlapping sets of three bases.
- There are thus three possible ways to translate any nucleotide sequence, each of which is called a reading frame
- These three reading frames give three different protein sequences.
- In the actual translation process, the detailed control signals ensure that only the appropriate reading frame is translated into protein.

Reading Frames



Translation Machinery

- There are three main classes of RNA in the cell: messenger RNA (mRNA), ribosomal RNA (rRNA), and transfer RNA (tRNA).
- rRNA and tRNA are involved in mRNA translation and protein synthesis



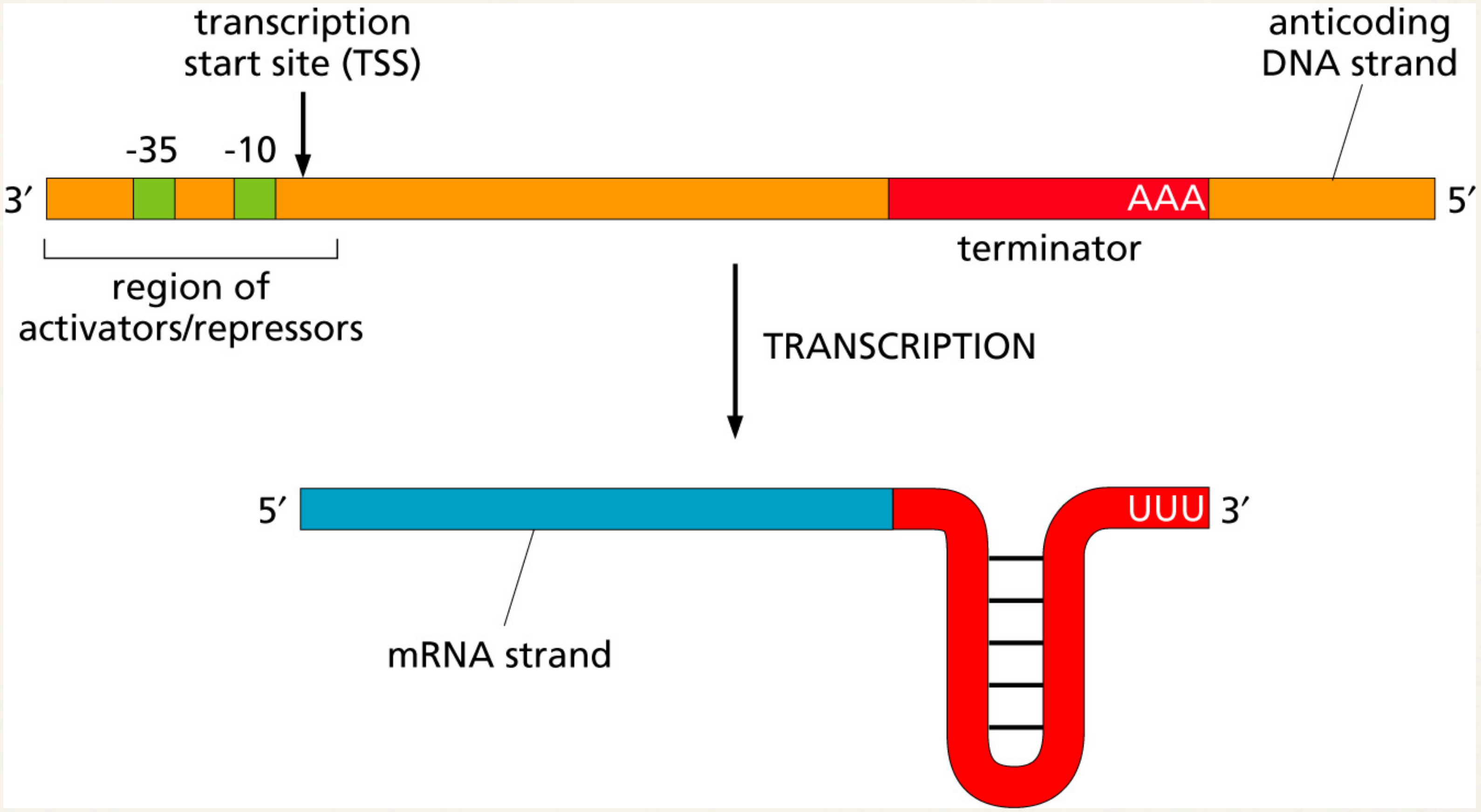
Gene Structure and Control

- The regulation of many processes that interpret the information contained in a DNA sequence relies on the presence of short signal sequences in the DNA.
- The general term for these signal sequences is regulatory elements.
- For example, the molecules involved in transcription and translation require signals to identify where they should start and stop.
- Gene structure and control differ between prokaryotes and eukaryotes

Transcription Regulation

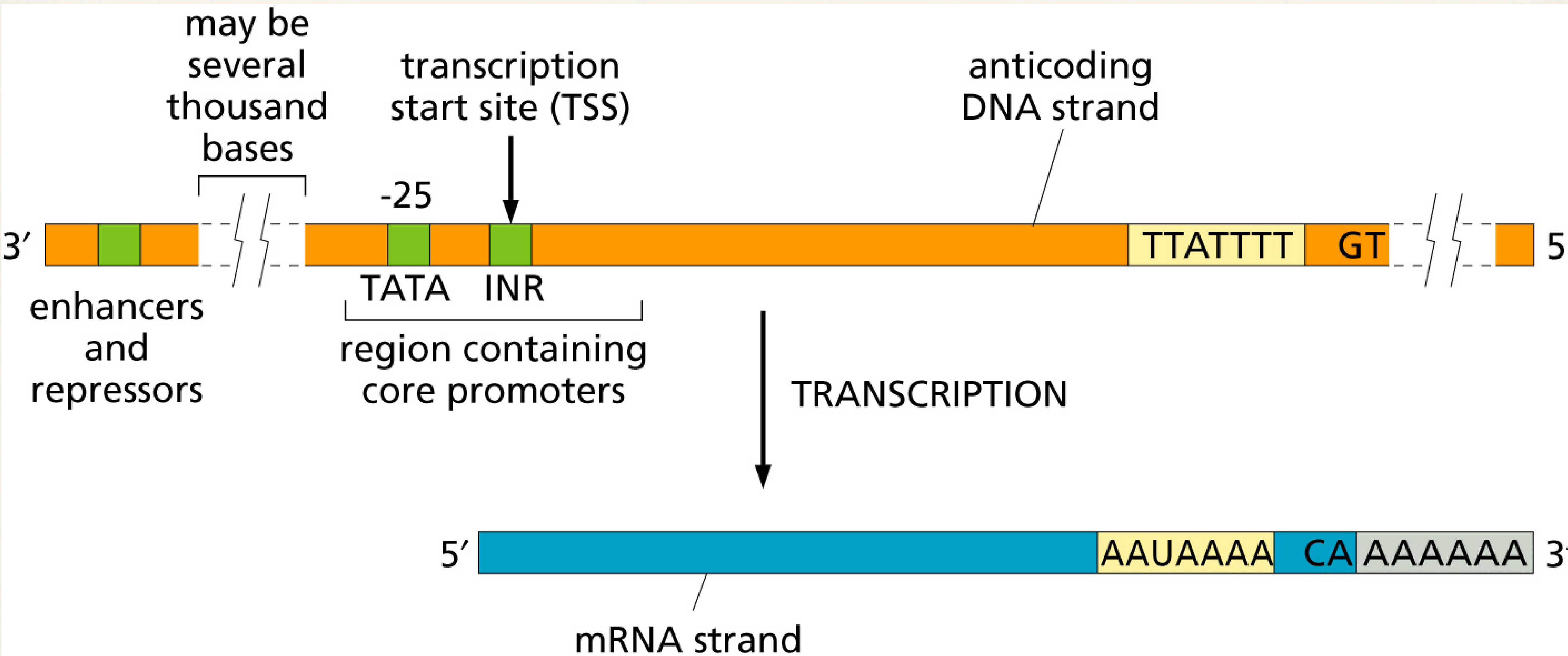
- The control regions at which RNA polymerase binds to initiate transcription are called promoters.
- RNA polymerase binds more tightly to these regions than to the rest of the DNA and this triggers the start of transcription.

Gene Structure in Prokaryotes



* Bacterial promoters typically occur immediately before the position of the transcription start site (TSS), and contain two characteristic short sequences, or motifs, that are almost the same in the promoters for different genes.
* The termination of transcription is controlled by the terminator signal which in bacteria differs from the promoter in that it is active when transcribed to form the end of the mRNA strand (forms a loop structure that prevents the transcription apparatus from continuing).
* Single type of RNA polymerase transcribes all genes.

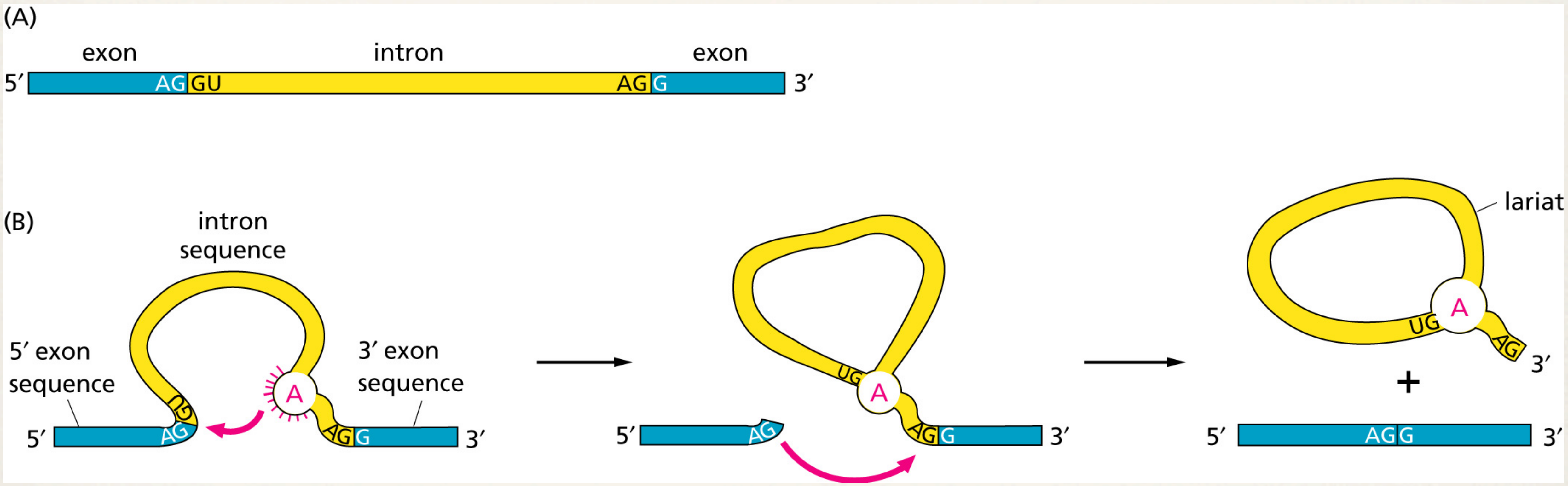
Gene Structure in Eukaryotes



* Regulatory elements in eukaryotes are more complex.

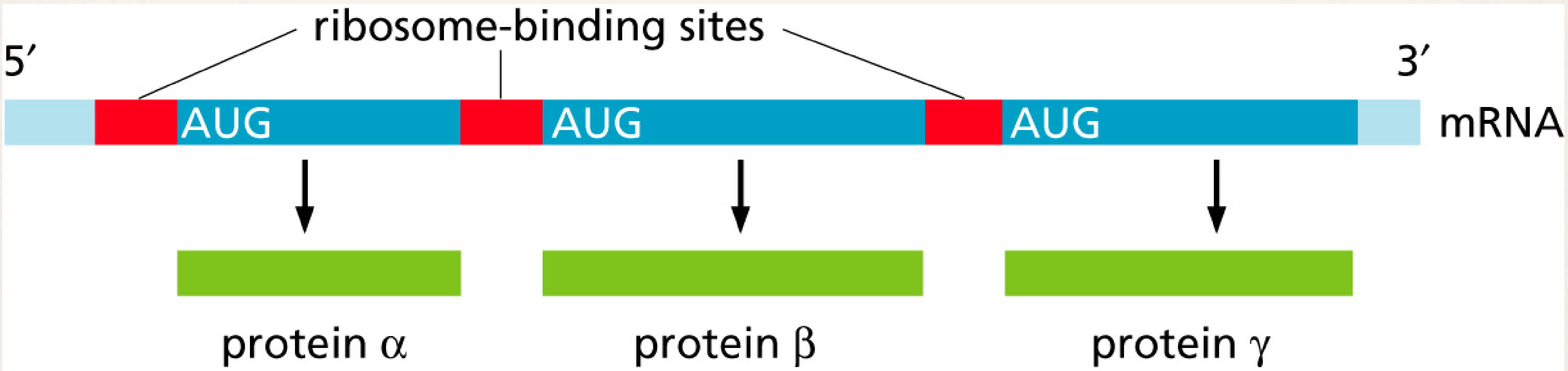
* Three types of RNA polymerase transcribe genes: RNA polymerase II transcribes all protein coding genes, where other RNA polymerase tpyes transcribe genes for tRNAs, rRNAs and other types of RNA

Splicing of an Intron



* The existence of introns necessitates an extra step between transcription and translation, which is known as RNA splicing: (1) the complete gene is initially transcribed into RNA, and (2) the introns are then excised and the exons spliced together to provide a functional mRNA that gives the correct protein sequence when translated. In most protein coding genes, this process is carried out by the spliceosome, which consists of small nuclear RNA (snRNA) and proteins.

Operon Structure

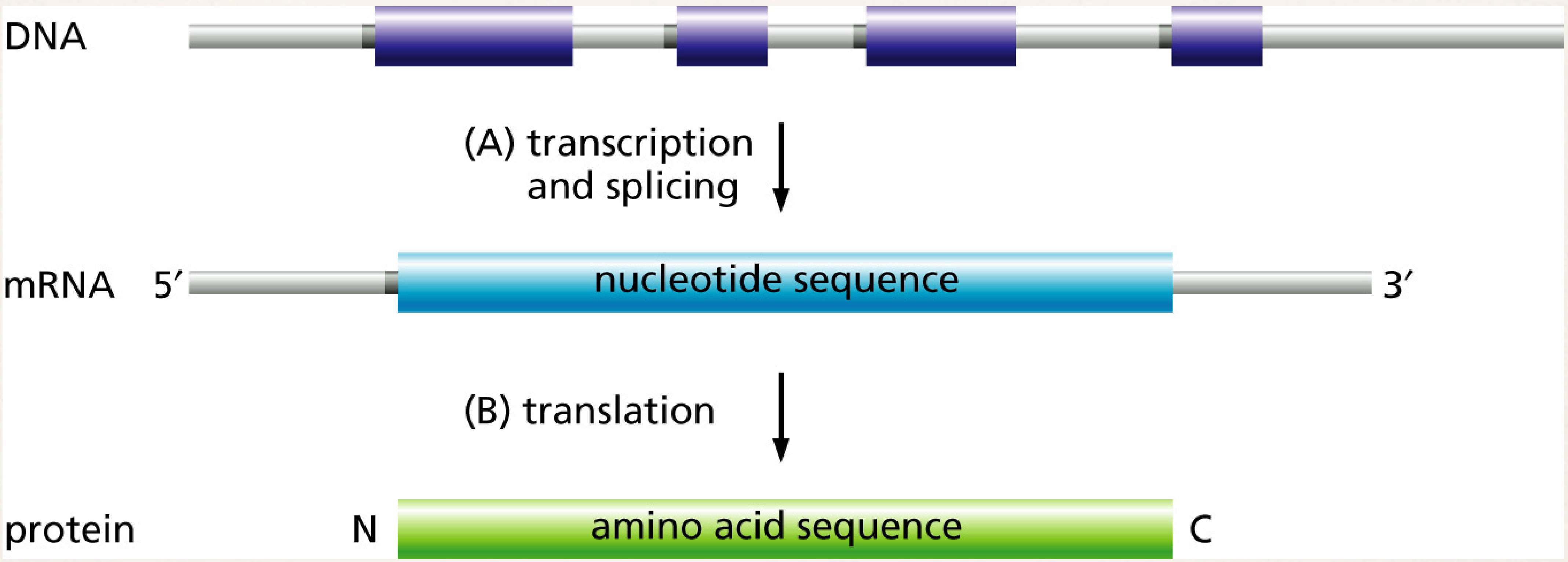


In bacteria, functionally related protein-coding sequences are often clustered together into operons. Each operon is transcribed as a single mRNA transcript and the proteins are then separately translated from this one long molecule.

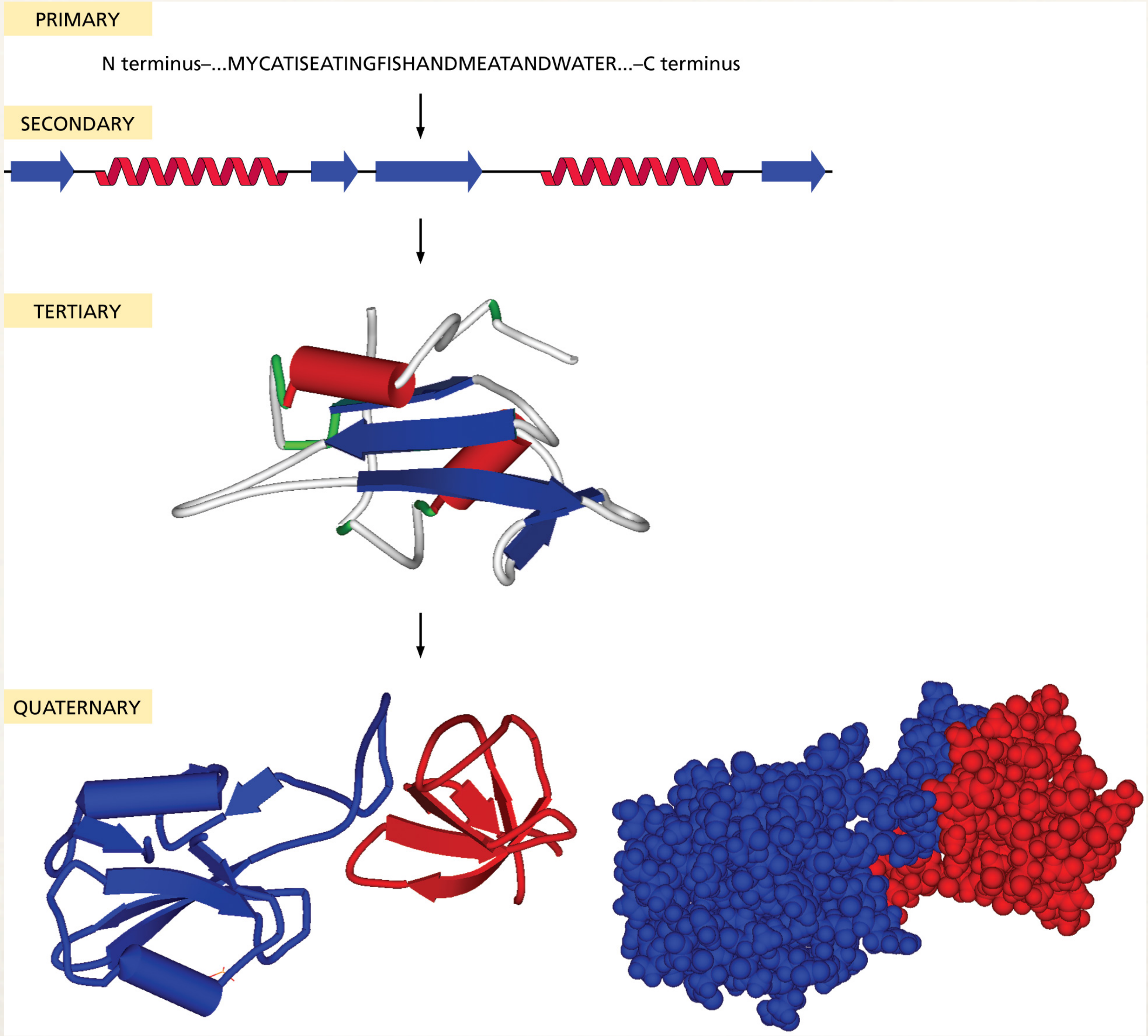
This has the advantage that only one control region is required to activate the simultaneous expression of all genes in the operon.

Not all bacterial genes are contained in operons; many are transcribed individually and have their own control regions.

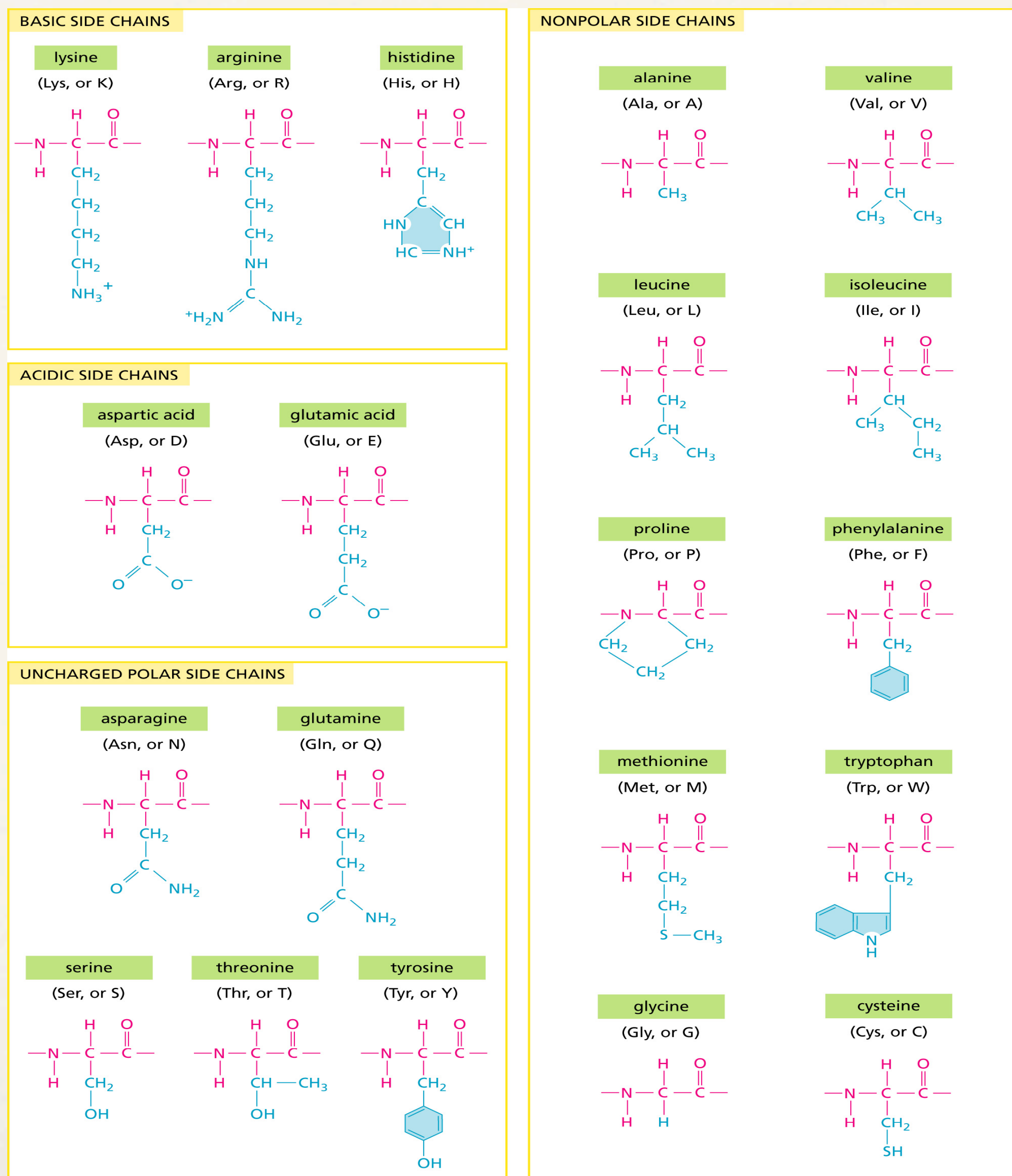
Proteins


DNA
(A) transcription
and splicing
mRNA
5′
nucleotide sequence
3′
(B) translation
protein
N
amino acid sequence
C

Levels of Protein Structure



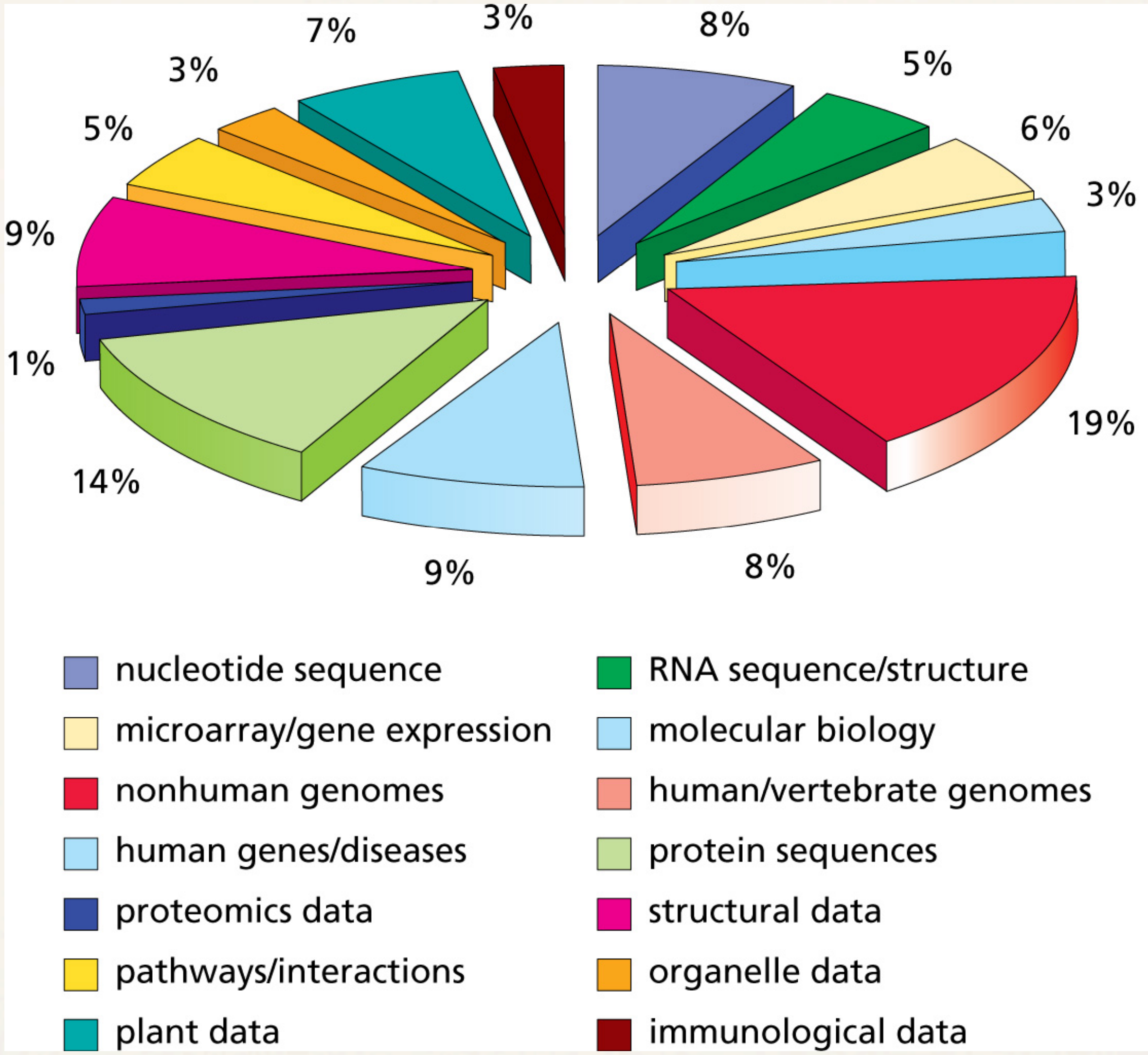
Side Chains of the Amino Acids



Organization of the DNA

- The genome is an organism's complete set of DNA
- Genomes vary widely in size
  - some bacteria have 600,000 base pairs
  - humans have about 3 billion base pairs
  - Except for mature red blood cells, all human cells contain a complete genome
- DNA in the human genome is arranged into 23 pairs of DNA molecules, called chromosomes (physically separate molecules, and vary widely in length)
- Each chromosome contains many genes

Available Data



858 Databases in total

(as classified at the NAR Molecular Biology Database Collection Website, 2006)

Acknowledgments

- Materials in this lecture are mostly based on the book “Understanding Bioinformatics”, by M. Zvelebil and J. Baum, 1st Edition, Garland Science, 2007.