

CAP 5510: Introduction to Bioinformatics
CGS 5166: Bioinformatics Tools

Giri Narasimhan

ECS 389; Phone: x3748

giri@cs.fiu.edu

www.cs.fiu.edu/~giri/teach/5166S05.html

Course Homepage

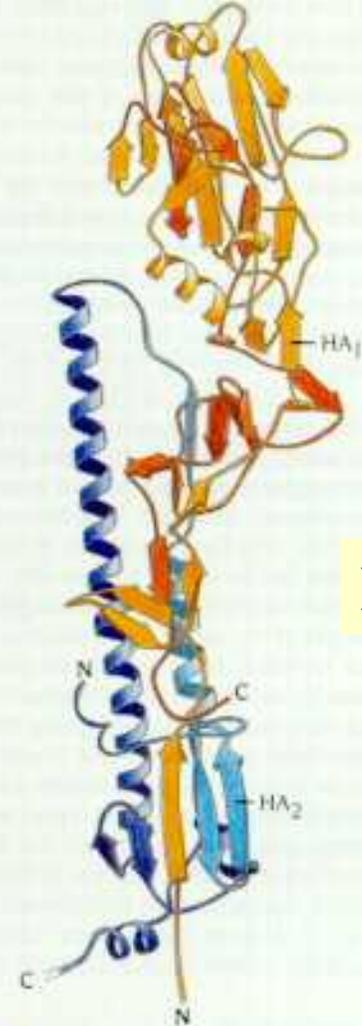
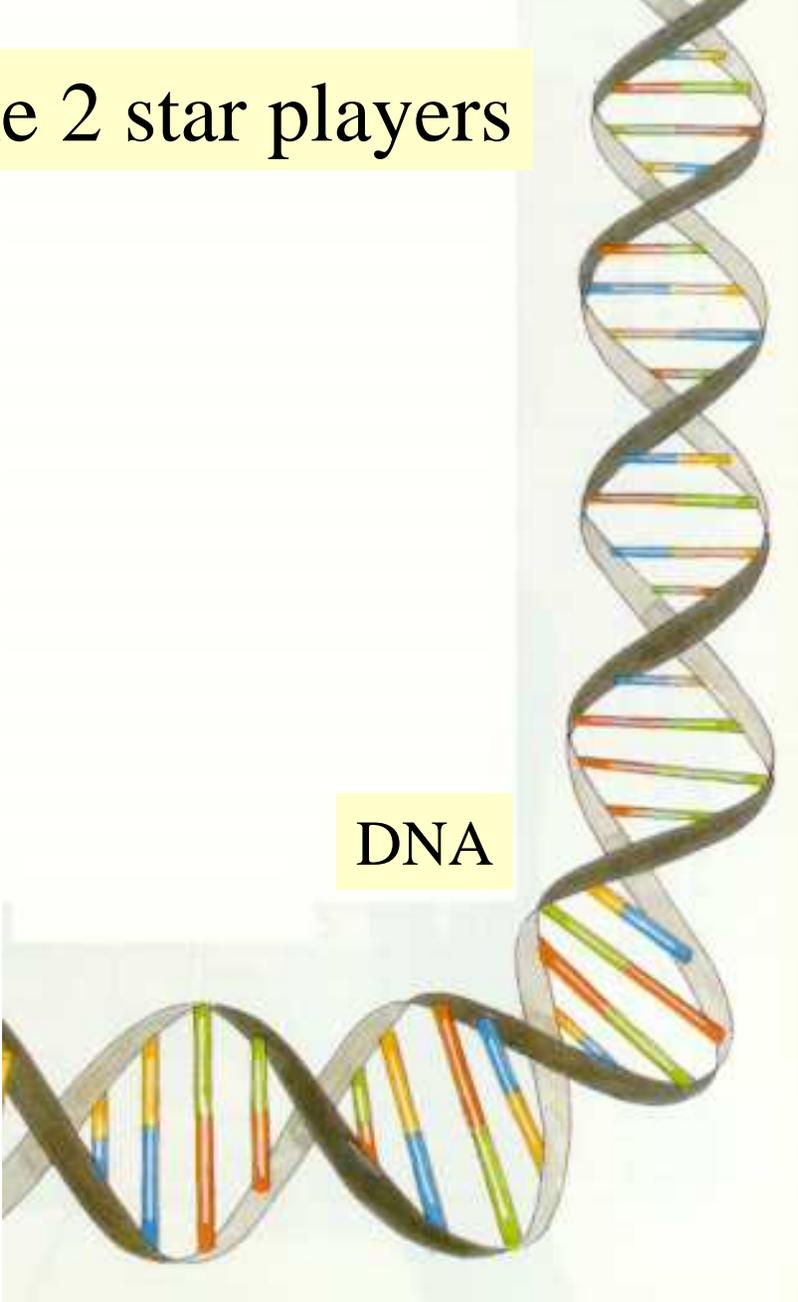
www.cs.fiu.edu/~giri/teach/5166S05.html

- Lecture notes, required reading material, homework, announcements, etc.

Molecular Biology Background

The 2 star players

DNA



Protein

Figure 8.21 Schematic diagram of the subunit structure of hemagglutinin from influenza virus. The structure comprises about 550 amino acids arranged in two chains HA₁ (red) and HA₂ (blue). The first half of each chain has a lighter color in the diagram. The subunit is very elongated with a long stemlike region built up by residues from both chains and includes one of the longest α helices known in a globular structure, about 75 Å long. The globular head is formed by residues only from HA₁. (Courtesy of Don Wiley, Harvard University.)

The Players

DNA

String with alphabet {A, C, G, T}

Nucleotides/Bases

RNA

String with alphabet {A, C, G, U} Bases

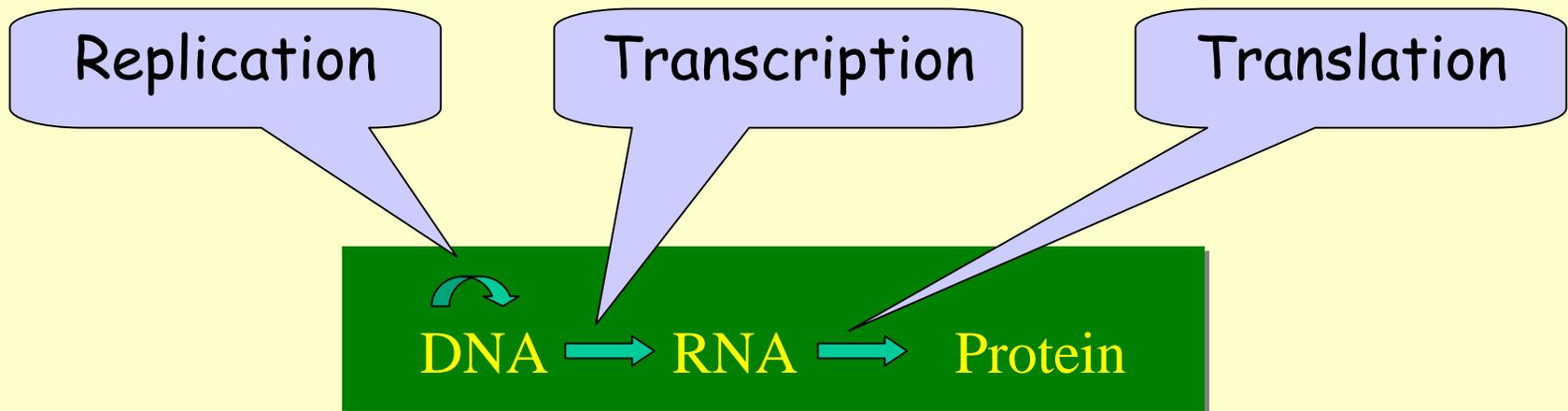
Protein

String with 20-letter alphabet

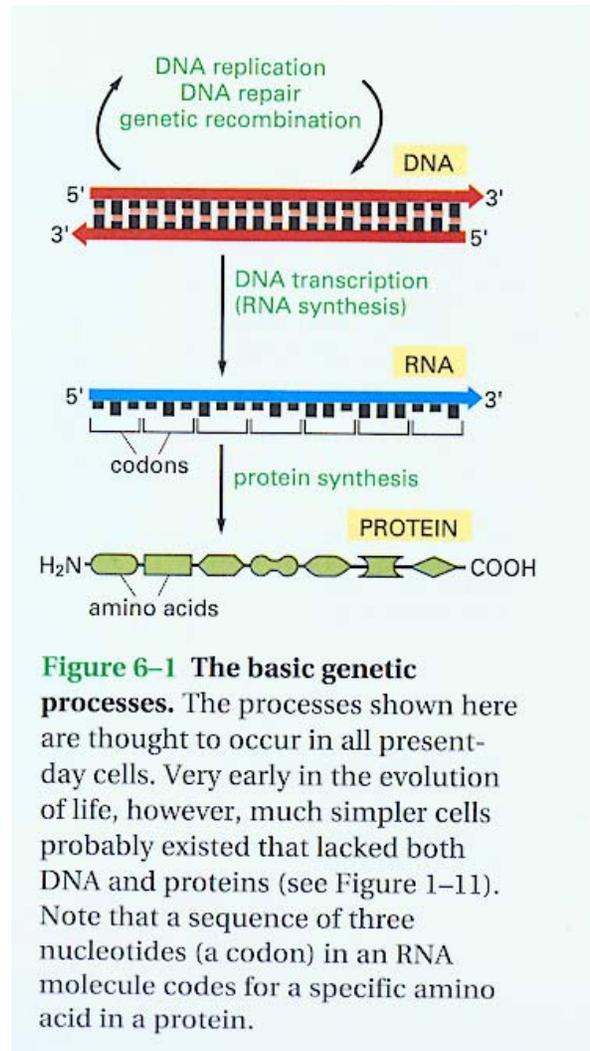
Amino acids/Residues

Central Dogma

- DNA acts as a template to replicate itself.
- DNA is transcribed into RNA.
- RNA is translated into **Protein**.

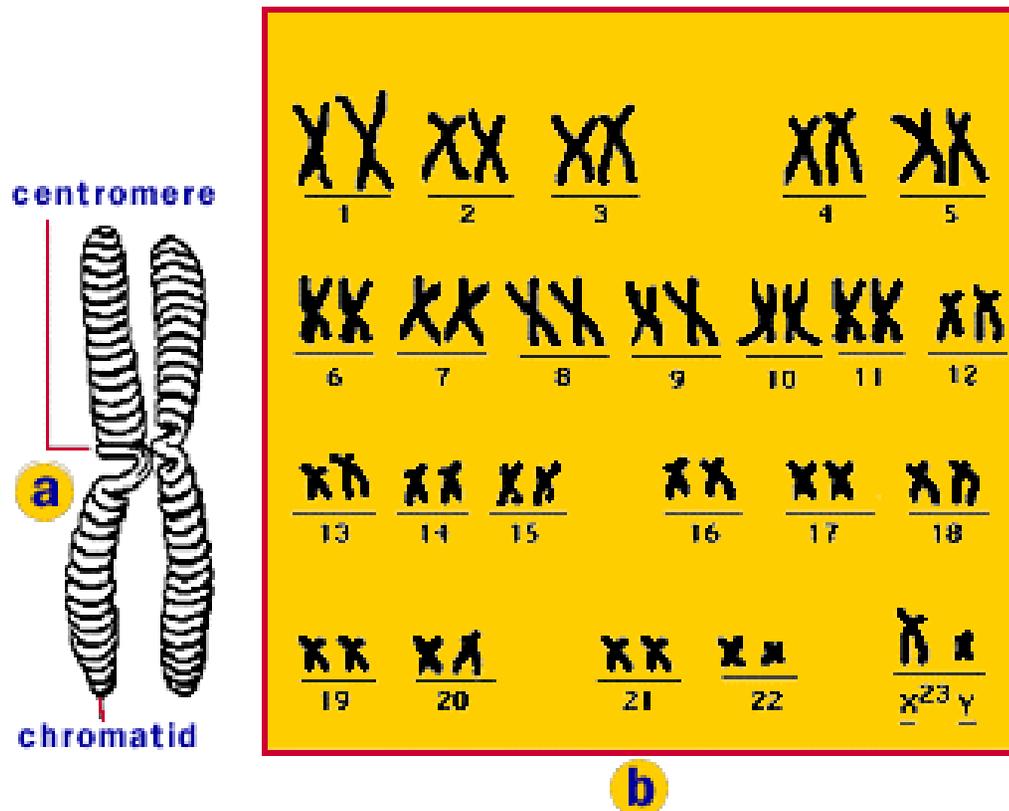


Basic Genetic Processes



Chromosomes

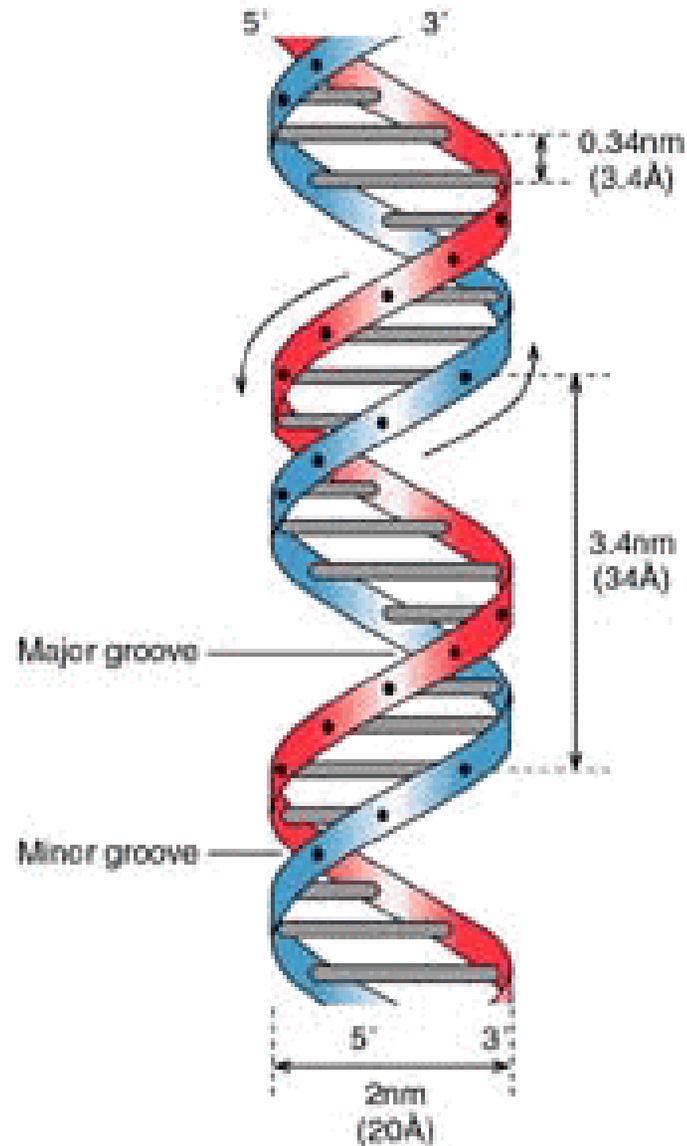
Human chromosomes!



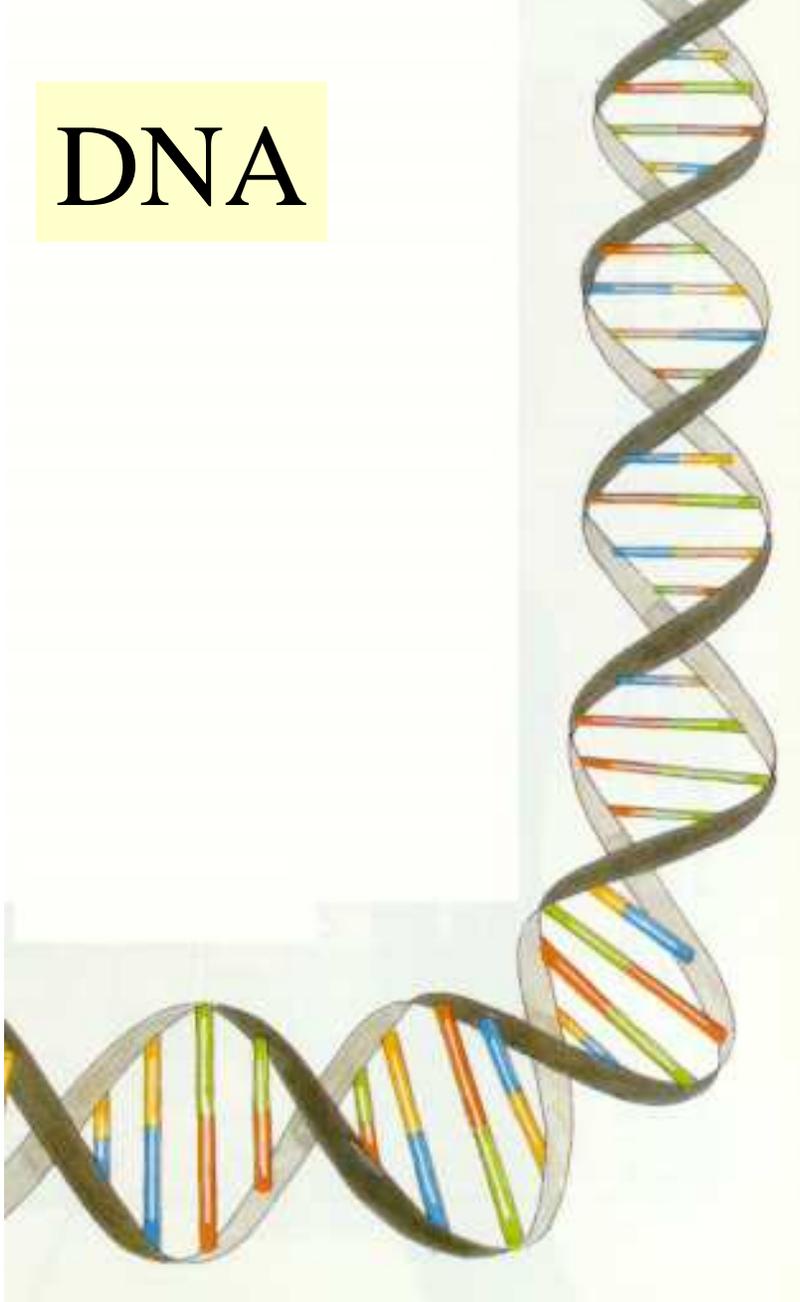
Chromosomes



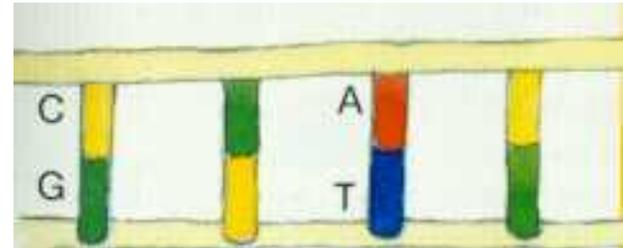
DNA Molecule



DNA



Complementary Bases

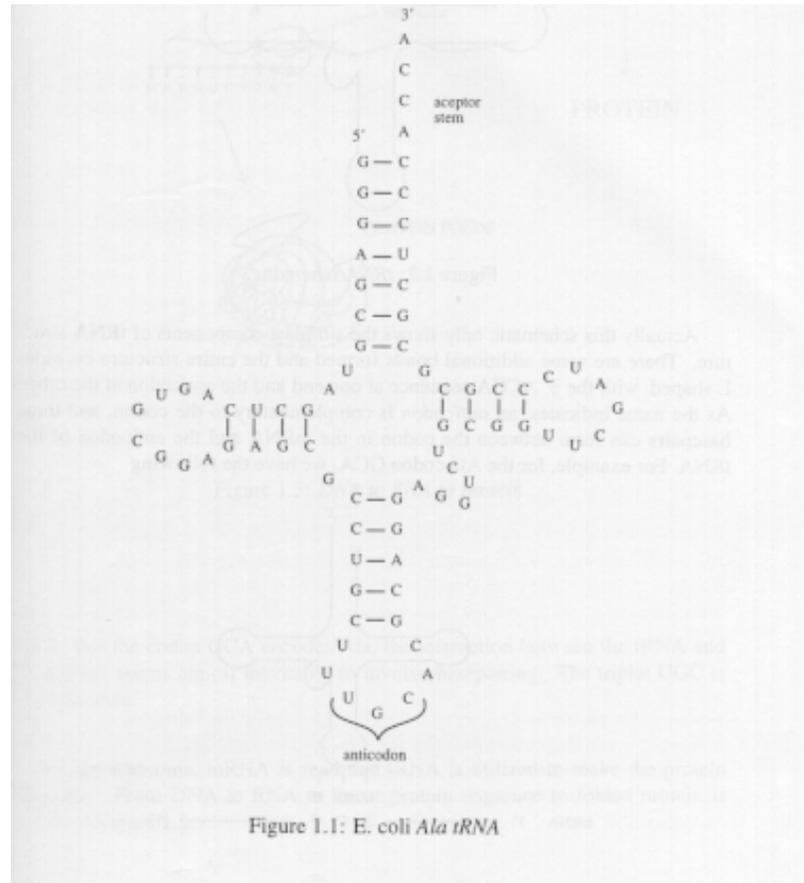


Proteins – Amino acids

amino acid	3 letter code	1 letter code
alanine	Ala	A
arginine	Arg	R
aspartic acid	Asp	D
asparagine	Asn	N
cysteine	Cys	C
glutamic acid	Glu	E
glutamine	Gln	Q
glycine	Gly	G
histine	His	H
isoleucine	Ile	I
leucine	Leu	L
lysine	Lys	K
methionine	Met	M
phenylalanine	Phe	F
proline	Pro	P
serine	Ser	S
threonine	Thr	T
tryptophan	Trp	W
tyrosine	Tyr	Y
valine	Val	V

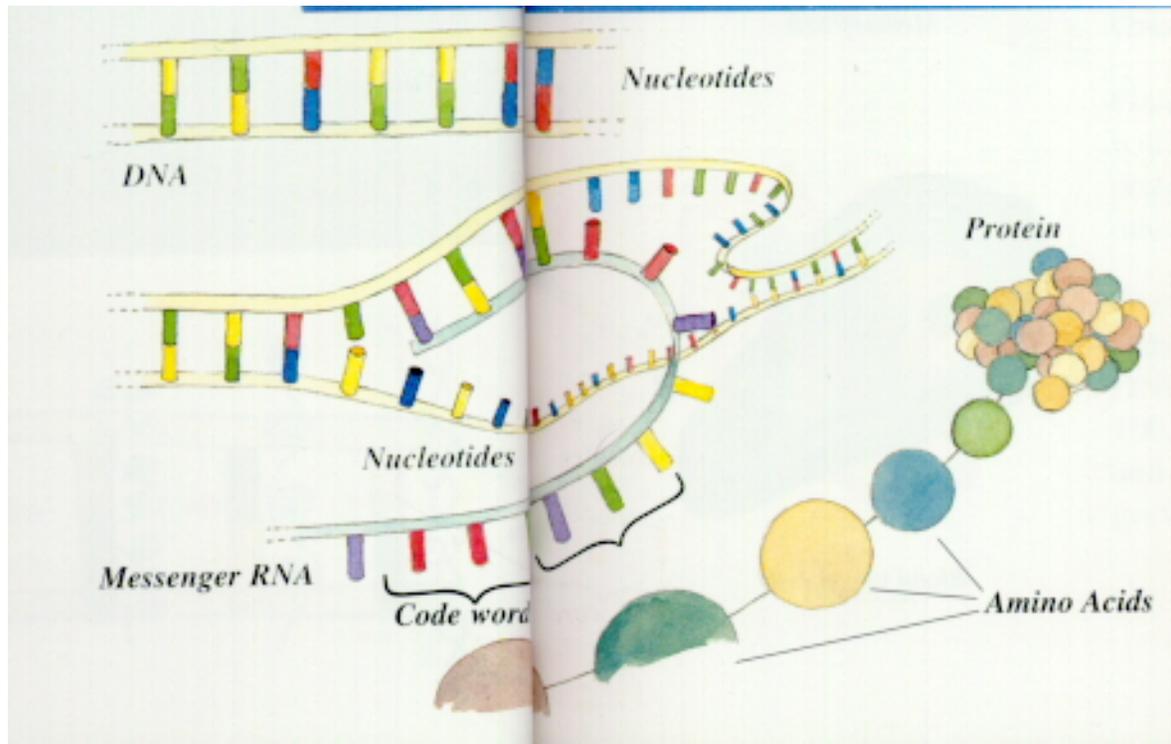
Table 1.1: *Amino acid abbreviations*

RNA

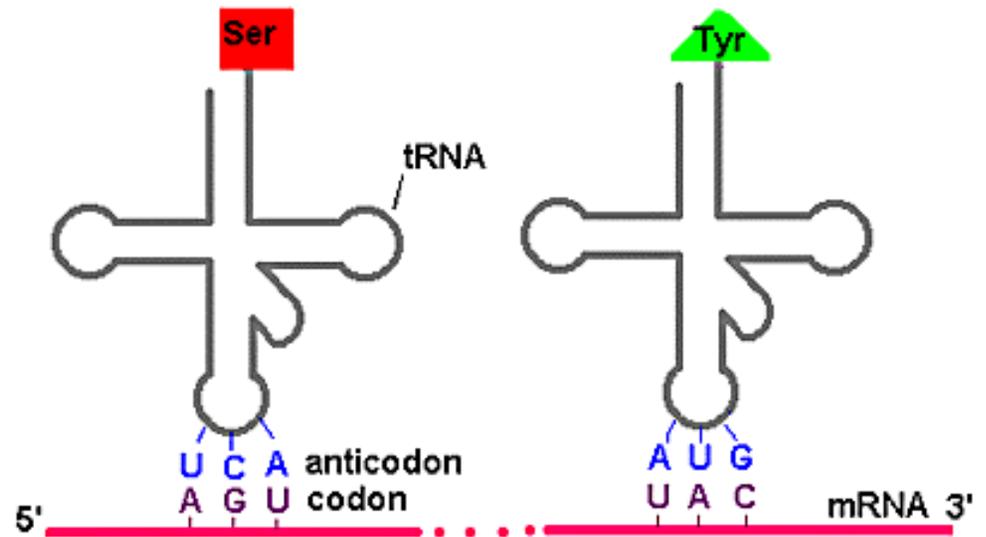


Genes





The Genetic Code



		2nd base in codon					
		U	C	A	G		
1st base in codon	U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr STOP STOP	Cys Cys STOP Trp	U C A G	3rd base in codon
	C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C A G	
	A	Ile Ile Ile Met	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G	
	G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	U C A G	

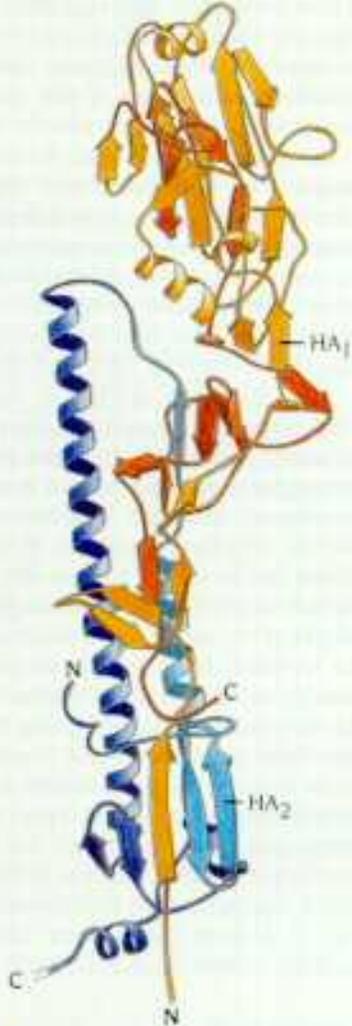
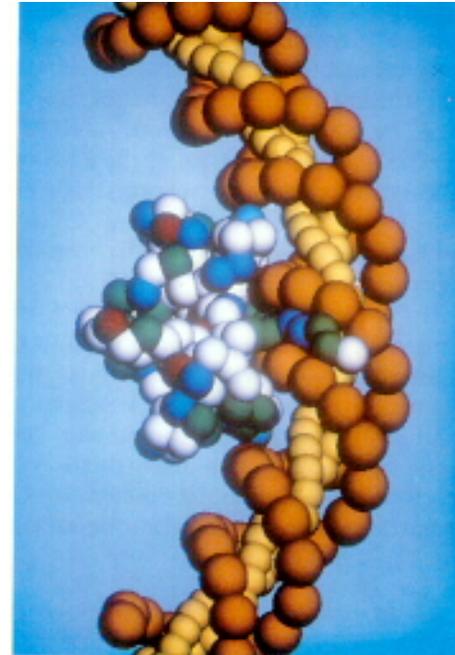
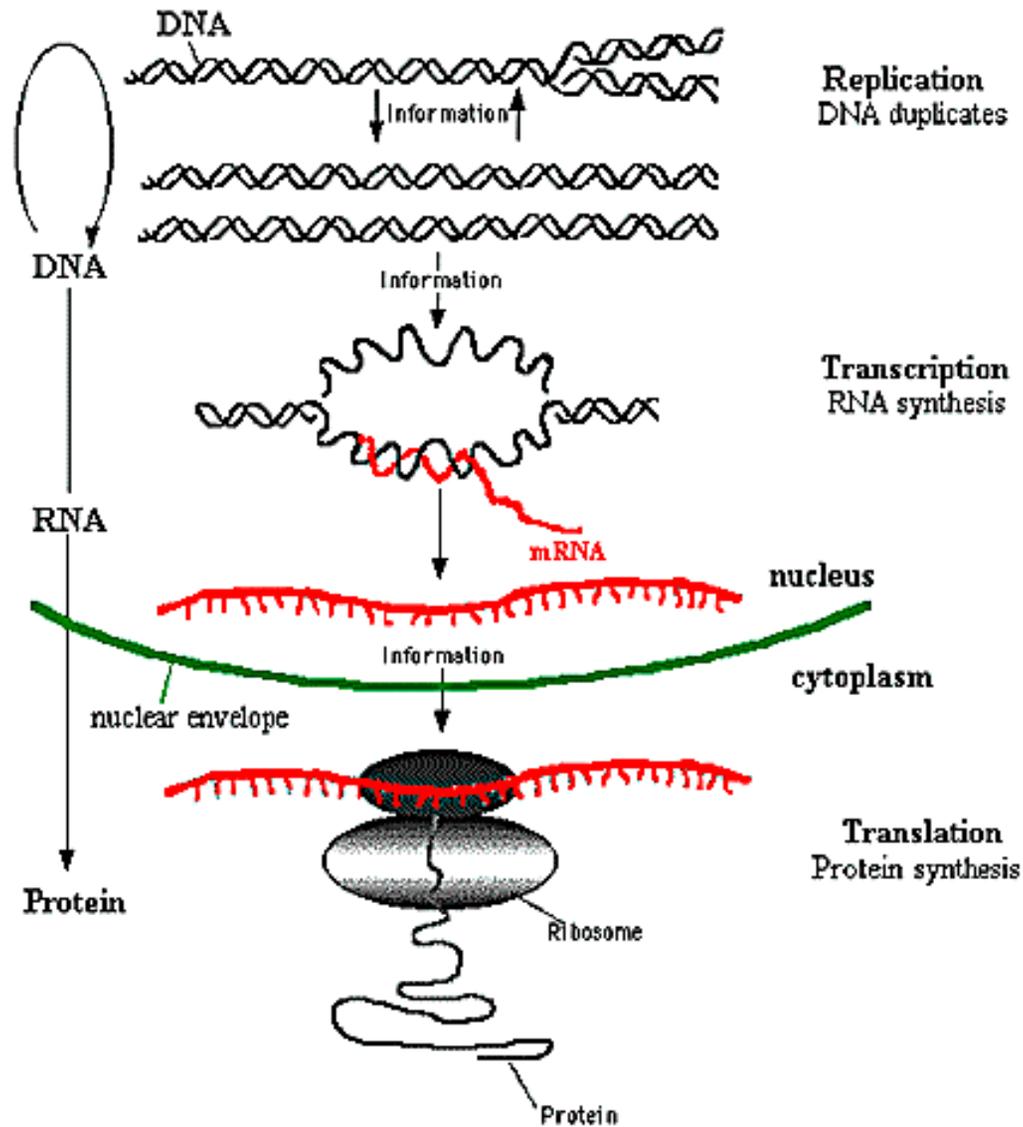
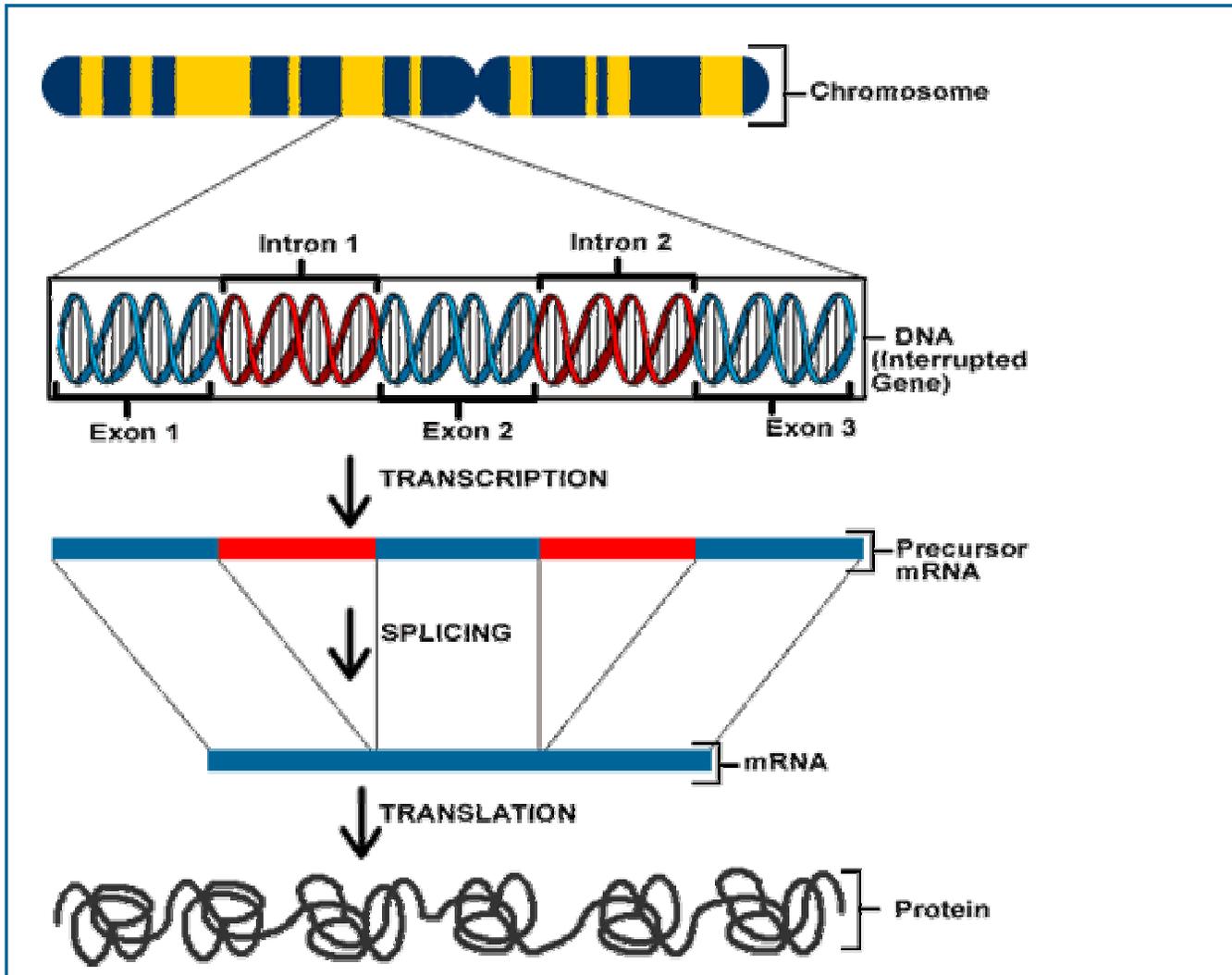


Figure 8.21 Schematic diagram of the subunit structure of hemagglutinin from influenza virus. The structure comprises about 550 amino acids arranged in two chains HA₁ (red) and HA₂ (blue). The first half of each chain has a lighter color in the diagram. The subunit is very elongated with a long stemlike region built up by residues from both chains and includes one of the longest α helices known in a globular structure, about 75 Å long. The globular head is formed by residues only from HA₁. (Courtesy of Don Wiley, Harvard University.)

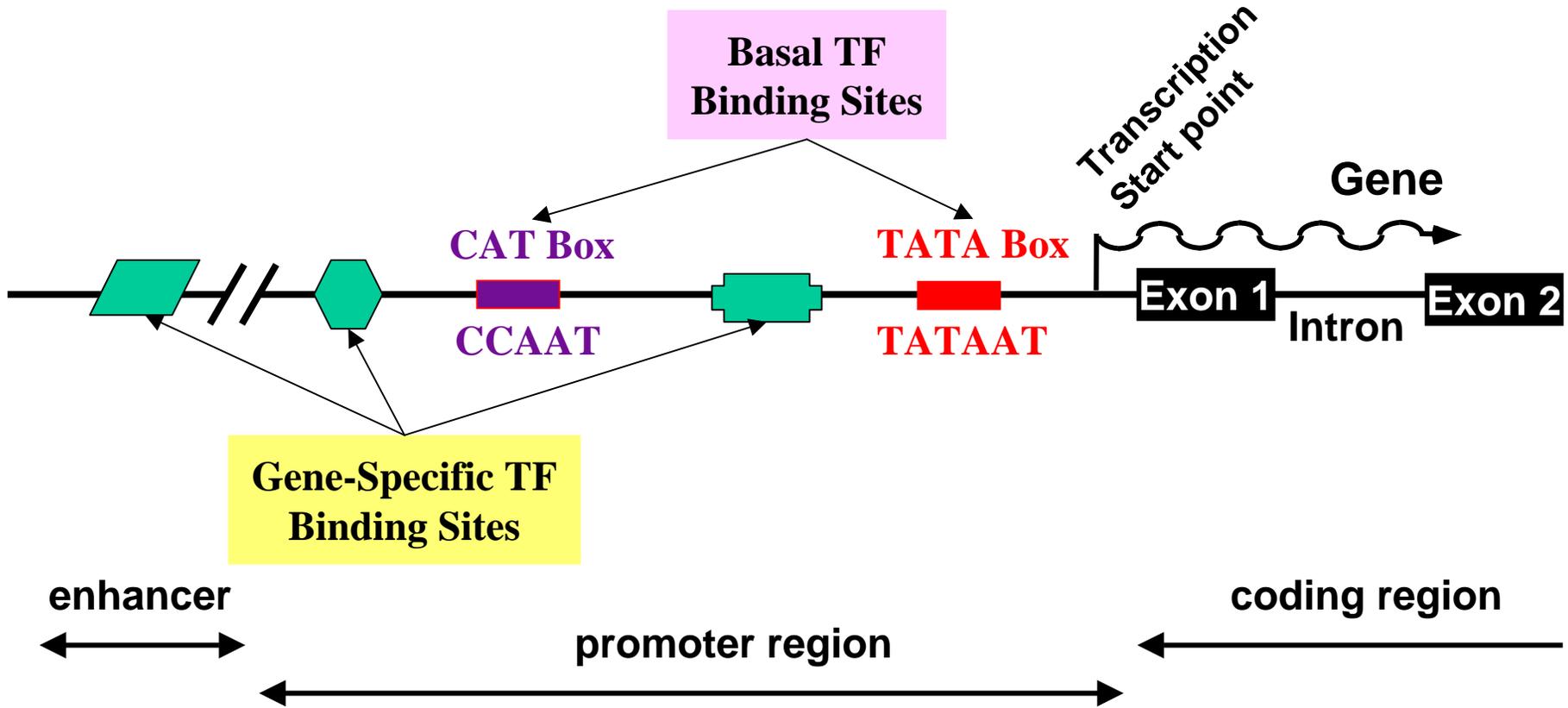




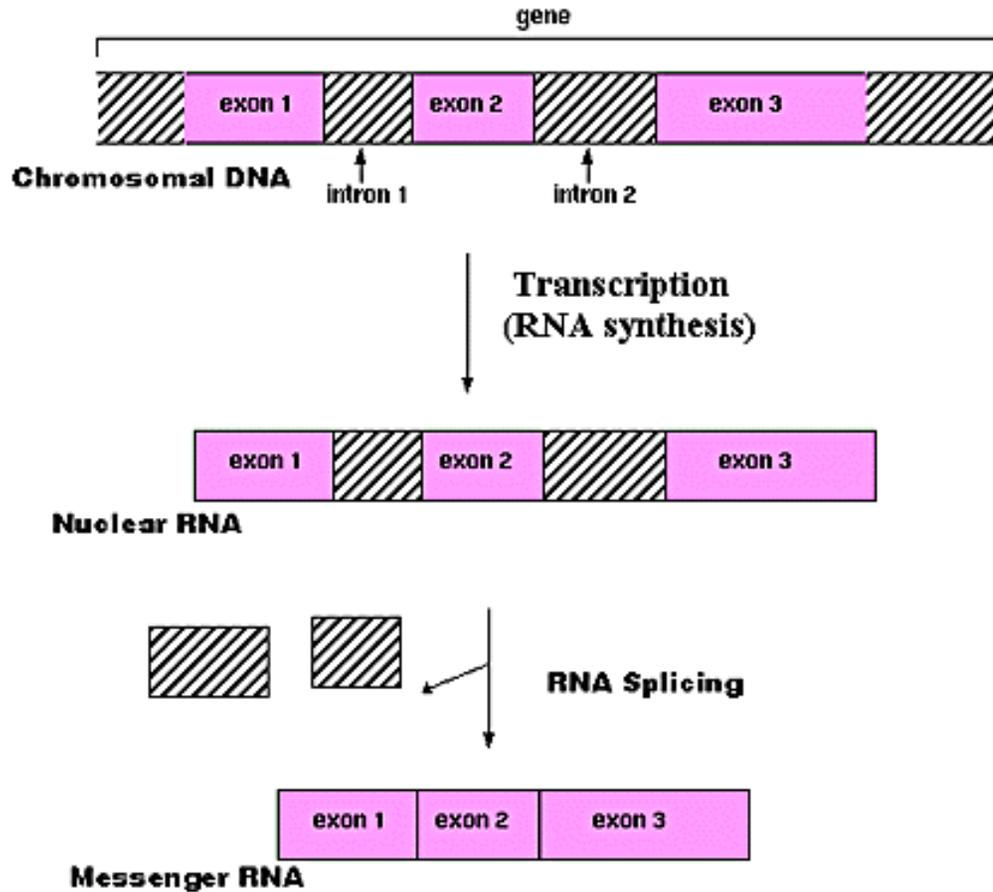
The Central Dogma of Molecular Biology



Transcription Regulation

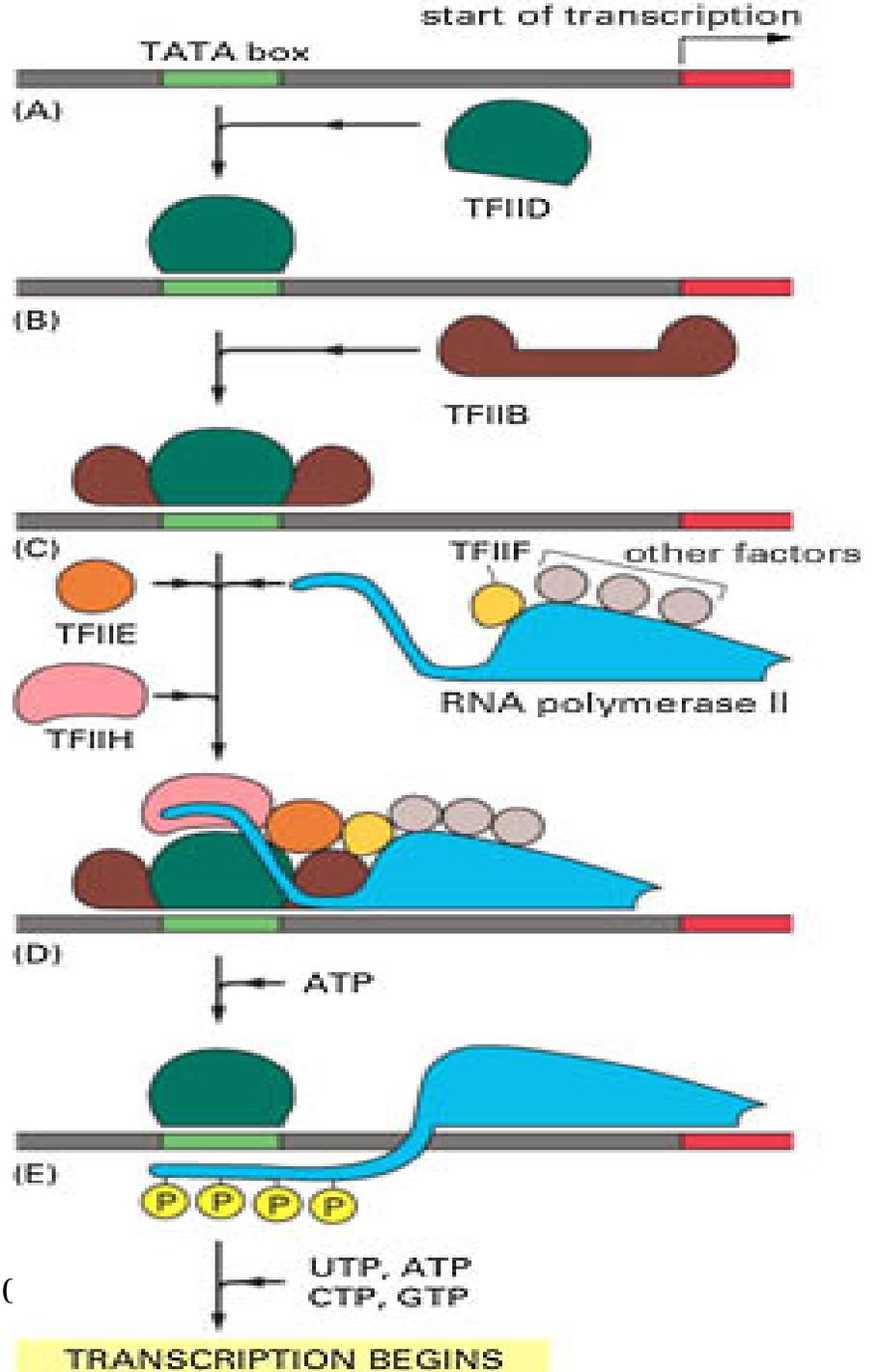


DNA Transcription



RNA synthesis and processing

Transcription Initiation



Transcription

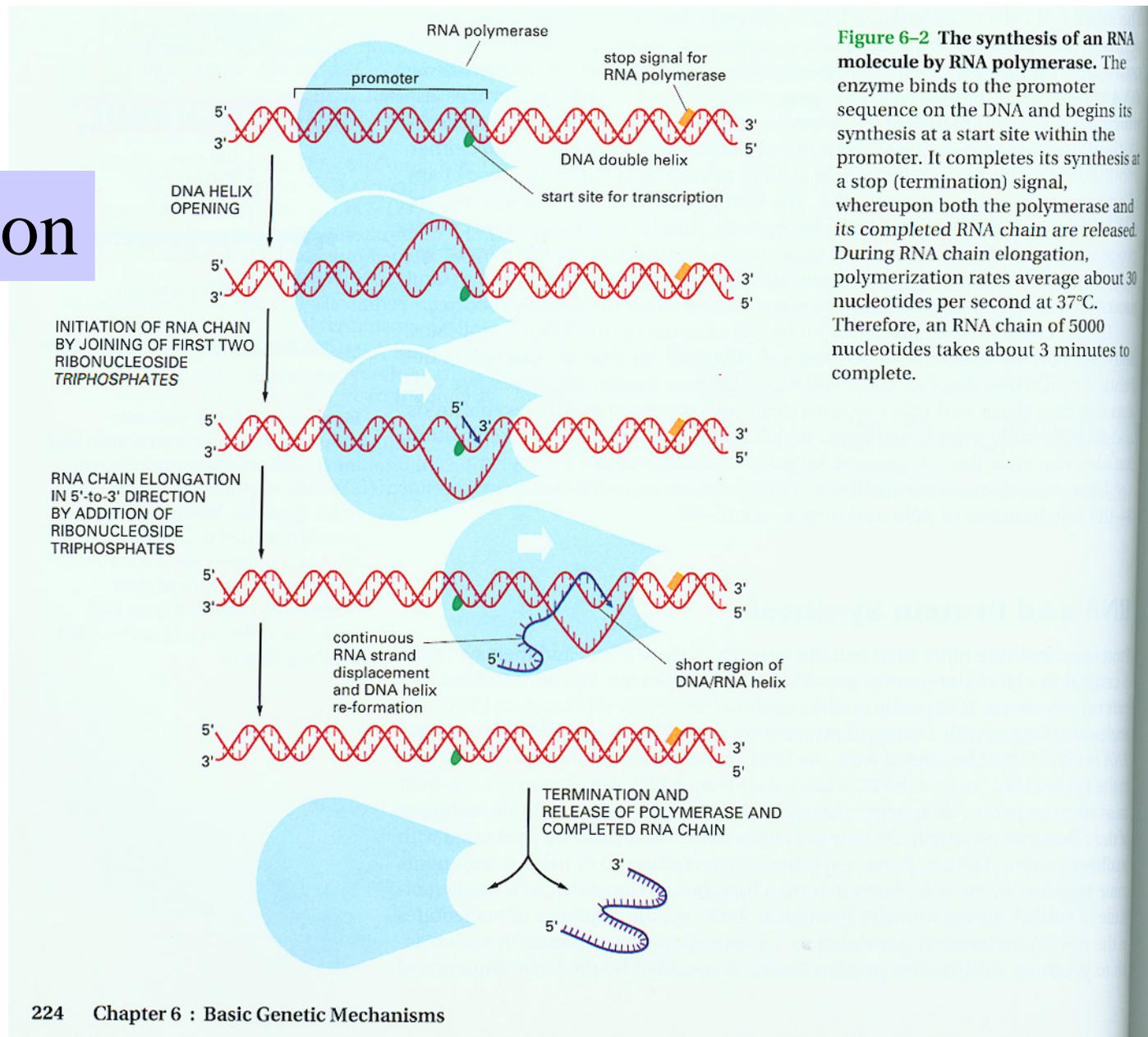


Figure 6-2 The synthesis of an RNA molecule by RNA polymerase. The enzyme binds to the promoter sequence on the DNA and begins its synthesis at a start site within the promoter. It completes its synthesis at a stop (termination) signal, whereupon both the polymerase and its completed RNA chain are released. During RNA chain elongation, polymerization rates average about 30 nucleotides per second at 37°C. Therefore, an RNA chain of 5000 nucleotides takes about 3 minutes to complete.

Transcription Steps

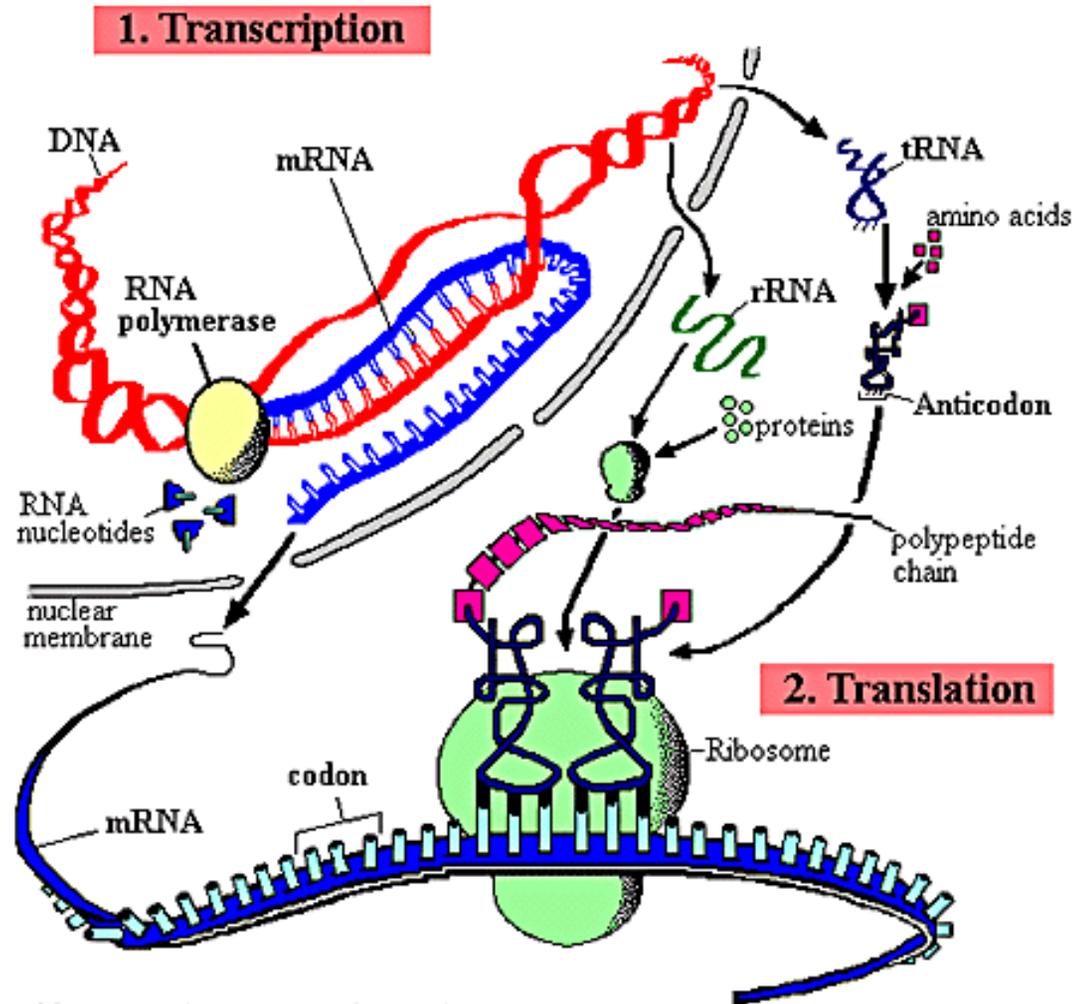
RNA polymerase needs many transcription factors (TFIIA, TFIIB, etc.)

- (A) The promoter sequence (TATA box) is located 25 nucleotides away from transcription initiation site.
- (B) The TATA box is recognized and bound by transcription factor TFIID, which then enables the adjacent binding of TFIIB. DNA is somewhat distorted in the process.
- (D) The rest of the general transcription factors as well as the RNA polymerase itself assemble at the promoter. What order?
- (E) TFIIF then uses ATP to phosphorylate RNA polymerase II, changing its conformation so that the polymerase is released from the complex and is able to start transcribing. As shown, the site of phosphorylation is a long polypeptide tail that extends from the polymerase molecule.

Transcription Factors

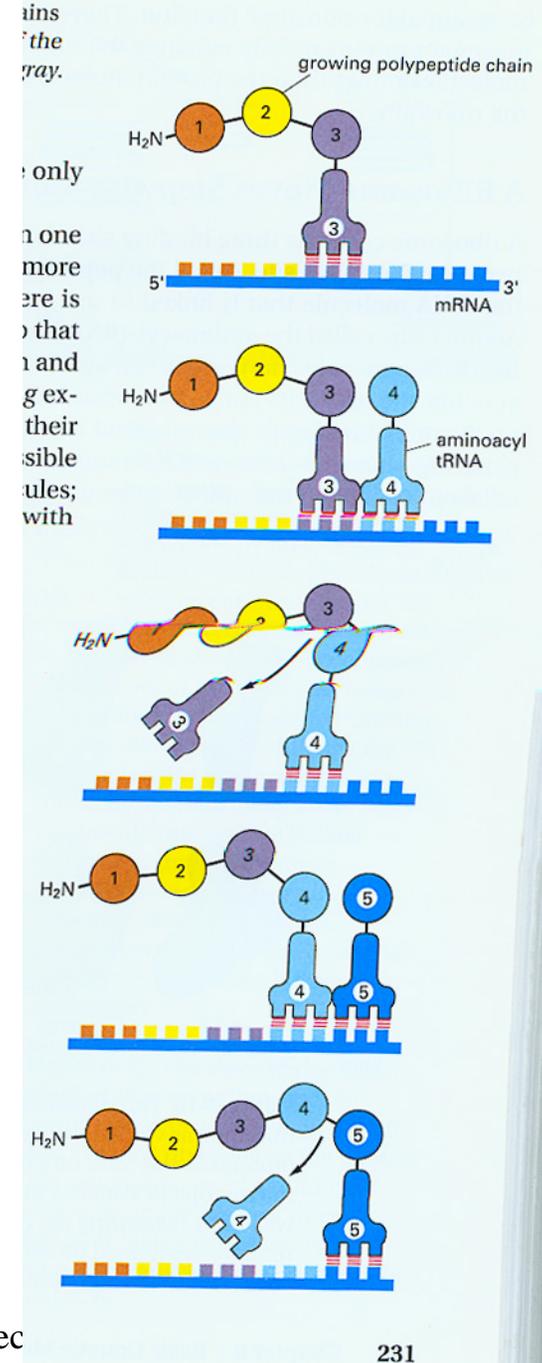
- The general transcription factors have been highly conserved in evolution; some of those from human cells can be replaced in biochemical experiments by the corresponding factors from simple yeasts.

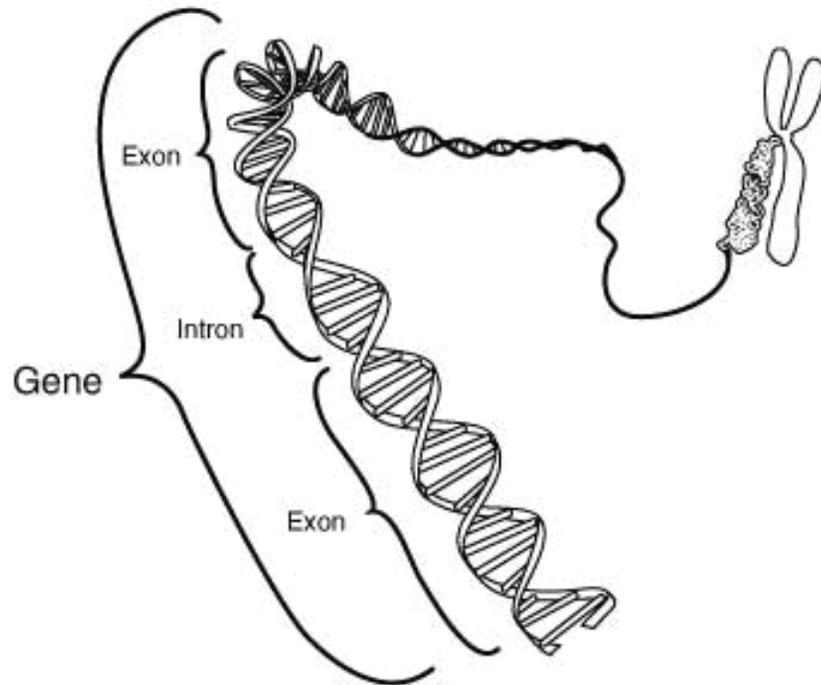
Protein Synthesis



Protein synthesis

Protein Synthesis: Incorporation of amino acid into protein





DNA $\xrightarrow{\text{Transcription}}$ mRNA $\xrightarrow{\text{Translation}}$ tRNA \rightarrow Amino Acid \rightarrow Polypeptide chain

