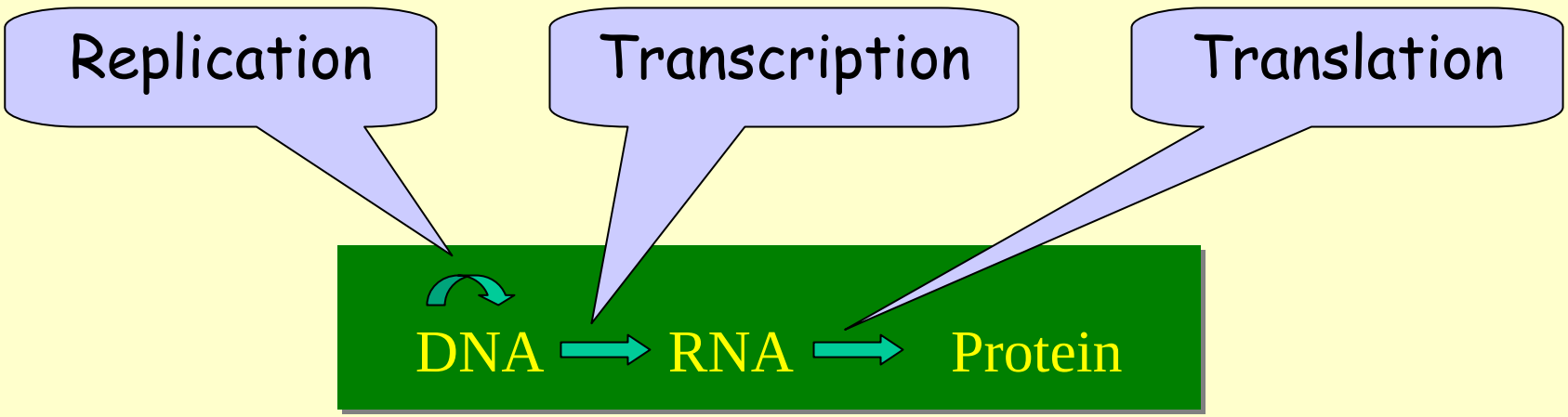
CAP 5510: Introduction to Bioinformatics

Giri Narasimhan
ECS 389; Phone: x3748
giri@cis.fiu.edu
www.cis.fiu.edu/~giri/teach/BioinfS07.html

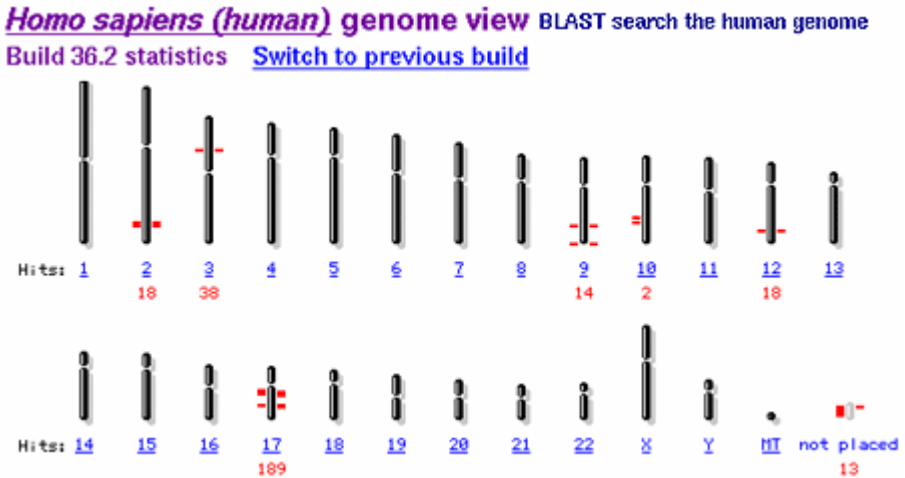
Molecular Biology Background

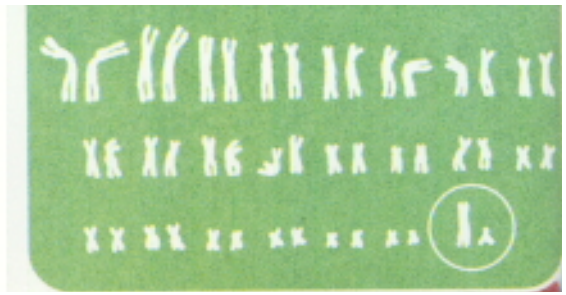
Central Dogma

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

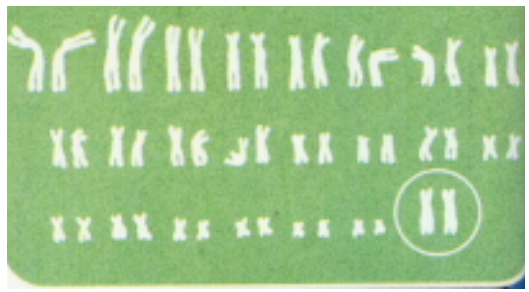


Chromosomes

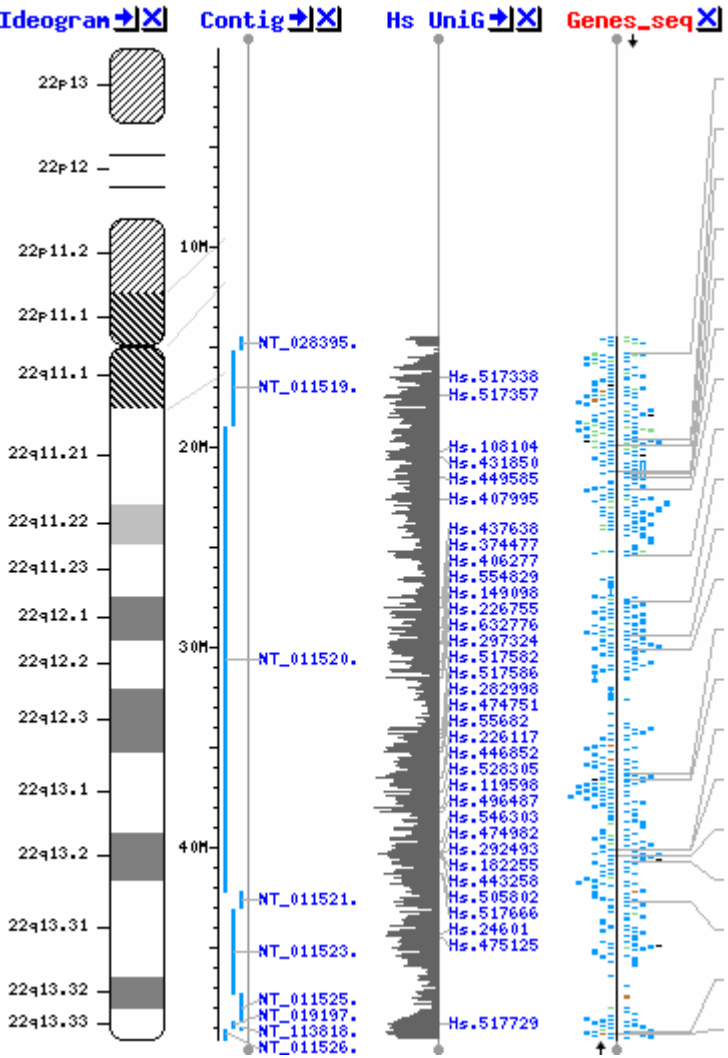




The chromosomal locations of several genes believed to be associated with the human BRCA1 gene implicated in breast cancer are highlighted.



Human Chr 22

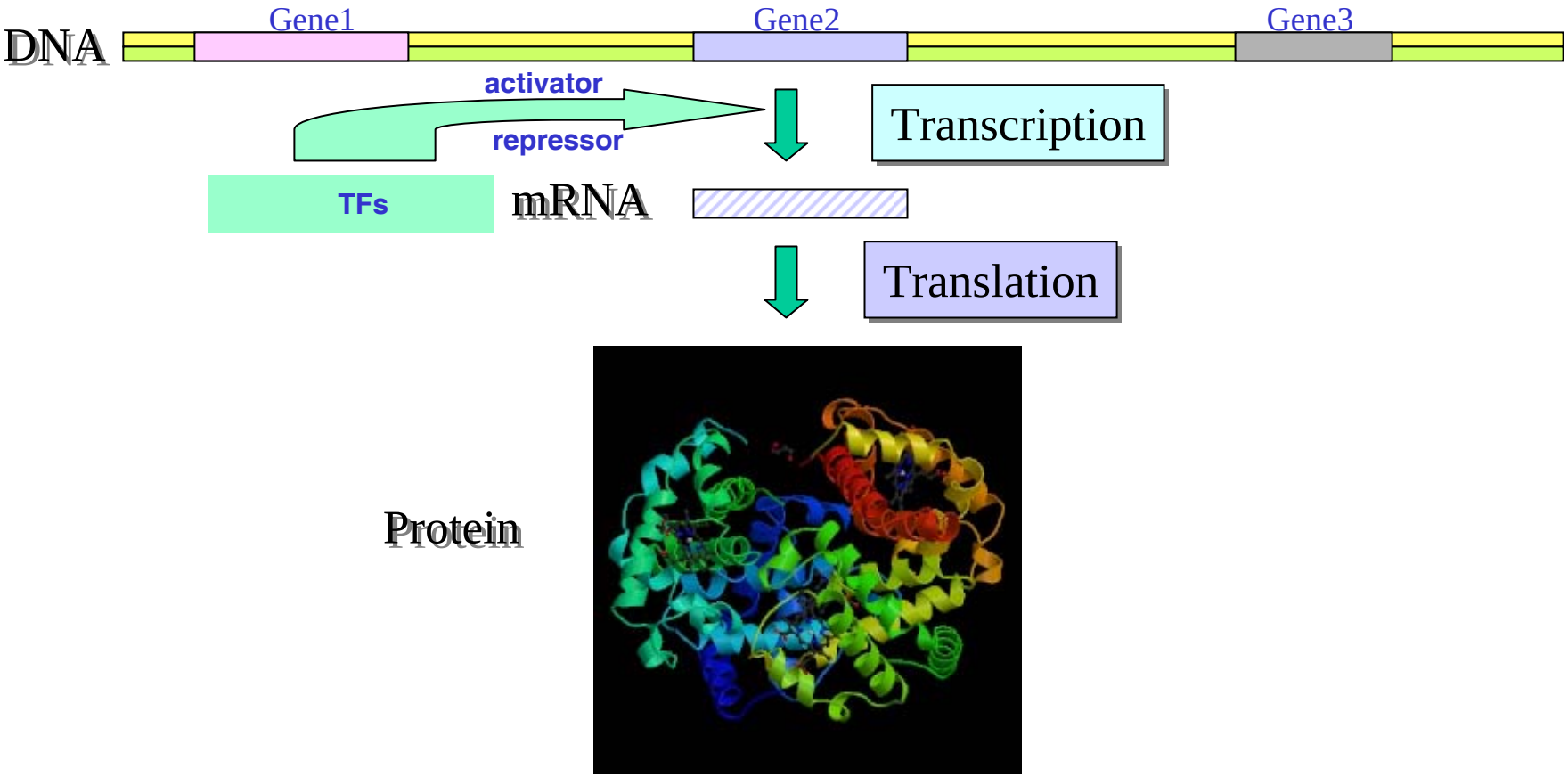


Symbol	Position	Description
ABCD1P4	22q11	ATP-binding cassette, sub-family D (ALD)
SNAP29	22q11.21	synaptosomal-associated protein

-
-
-

Genes

Central Dogma



RNA

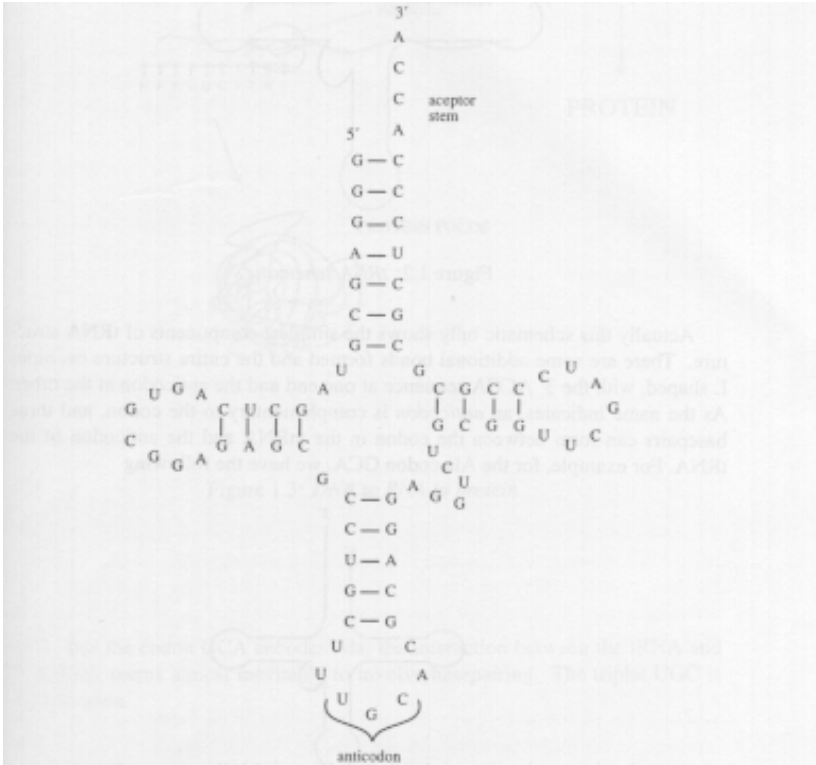


Figure 1.1: E. coli *Ala tRNA*

Basic Genetic Processes

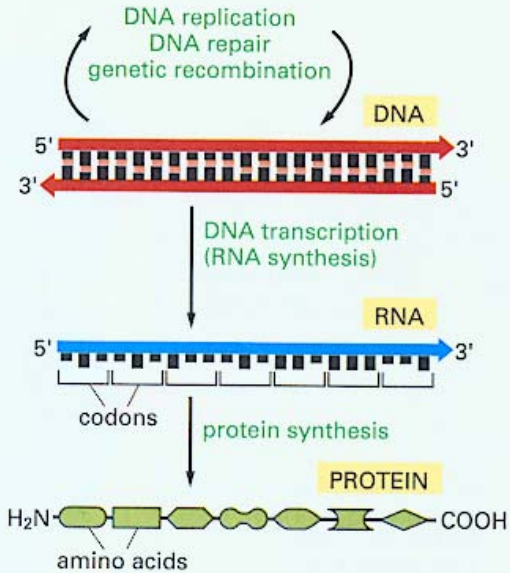
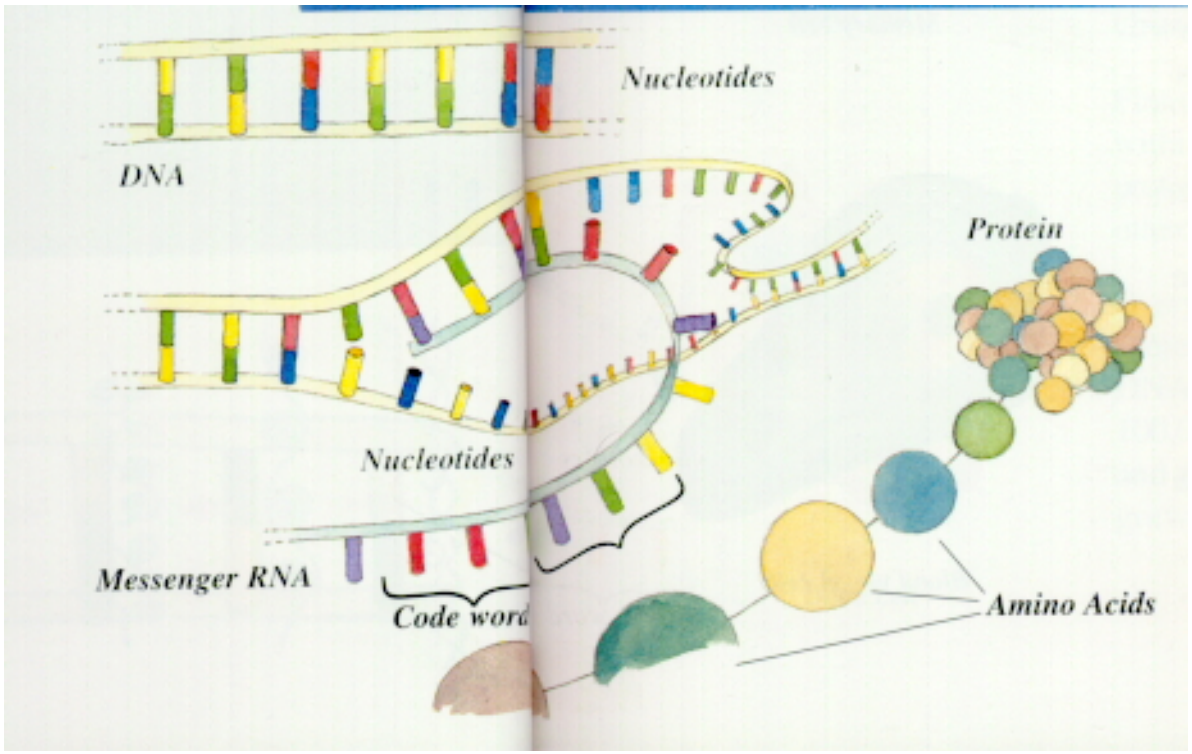


Figure 6–1 The basic genetic processes. The processes shown here are thought to occur in all present-day cells. Very early in the evolution of life, however, much simpler cells probably existed that lacked both DNA and proteins (see Figure 1–11). Note that a sequence of three nucleotides (a codon) in an RNA molecule codes for a specific amino acid in a protein.

DNA → RNA → Protein



Enterobacteria phage lambda

```
ORIGIN      5' end of the ecori-g fragment, polarity of l strand.
        1 aattcttttg ctttttaccc tggaagaaat actcataagc cacctctgtt atttaccccc
       61 aatcttcaca agaaaaactg tatttgacaa acaagataca ttgtatgaaa atacaagaaa
      121 gtttgttgat ggaggcgata tgcaaactct ttctgaacgc ctcaagaaga ggcgaattgc
      181 gttaaaaatg acgcaaaccg aactggcaac caaagccggt gttaaacagc aatcaattca
      241 actgattgaa gctggagtaa ccaagcgacc gcgcttcttg tttgagattg ctatggcgct
      301 taactgtgat ccggtttggt tacagtacgg aactaaacgc ggtaaagccg cttaagacat
      361 tcccgctctt acacattcca gccctgaaaa agggcatcaa attaaaccac acctatggtg
      421 tatgcattta tttgcataca ttcaatcaat tgttatctaa ggaaatactt acatatggtt
      481 cgtgcaaaca aacgcaacga ggctctacga atcgagagtg cgttgcttaa caaaatcgca
      541 atgcttggaa ctgagaagac agcggaagct gtgggcgttg ataagtcgca gatcagcagg
      601 tggaagaggg actggattcc aaagttctca atgctgcttg ctgttcttga atgggggtc
      661 gttgacgacg acatggctcg attggcgcga caagttgctg cgattctcac caataaaaaa
      721 cgcccggcgg caaccgagcg ttctgaacaa atccagatgg agttctgagg tcattactgg
      781 atctatcaac aggagtcatt atgacaaata cagcaaaaat actcaacttc ggcagaggta
      841 actttgccgg acaggagcgt aatgtggcag atctcgatga tggttacgcc agactatcaa
      901 atatgctgct tgaggcttat tcgggcgcag atctgaccaa gcgacagttt aaagtgctgc
      961 ttgccattct gcgtaaaacc tatgggtgga ataaaccaat ggacagaatc accgattctc
     1021 aacttagcga gattacaaag ttacctgtca aacggtgcaa tgaagccaag ttagaactcg
     1081 tcagaatgaa tattatcaag cagcaaggcg gcatgtttgg accaaataaa aacatctcag
     1141 aatggtgcat ccctcaaaac gagggaaaat ccccctaaaac gagggataaa acatccctca
     1201 aattggggga ttgctatccc tcaaaacagg gggacacaaa agacactatt acaaaagaaa
     1261 aaagaaaaga ttattcgtca gagaatt
//
```

DNA

Gene for protein Cro
140-355

Enterobacteria phage lambda

```
ORIGIN      5' end of the ecori-g fragment, polarity of 1 strand.
        1 aattcttttg ctttttaccc tggaagaaat actcataagc cacctctgtt atttaccccc
       61 aatcttcaca agaaaaactg tatttgacaa acaagataca ttgtatgaaa atacaagaaa
      121 gtttgttgat ggaggcgata tgcaaactct ttctgaacgc ctcaagaaga ggcgaattgc
      181 gttaaaaatg acgcaaaccg aactggcaac caaagccggt gttaaacagc aatcaattca
      241 actgattgaa gctggagtaa ccaagcgacc gcgcttcttg tttgagattg ctatggcgct
      301 taactgtgat ccggtttggt tacagtacgg aactaaacgc ggtaaagccg cttaagacat
      361 tcccgctctt acacattcca gccctgaaaa agggcatcaa attaaaccac acctatggtg
      421 tatgcattta tttgcataca ttcaatcaat tgttatctaa ggaaatactt acatatggtt
      481 cgtgcaaaca aacgcaacga ggctctacga atcgagagtg cgttgcttaa caaaatcgca
      • • •
     1141 aatggtgcat ccctcaaaac gagggaaaat cccctaaaac gagggataaa acatccctca
     1201 aattggggga ttgctatccc tcaaaacagg gggacacaaa agacactatt acaaaagaaa
     1261 aaagaaaaga ttattcgtca gagaatt
//
```

DNA

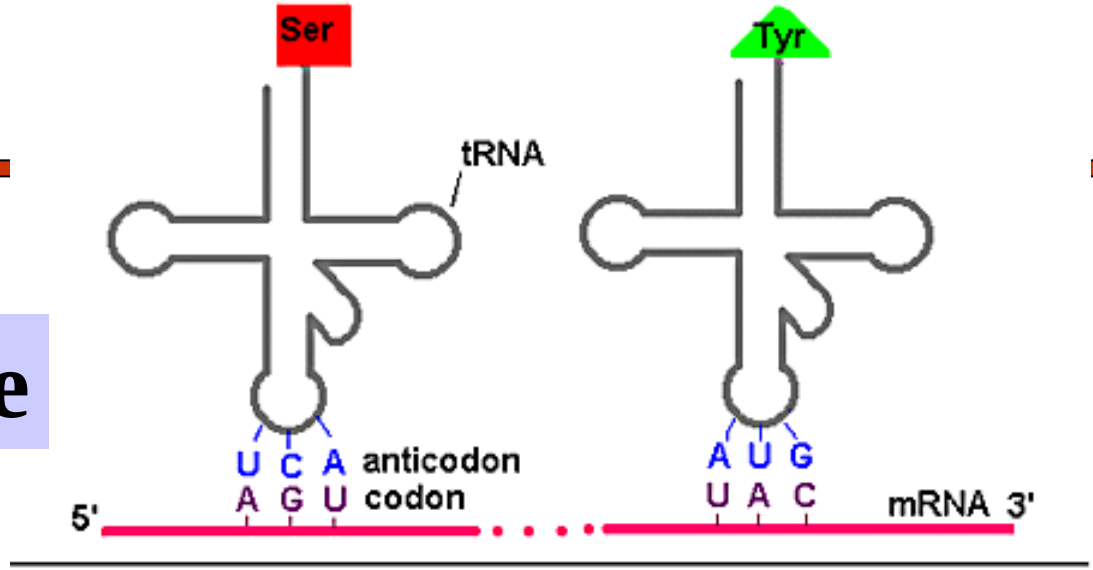
Gene for protein Cro
140-355

```
                                a ugcaaacucu uucugaacgc cucaagaaga ggcgaauugc
          guuaaaaaug acgcaaaccg aacuggcaac caaagccggu guuaaacagc aaucaauuca
mRNA      acugauugaa gcuggaguaa ccaagcgacc gcgcuucuug uuugagauug cuauggcgcu
          uaacugugau ccgguuuggu uacaguacgg aacuaaacgc gguaaagccg cuuaa
```

```
ORIGIN
        1 mqtlserlkk rrialkmtqt elatkagvkq qsiqlieagv tkrprflfei amalncdpvw
       61 lqygtkrgka a
//
```

Protein

The Genetic Code



2nd base in codon

1st base in codon	U	C	A	G	3rd base in codon
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

The Genetic Code

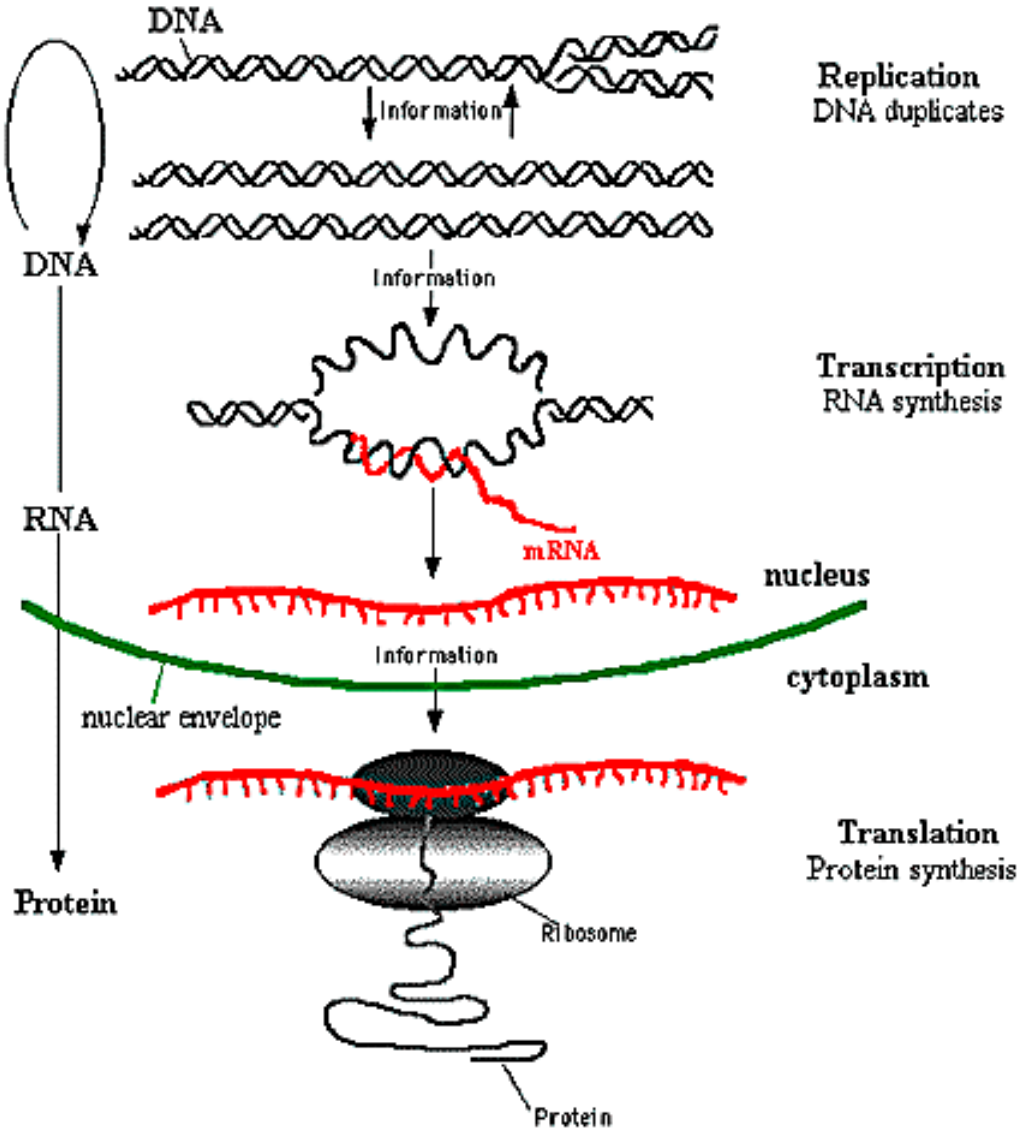
Enterobacteria phage lambda

mRNA

```
                        a ugcaaacucu uucugaacgc cucaagaaga ggcgaauugc
guuaaaaaug acgcaaaccg aacuggcaac caaagccggu guuaaacagc aaucaauuca
acugauugaa gcuggaguaa ccaagcgacc gcgcuucuug uuugagauug cuauggcgcu
uaacugugau ccgguuuggu uacaguacgg aacuaaacgc gguaaagccg cuuaa
```

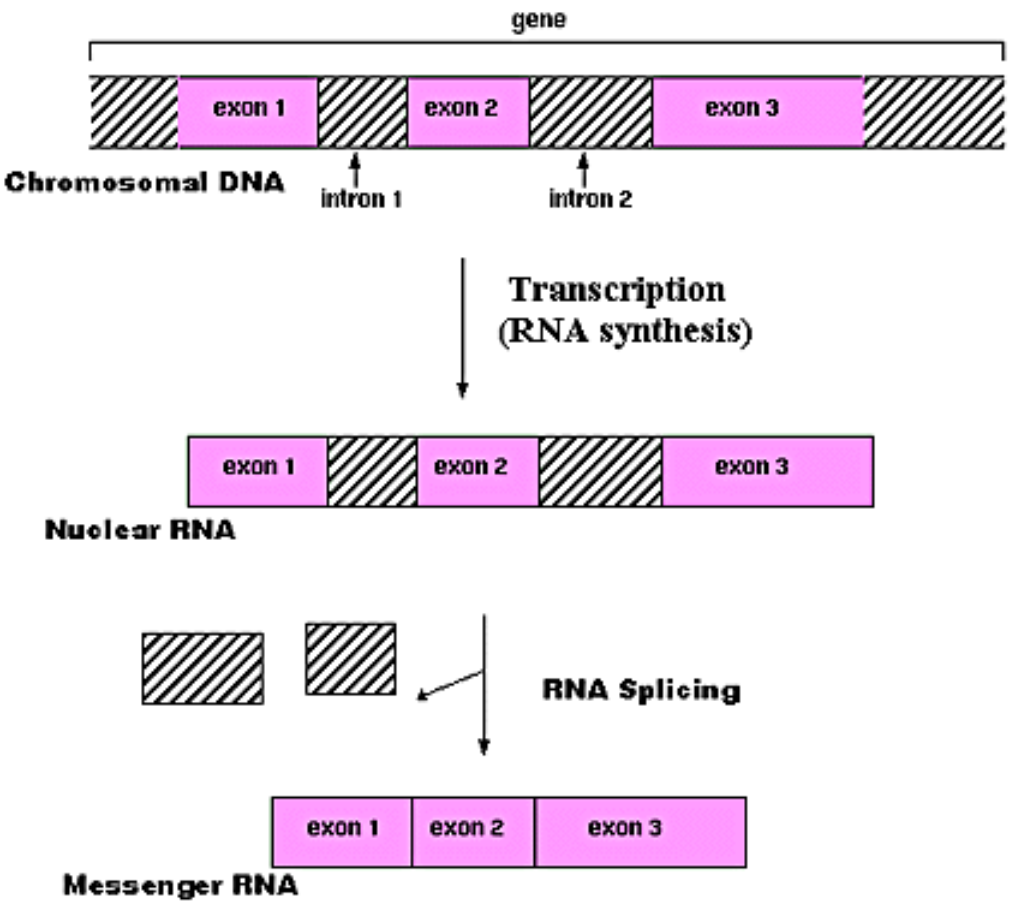
Protein

```
  1 mqtlserlkk rrialkmtqt elatkagvkq qsiqlieagv tkrprflfei amalncdpvw
 61 lqygtkrgka a
```

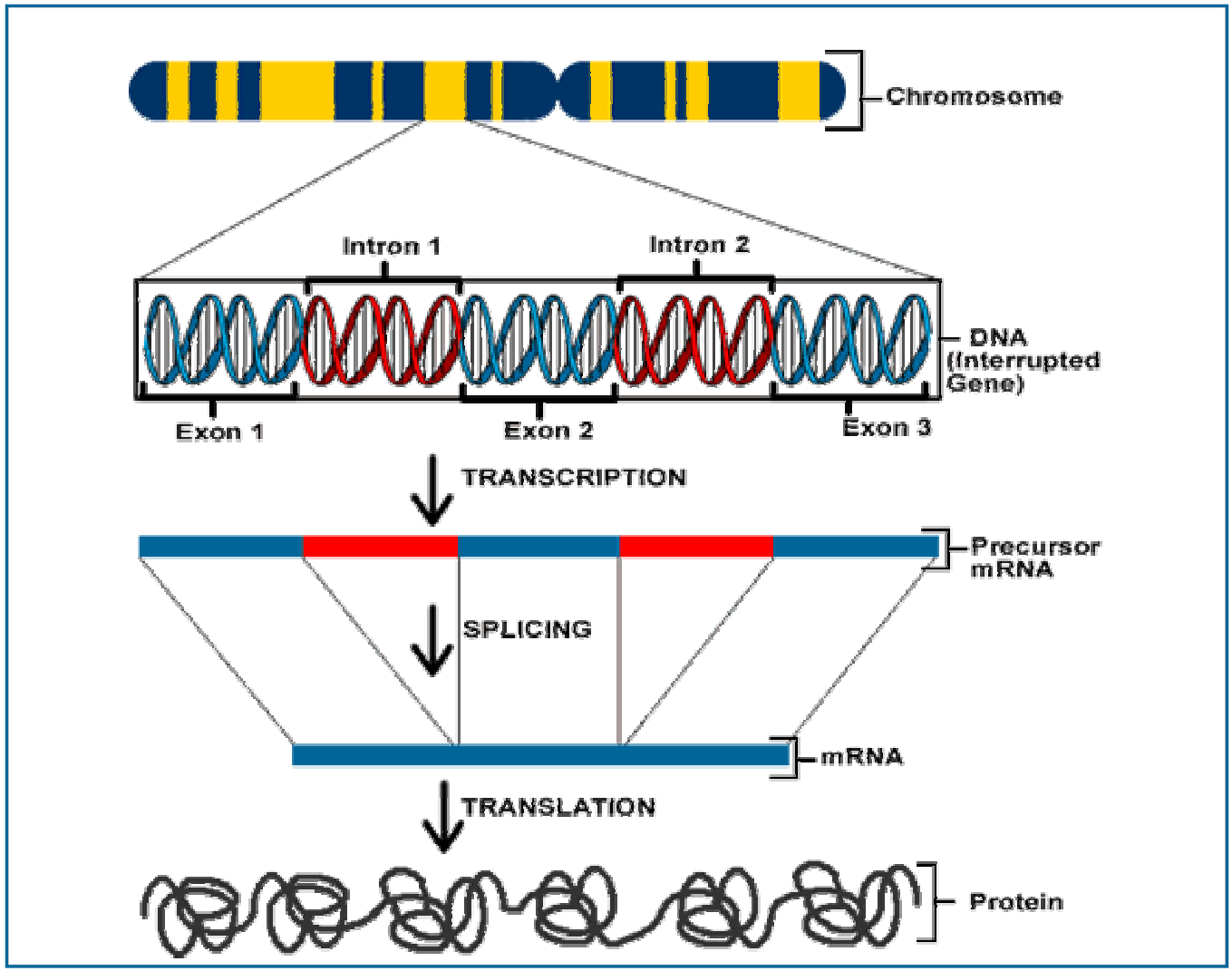


The Central Dogma of Molecular Biology

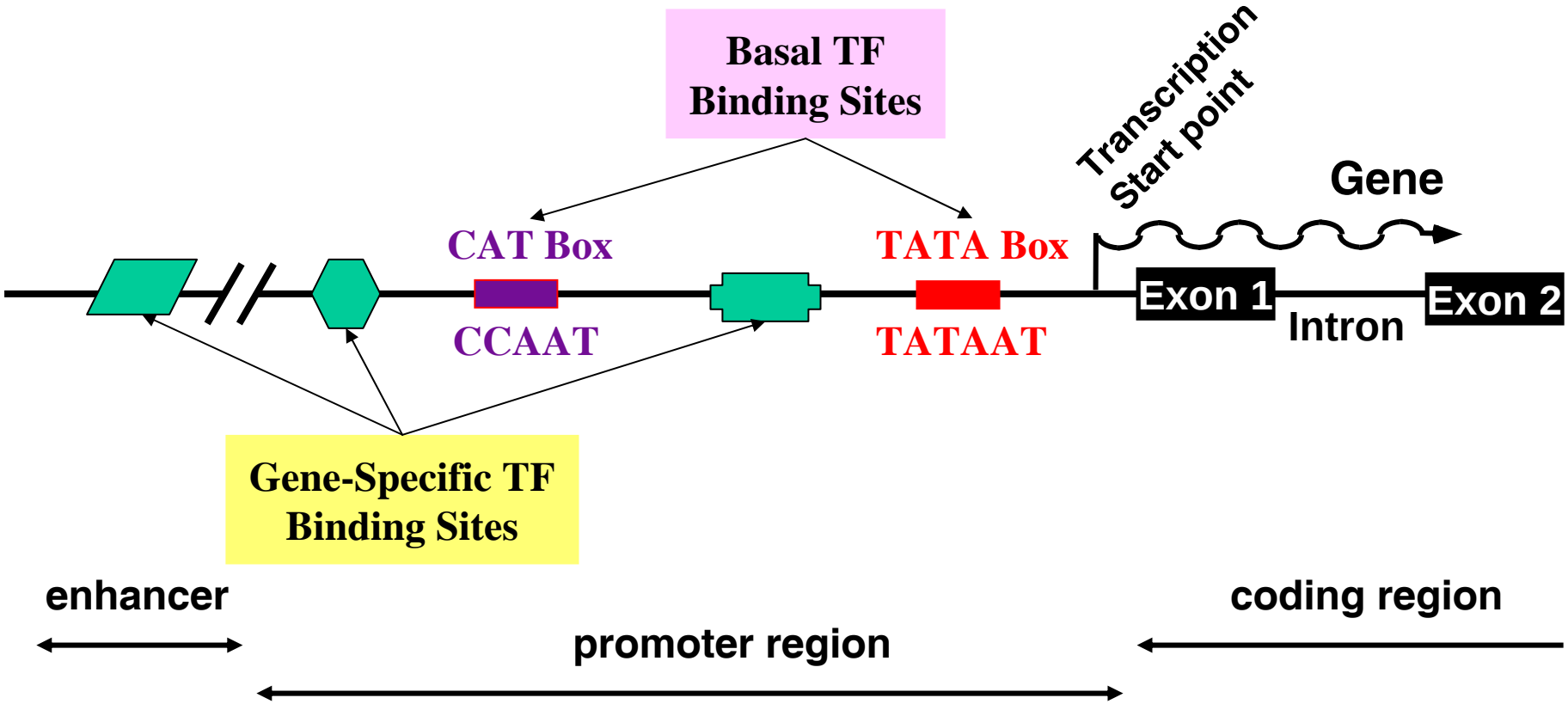
DNA Transcription



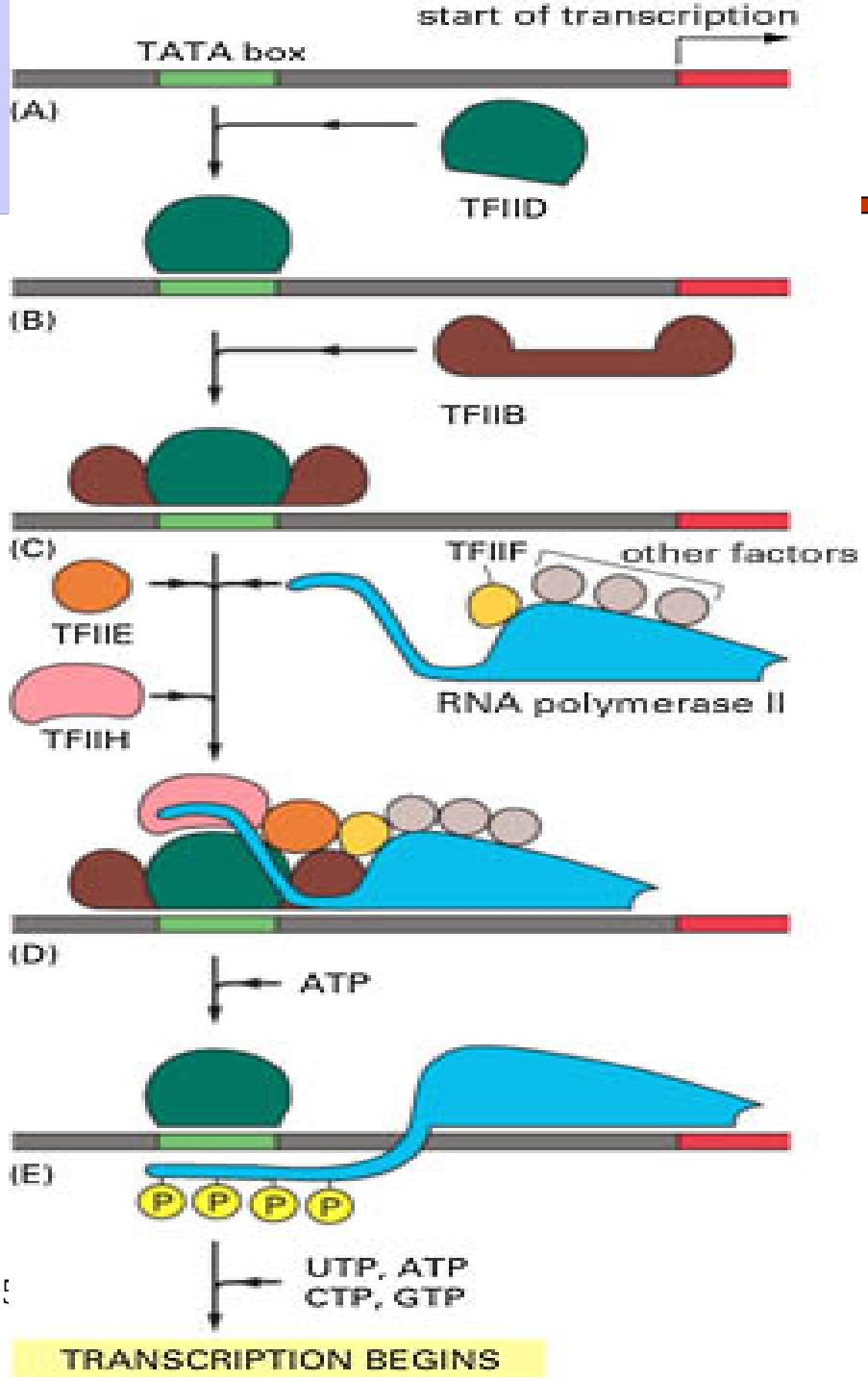
RNA synthesis and processing


Chromosome
Intron 1
Intron 2
DNA
(Interrupted
Gene)
Exon 1
Exon 2
Exon 3
TRANSCRIPTION
Precursor
mRNA
SPLICING
mRNA
TRANSLATION
Protein

Transcription Regulation



Transcription Initiation



Transcription

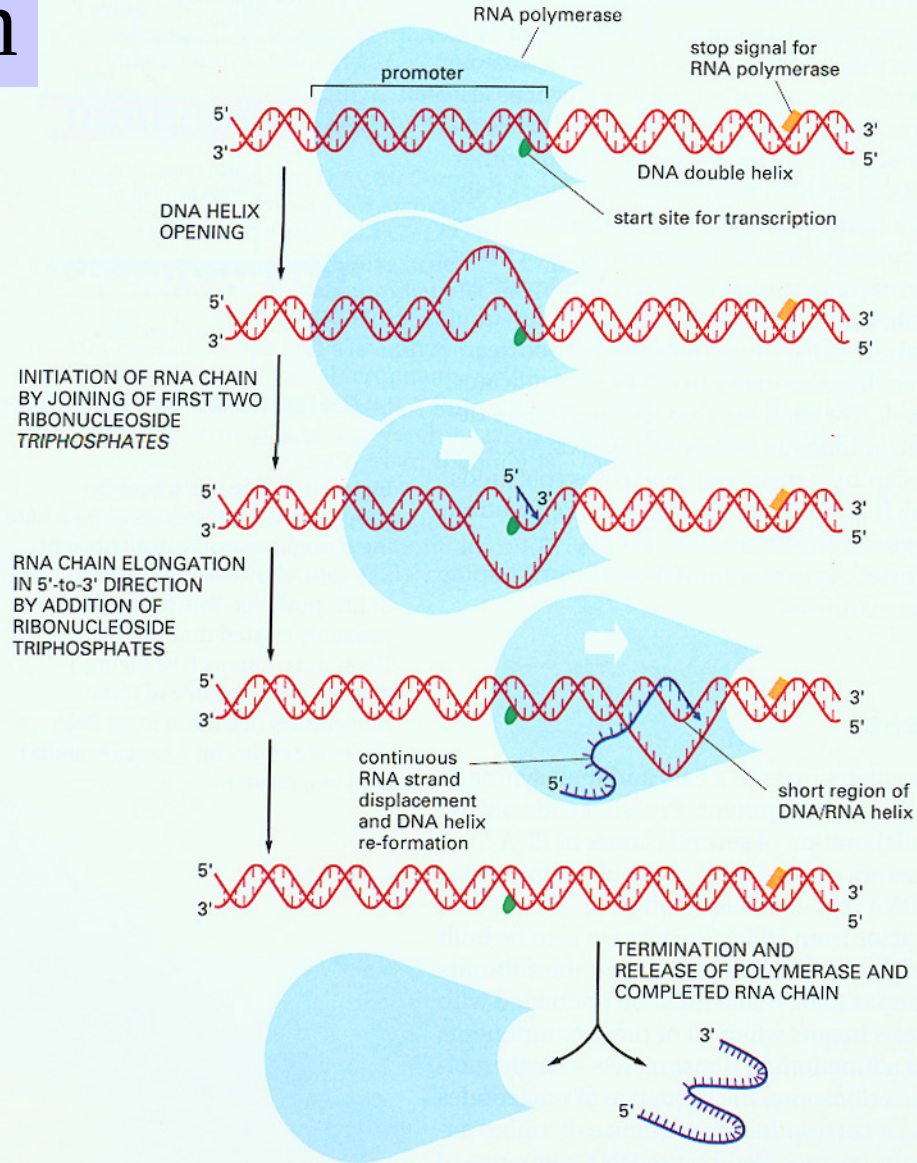


Figure 6–2 The synthesis of an RNA molecule by RNA polymerase. The enzyme binds to the promoter sequence on the DNA and begins its synthesis at a start site within the promoter. It completes its synthesis at a stop (termination) signal, whereupon both the polymerase and its completed RNA chain are released. 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.

224 Chapter 6 : Basic Genetic Mechanisms

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.

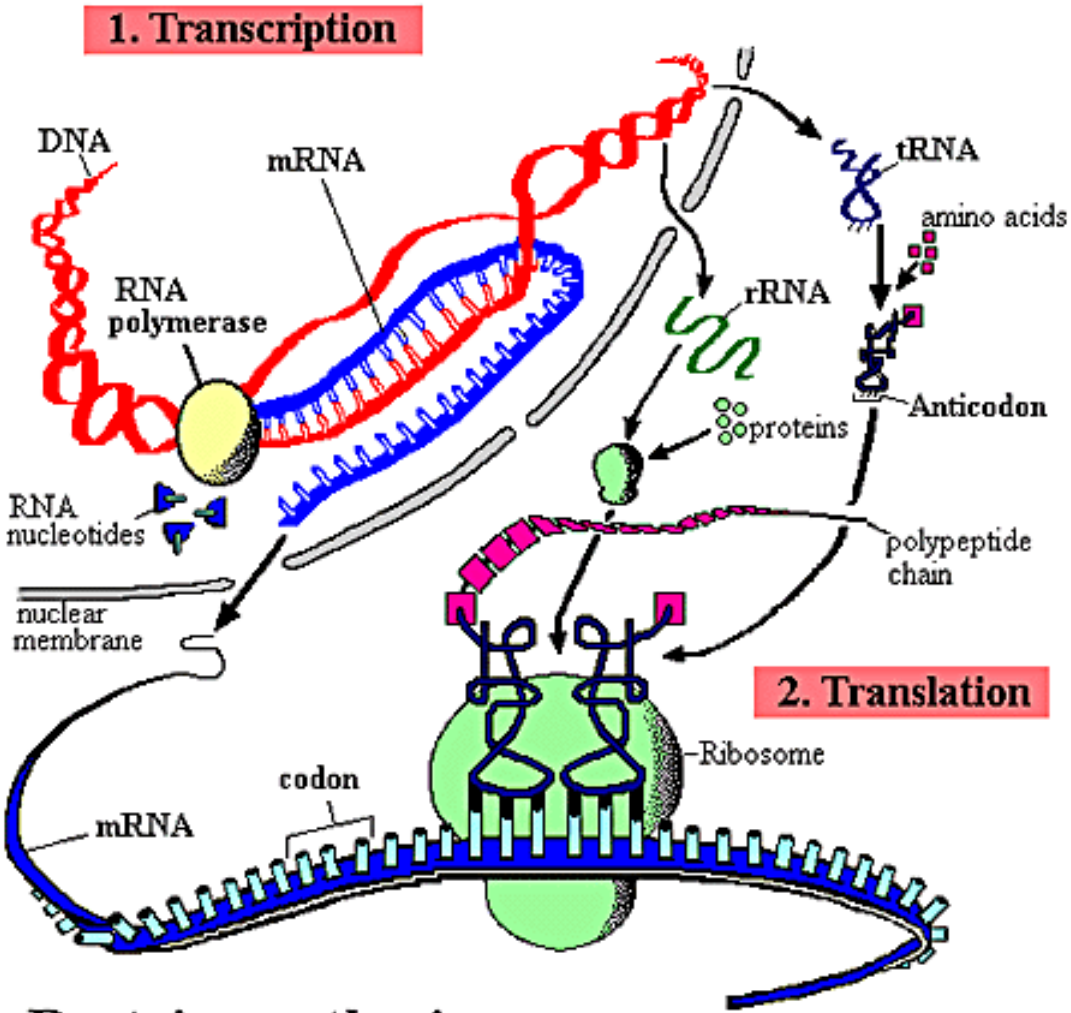
(D) The rest of the general transcription factors as well as the RNA polymerase itself assemble at the promoter. What order?

(E) TFIIH 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

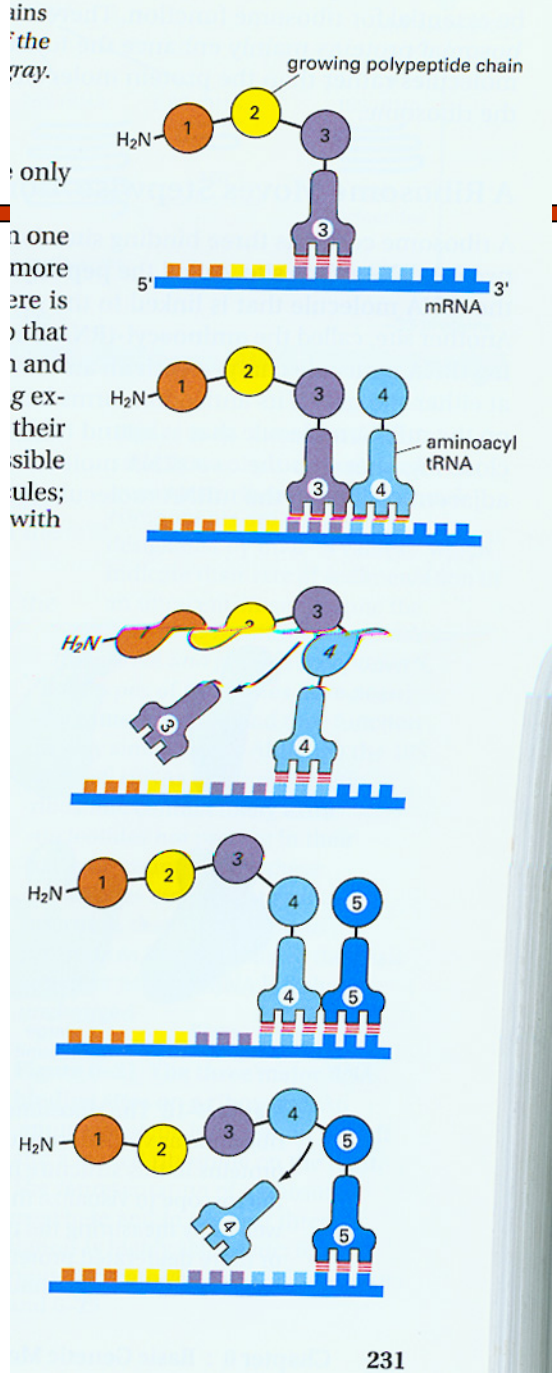
- The general transcription factors have been highly conserved in evolution; some of those from human cells can be replaced in biochemical experiments by the corresponding factors from simple yeasts.

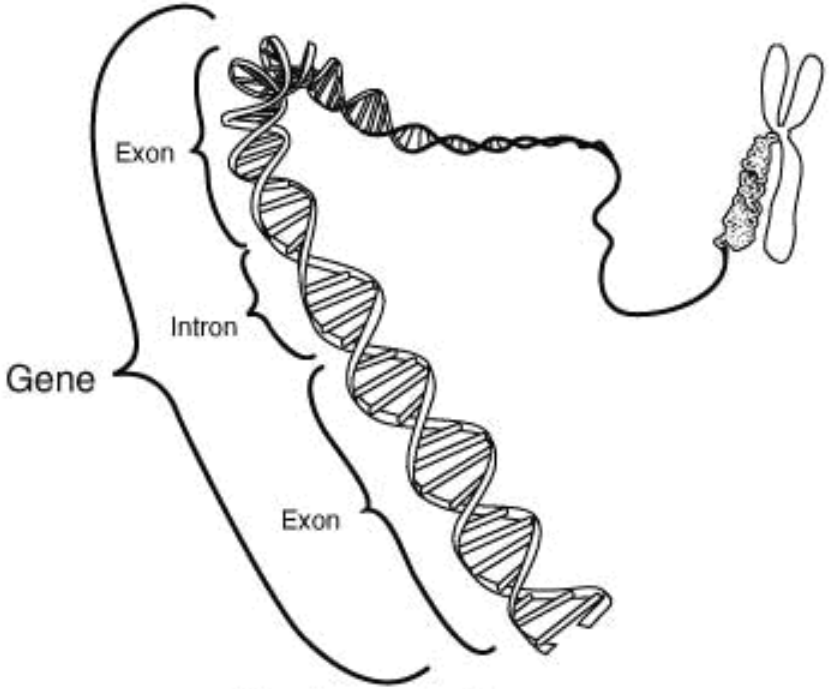
Protein Synthesis

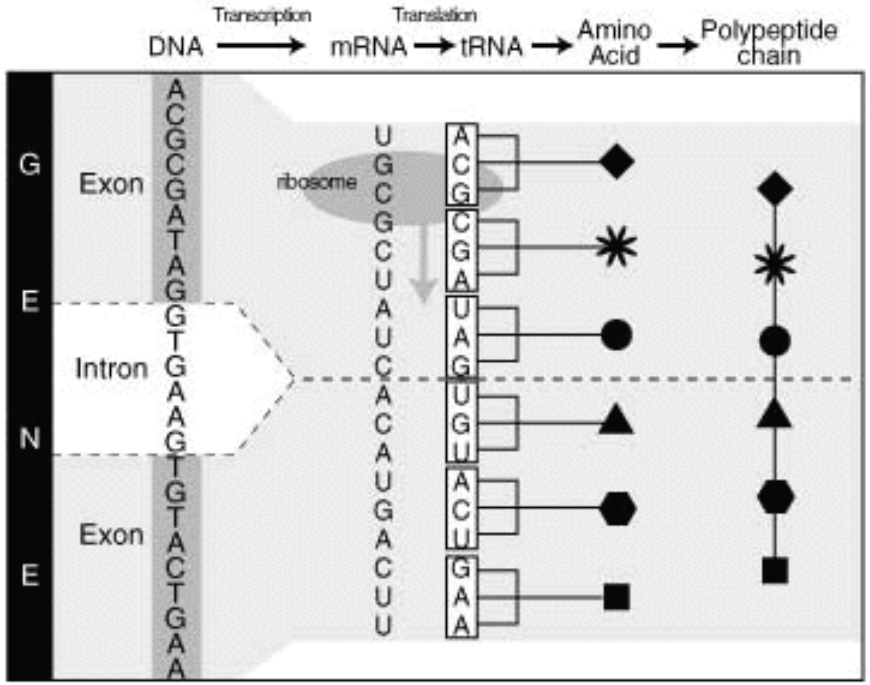


Protein Synthesis:

Incorporation of amino acid into protein




Exon
Intron
Gene
Exon


Transcription
Translation
DNA
mRNA
tRNA
Amino Acid
Polypeptide chain
G
E
N
E
Exon
Intron
Exon
ACGCGATAGGTGAAGTGTACTGAA
ribosome
UGCGCUAUCCACAUGACUU
ACG
CGA
UAG
UGU
ACU
GAA