

## **GENE ACTIVITY**

**Gene structure**

**Transcription**

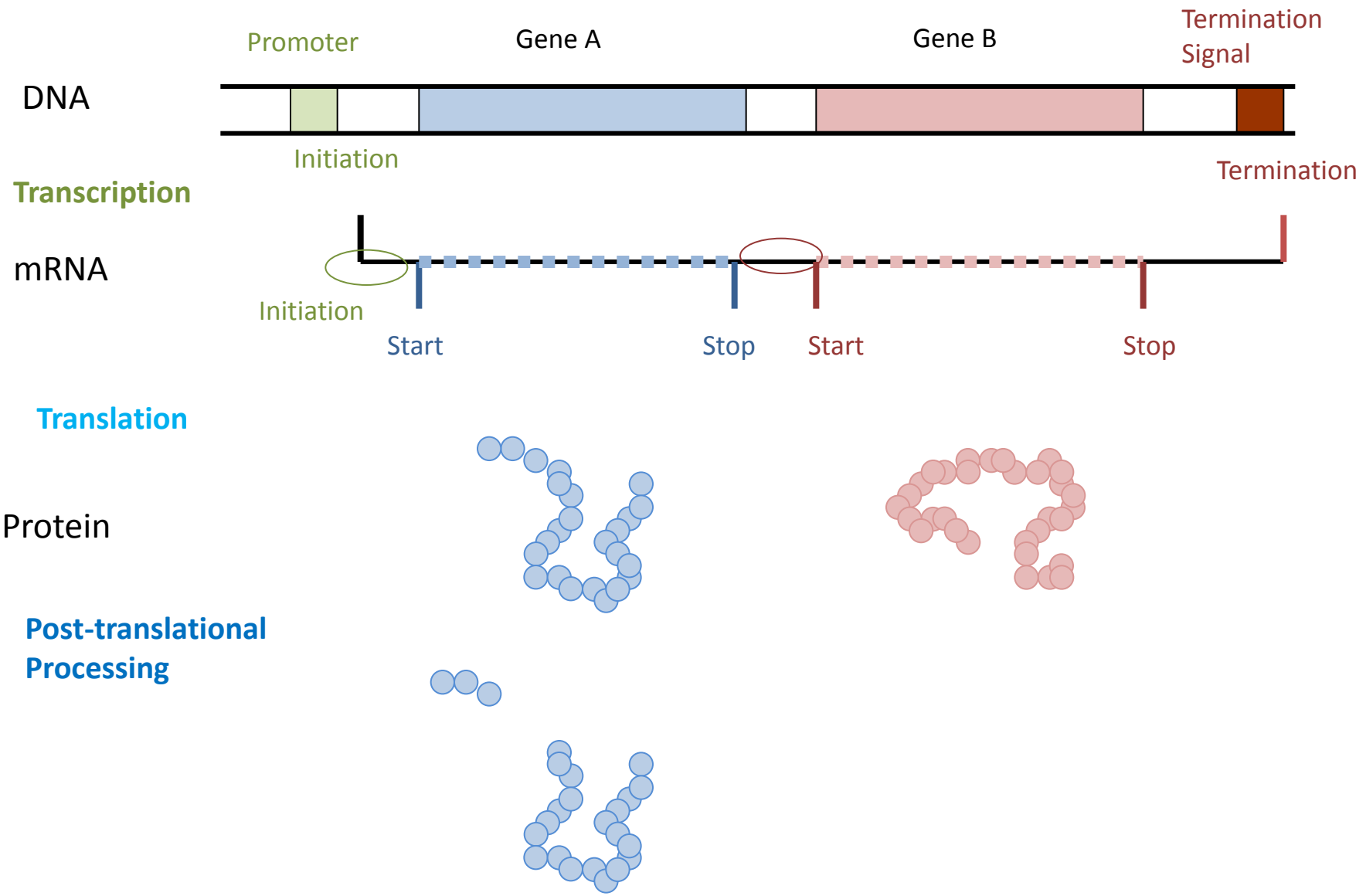
**Transcript processing**

**mRNA transport**

**mRNA stability**

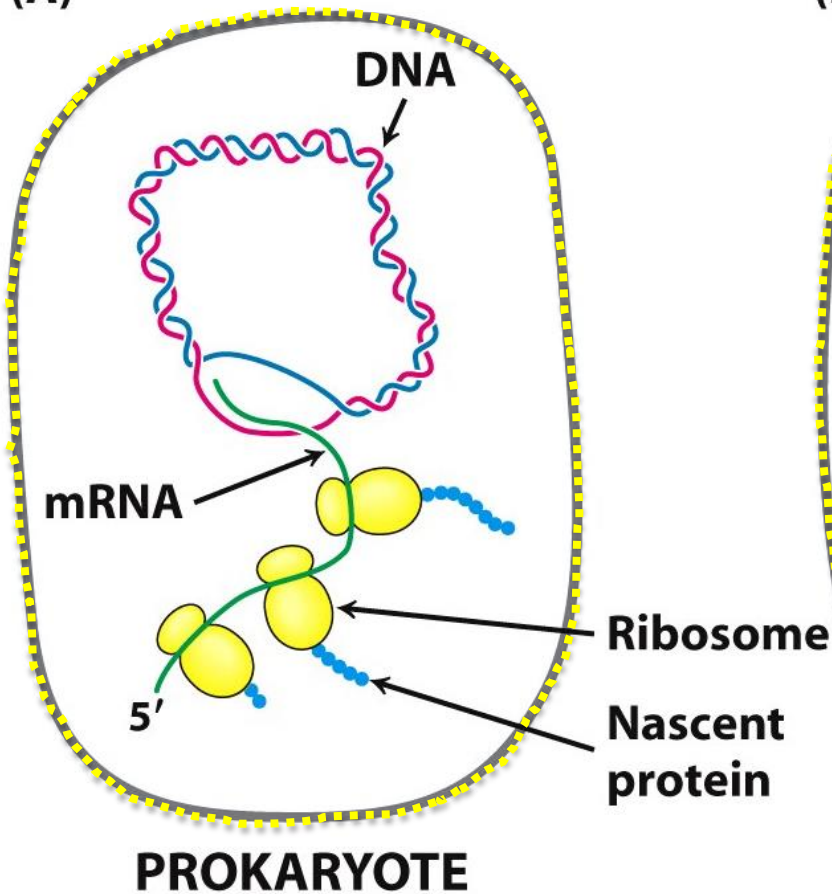
**Translation**

**Posttranslational modifications**

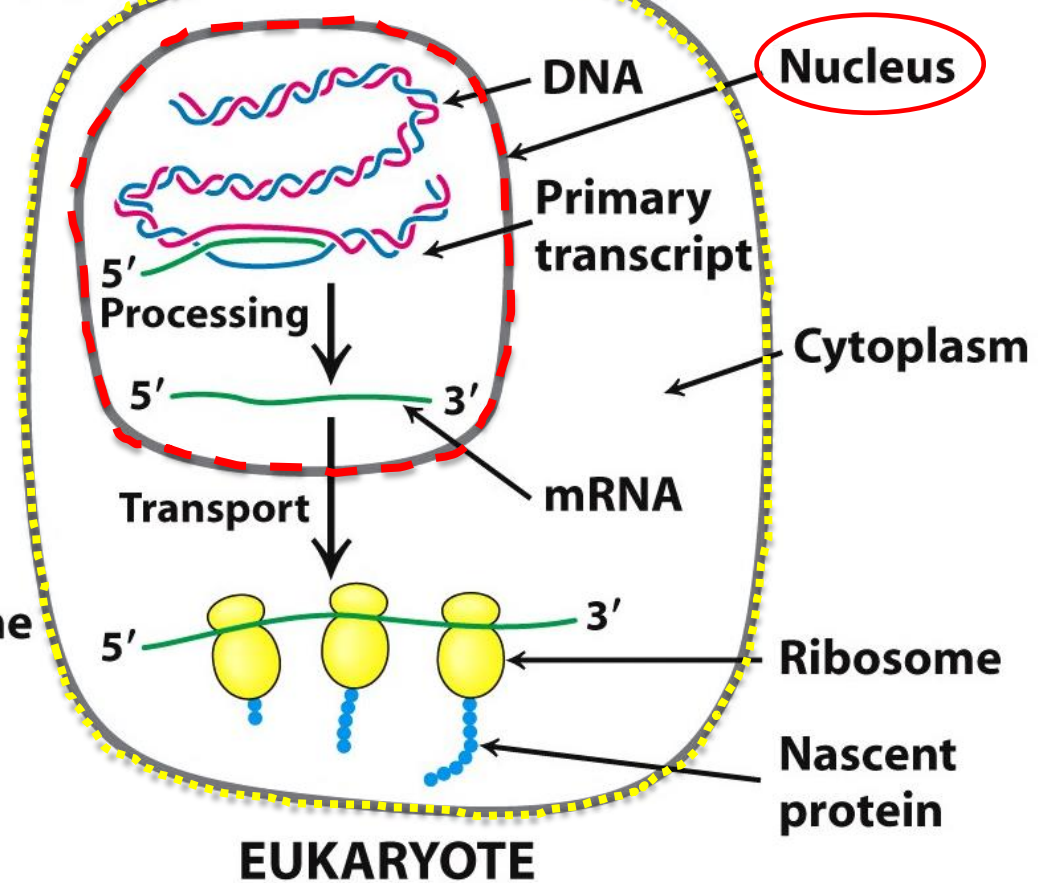


3

(A)

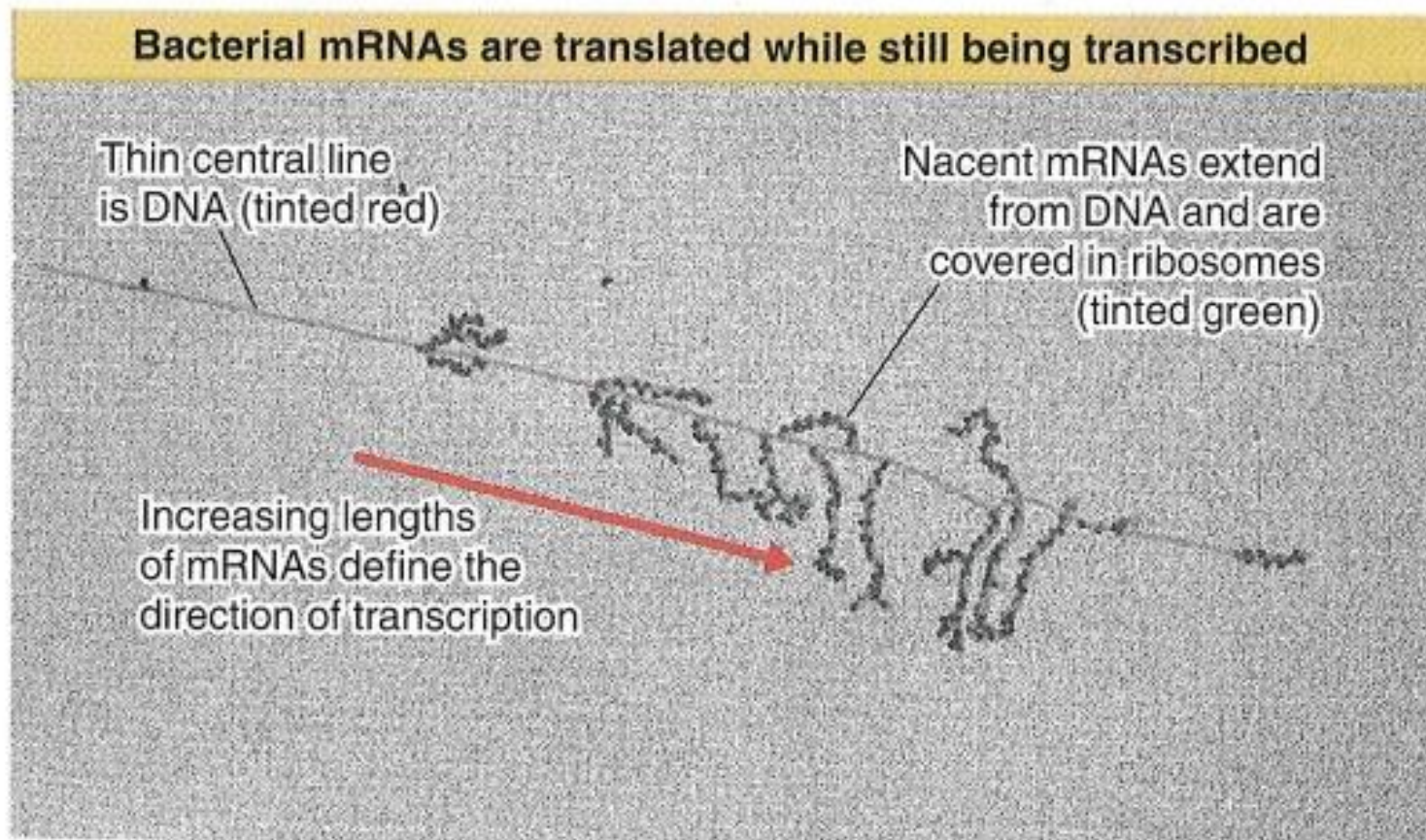


(B)

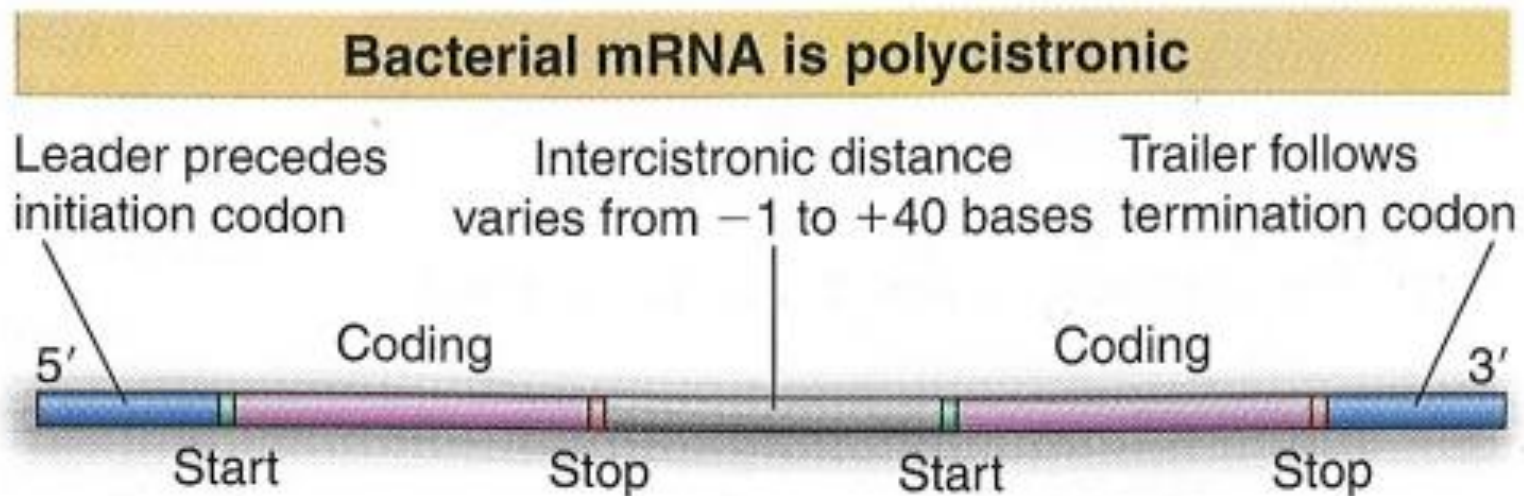
**Figure 29.21***Biochemistry, Seventh Edition*

© 2012 W. H. Freeman and Company

The production of functioning mRNA is very different in prokaryotes and eukaryotes. In prokaryotes, the RNA transcript serves directly as the mRNA and translation begins before transcription is completed; that is, transcription and translation are coupled. In eukaryotes, the primary RNA transcript must be modified in the cell nucleus to form mRNA. Translation takes place only after the completed mRNA is delivered to the cytoplasm.



**Figure 7.14** Transcription units can be visualized in bacteria. Photograph kindly provided by Oscar L. Miller, Department of Biology, University of Virginia.



**Figure 7.15** Bacterial mRNA includes untranslated as well as translated regions. Each coding region has its own initiation and termination signals. A typical bacterial mRNA has several coding regions.

**Tab. 3.1 Die drei RNA-Arten.**

	Größe (ungefähre Angaben)	Funktion
transfer-RNA (tRNA)	80–90 Nucleotide	Übertragung von Aminosäuren zum Proteinsynthese-Apparat der Zelle
ribosomale RNA (rRNA)	4 Arten (bei Eukaryoten) mit je ca. 120, 150, 1700, 3500 Nucleotiden	Struktur und Funktionselemente der Ribosomen
messenger-RNA (mRNA)	sehr verschieden (einige 100 bis über 10000 Nucleotide)	die Boten-( <i>messenger</i> -)RNA überbringt dem Proteinsynthese-Apparat eine Abschrift des Gens

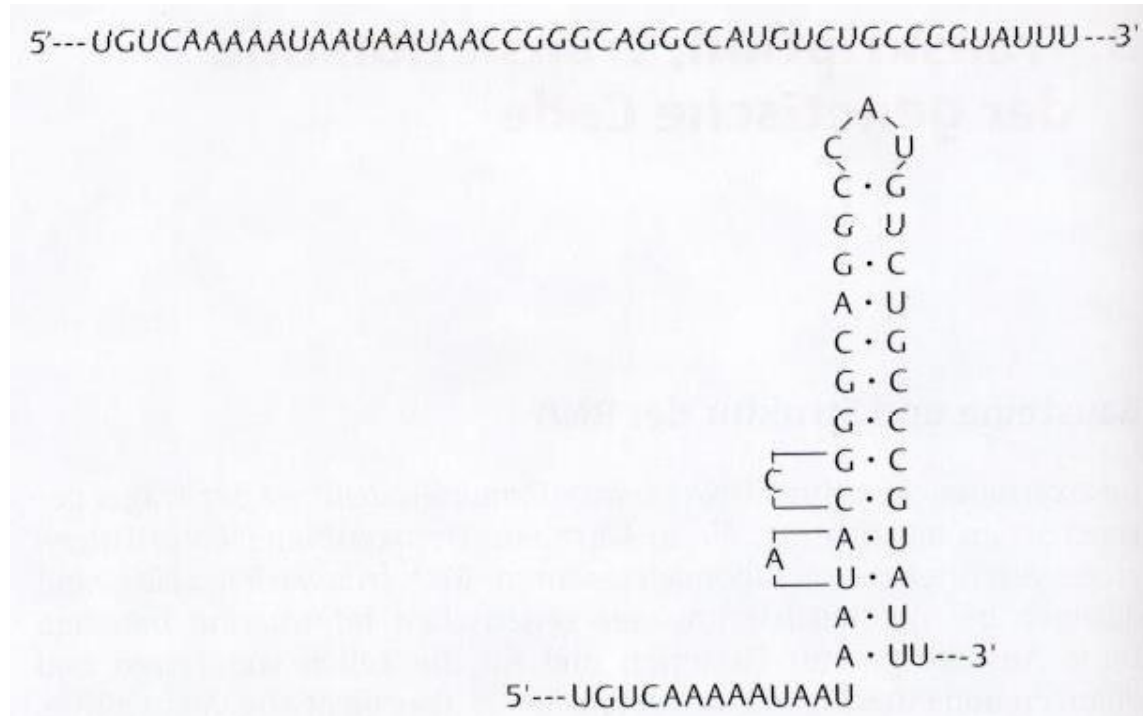
## The 3 RNA types

	Size (approximate)	Function
transfer-RNA (tRNA)	80-90 nucleotides	Transfer of amino acids to the protein synthesis apparatus of the cell
Ribosomal RNA (rRNA)	4 types (in eukaryotes) with each ca. 120, 150, 1700 and 3500 nucleotides	Structural and functional elements of the ribosomes
messenger-RNA (mRNA)	Very different (from several 100 to more than 10000 nucleotides)	The mRNA delivers a gene copy to the protein synthesis apparatus of the cell

Types of RNA	Characteristics and key functions
messenger RNA (mRNA)	<ul style="list-style-type: none"><li>• varies in length, depending on the gene that has been copied</li><li>• acts as the intermediary between DNA and the ribosomes</li><li>• translated into protein by ribosomes</li><li>• RNA version of the gene encoded by DNA</li></ul>
transfer RNA (tRNA)	<ul style="list-style-type: none"><li>• functions as the delivery system of amino acids to ribosomes as they synthesize proteins</li><li>• very short, only 70–90 base pairs long</li></ul>
ribosomal RNA (rRNA)	<ul style="list-style-type: none"><li>• binds with proteins to form the ribosomes</li><li>• varies in length</li></ul>



## RNA is single stranded but is organized partly as ds RNA by internal base pairing

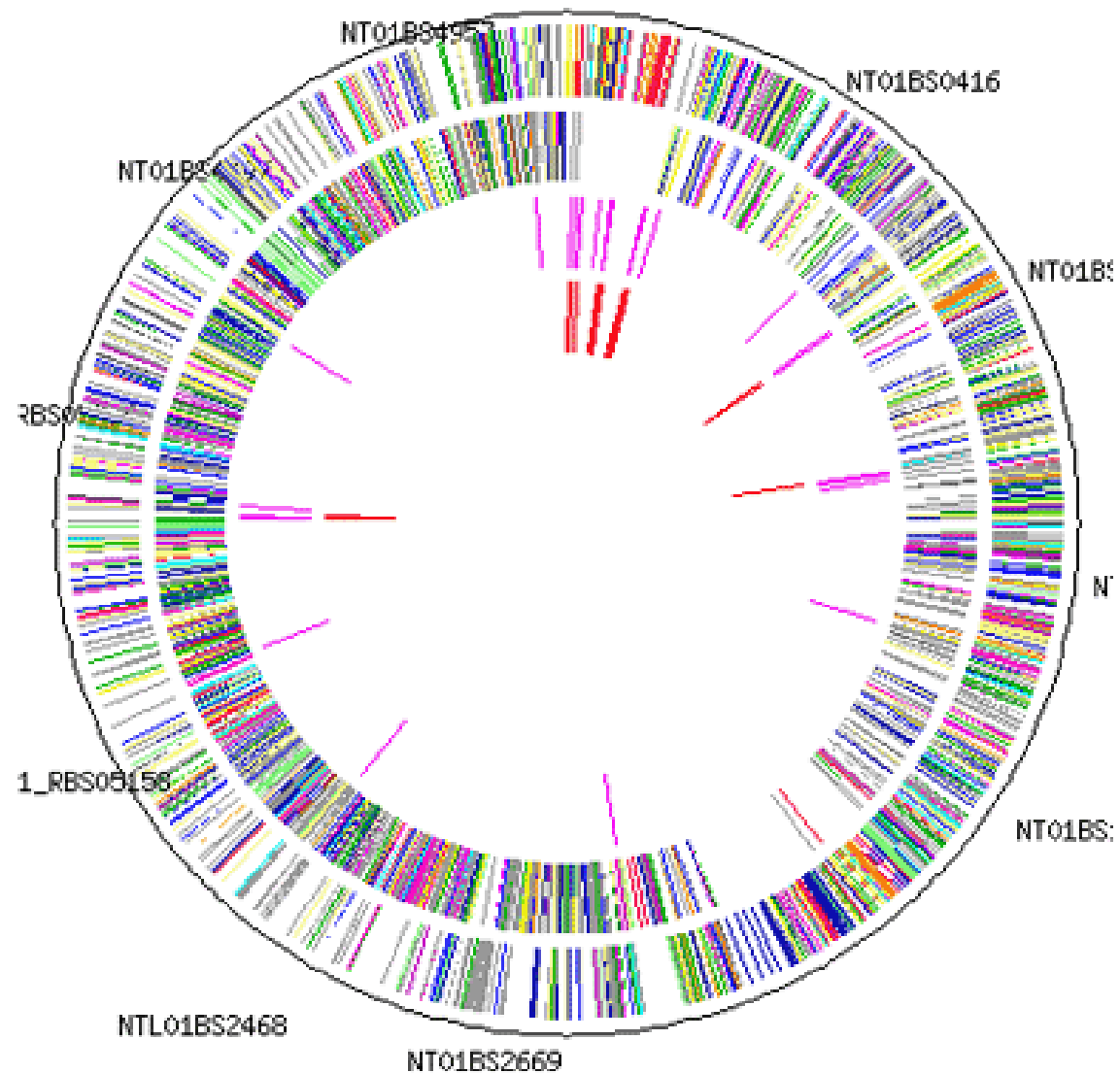


### Loop (secondary structure) formation in the *E.coli* mRNA.

The mRNA has a length of several thousand nucleotides, only a short part is shown here. This part contains complementary nucleotide sequences that can combine to a double stranded region so that a loop is created at this site. Cytosine pairs with Guanine and Uracil with Adenine.

10 Genome map of *Bacillus subtilis*

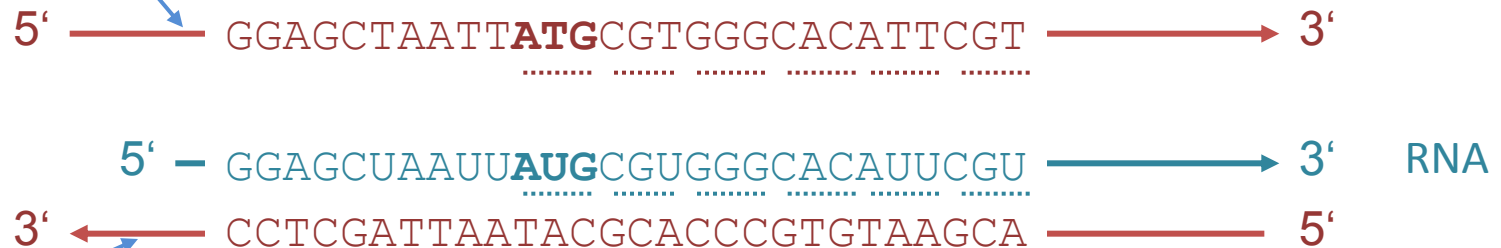
Genes are transcribed from both strands



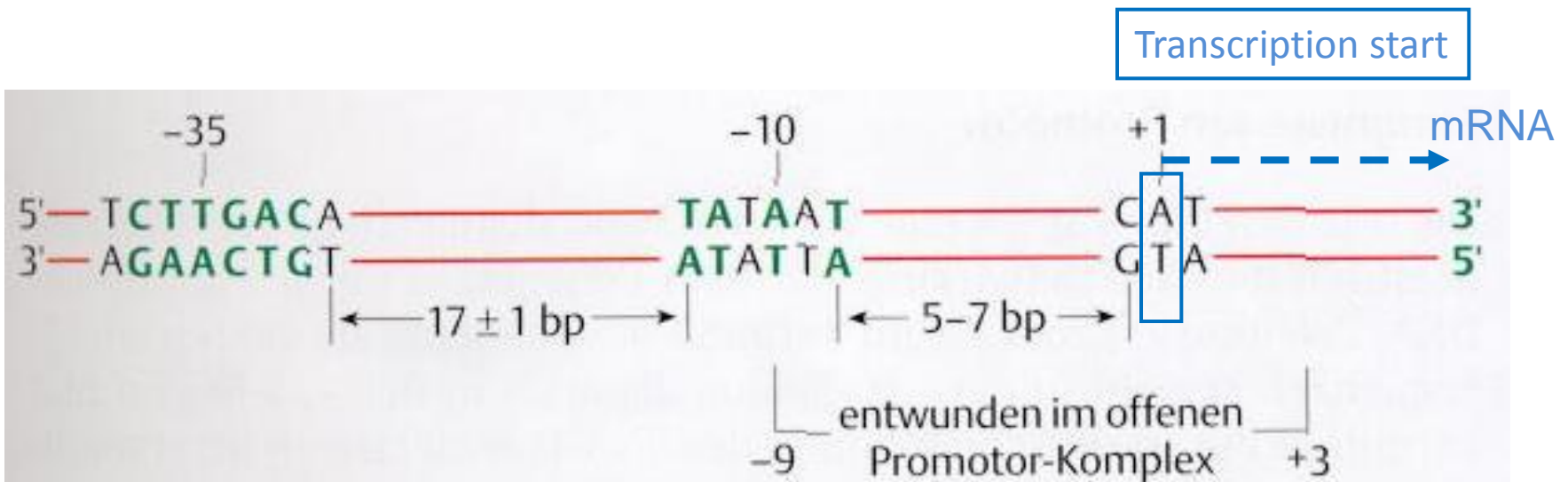
## Transcription: Only one strand is transcribed



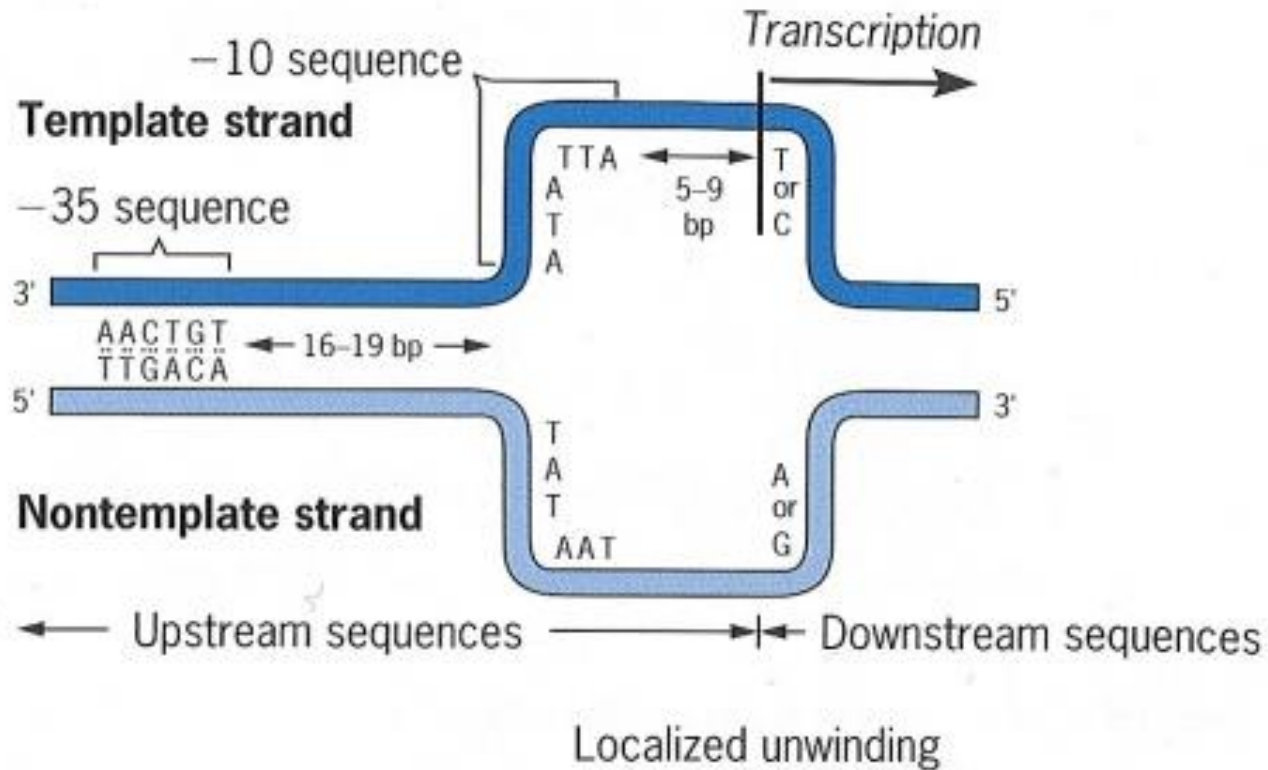
Codon sequence of encoded protein is reflected in complementary strand



Template for transcription



**Abb. 3.5 Ein Musterpromotor des *E. coli*-Genoms.** Der Abstand zwischen dem Transkriptionsstart und dem ersten Nucleotid der -10-Region beträgt 5-7 Basenpaare (bp); der Abschnitt zwischen der -10-Region und der -35-Region  $17 \pm 1$  bp. Der untere der beiden DNA-Stränge ist der transkribierte „codogene“ oder Sinnstrang, der obere der nichttranskribierte Gegensinn-Strang [nach 13].



**Figure 11.9** ▶ Structure of a typical promoter in *E. coli*. RNA polymerase binds to the -35 sequence of the promoter and initiates unwinding of the DNA strands at the AT-rich -10 sequence. Transcription begins within the transcription bubble at a site five to nine base pairs beyond the -10 sequence.

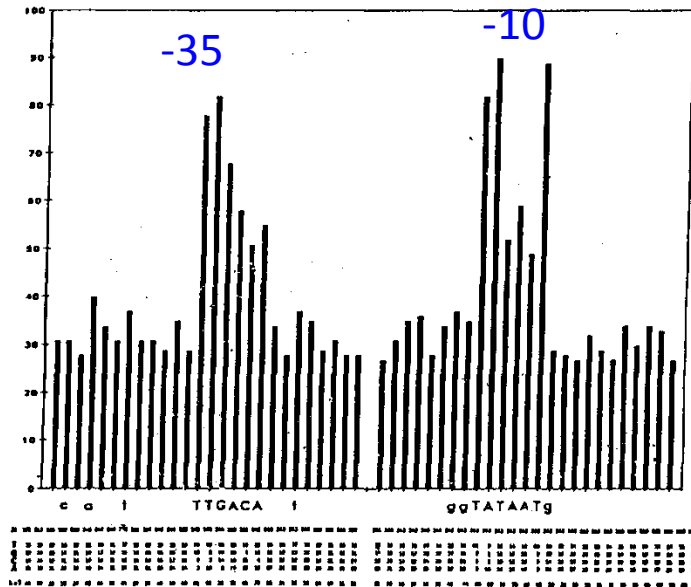
Gene	Factor	Use
<i>rpoD</i>	$\sigma^{70}$	most required functions
<i>rpoS</i>	$\sigma^S$	stationary phase/some stress responses
<i>rpoH</i>	$\sigma^{32}$	heat shock
<i>rpoE</i>	$\sigma^E$	periplasmic/extracellular proteins
<i>rpoN</i>	$\sigma^{54}$	nitrogen assimilation
<i>rpoF</i>	$\sigma^F$	flagellar synthesis/chemotaxis
<i>fecl</i>	$\sigma^{fecl}$	iron metabolism/transport

Fig. 19.37: In addition to  $\sigma^{70}$ , *E. coli* has several sigma factors that are induced by particular environmental conditions.

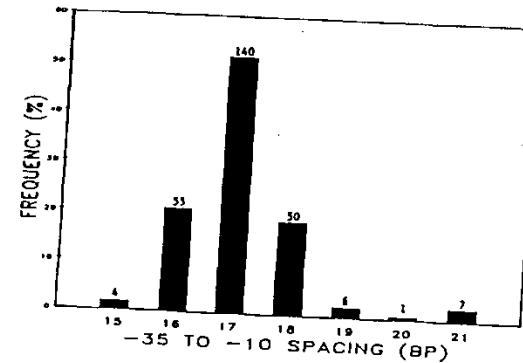
Subunit/gene	Size (# aa)	Approx. # of promoters	Promoter sequence recognized
Sigma 70 ( <i>rpoD</i> )	613	1000	TTGACA-16 to 18-bp-TATAAT
Sigma 54 ( <i>rpoN</i> )	477	5	CTGGNA-6 bp-TTGCA
Sigma S ( <i>rpoS</i> )	330	100	TTGACA-16 to 18-bp-TATAAT
Sigma 32 ( <i>rpoH</i> )	284	30	CCCTTGAA-13 to 15-bp- CCCGATNT
Sigma F ( <i>rpoF</i> )	239	40	CTAAA-15 bp-GCCGATAA
Sigma E ( <i>rpoE</i> )	202	20	GAA-16 bp-YCTGA
Sigma Fecl ( <i>fecl</i> )	173	1–2	?

Fig. 19.15: *E. coli* sigma factors recognize promoters with different consensus sequences.

## Effect of mutations on promoter function

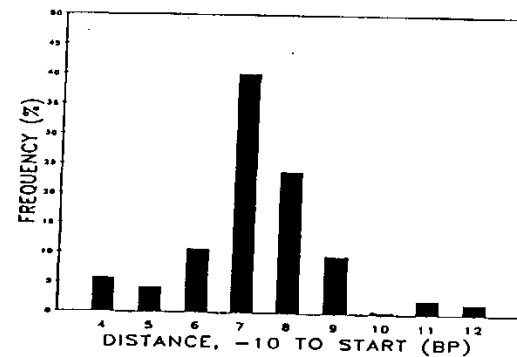


**Figure 1.** Base distribution of 263 analyzed promoters from Table 1. (a) Frequency histogram of the most highly conserved base on the non-template strand from 12 bp upstream of the -35 hexamer to 11 bp downstream of the -10 hexamer. Highly conserved (upper case) and weakly conserved (lower case) bases, as defined in the text, are shown below the histogram. (b) Frequency of bases (T, C, G, A and T+A) in aligned promoters as a percentage of total number of bases (N) at each position.



**Figure 2.** Distribution of promoters with 15-21 bp separating the -35 and -10 hexamers. The number of promoters in each group is indicated on top of the bars.

### Spacing between -35 and -10 region



**Figure 3.** Distribution of promoters with transcription start points initiating 4-12 bases downstream of the -10 hexamer. Only promoters with uniquely defined start points are included in this analysis.

### Spacing between -10 region and transcription start



Gene	Product	Functions
<i>rpoA</i>	2 $\alpha$ subunits (37 kD each)	enzyme assembly promoter recognition binds some activators
<i>rpoB</i>	$\beta$ subunit (151 kD)	catalytic center
<i>rpoC</i>	$\beta'$ subunit (155 kD)	
<i>rpoD</i>	$\sigma$ subunit (18–70 kD)	promoter specificity
<i>rpoZ</i>	$\omega$ subunit (10 kD)	
<i>E. coli</i> enzyme = 460 kD		

Fig. 19.7: Eubacterial RNA polymerases have five types of subunits.

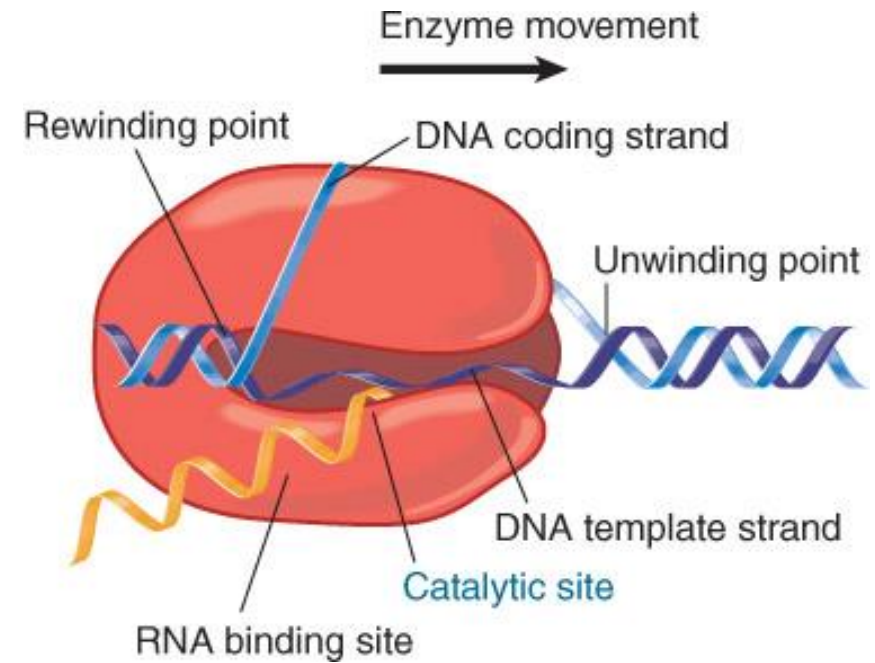
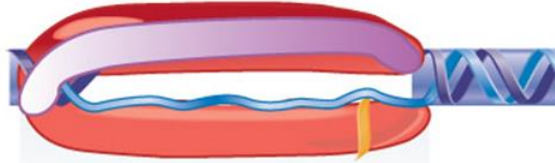


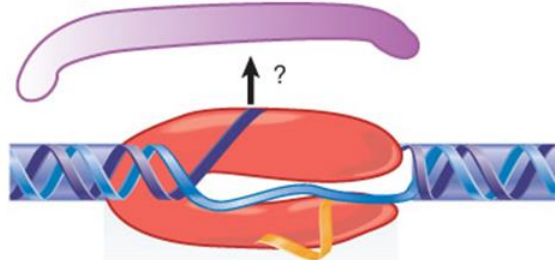
Fig. 19.5: During transcription, the bubble is maintained within bacterial RNA polymerase, which unwinds and rewinds DNA and synthesizes RNA.

Initiation complex contains sigma and covers ~75 bp



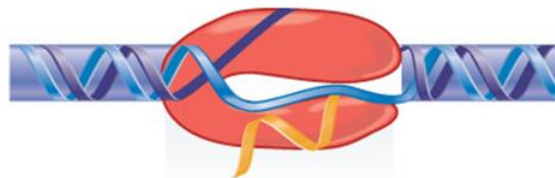
-50 -40 -30 -20 -10 1 +10 +20 +30

Initial elongation complex forms at 10 bases, may lose sigma, and loses contacts from -35 to -55



-50 -40 -30 -20 -10 1 +10 +20 +30

General elongation complex forms at 15–20 bases and covers 30–40 bp



-50 -40 -30 -20 -10 1 +10 +20 +30

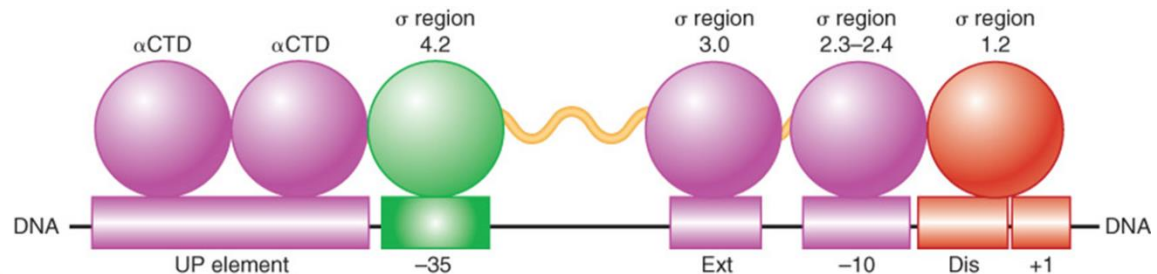


Fig. 19.14: DNA elements and RNA polymerase modules that contribute to promoter recognition by sigma factor.

Fig. 19.13: RNA polymerase initially contacts the region from -55 to +20. When sigma dissociates, the core enzyme contracts to -30; when the enzyme moves a few base pairs, it becomes more compactly organized into the general elongation complex.

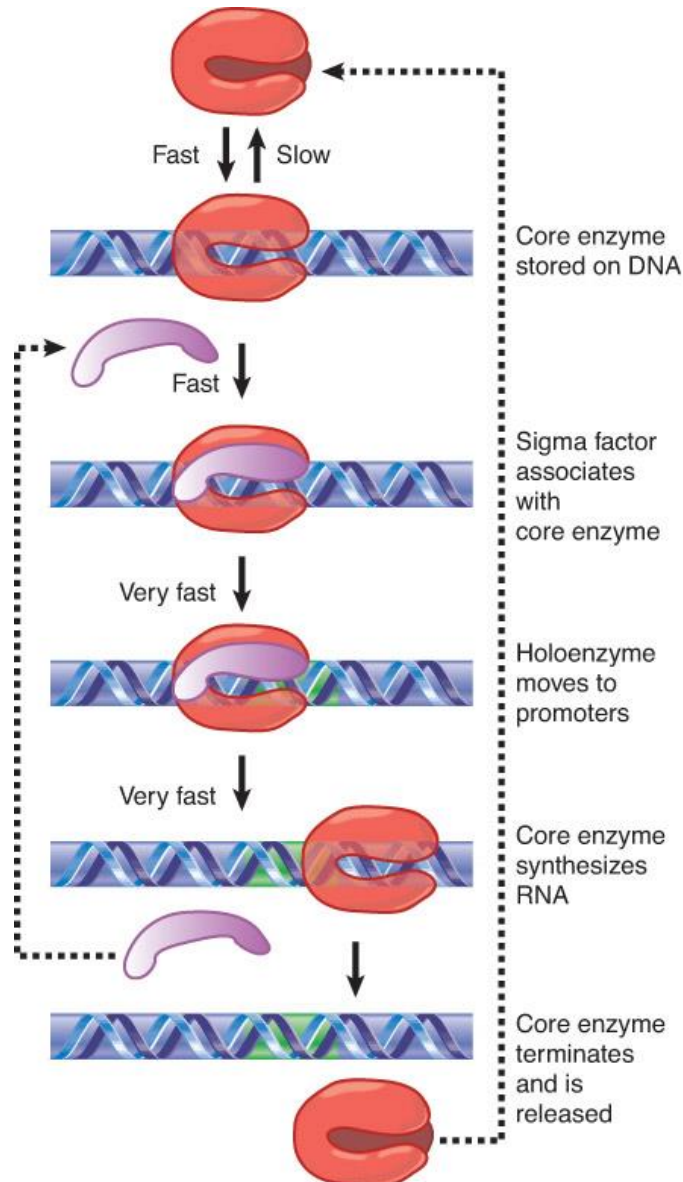
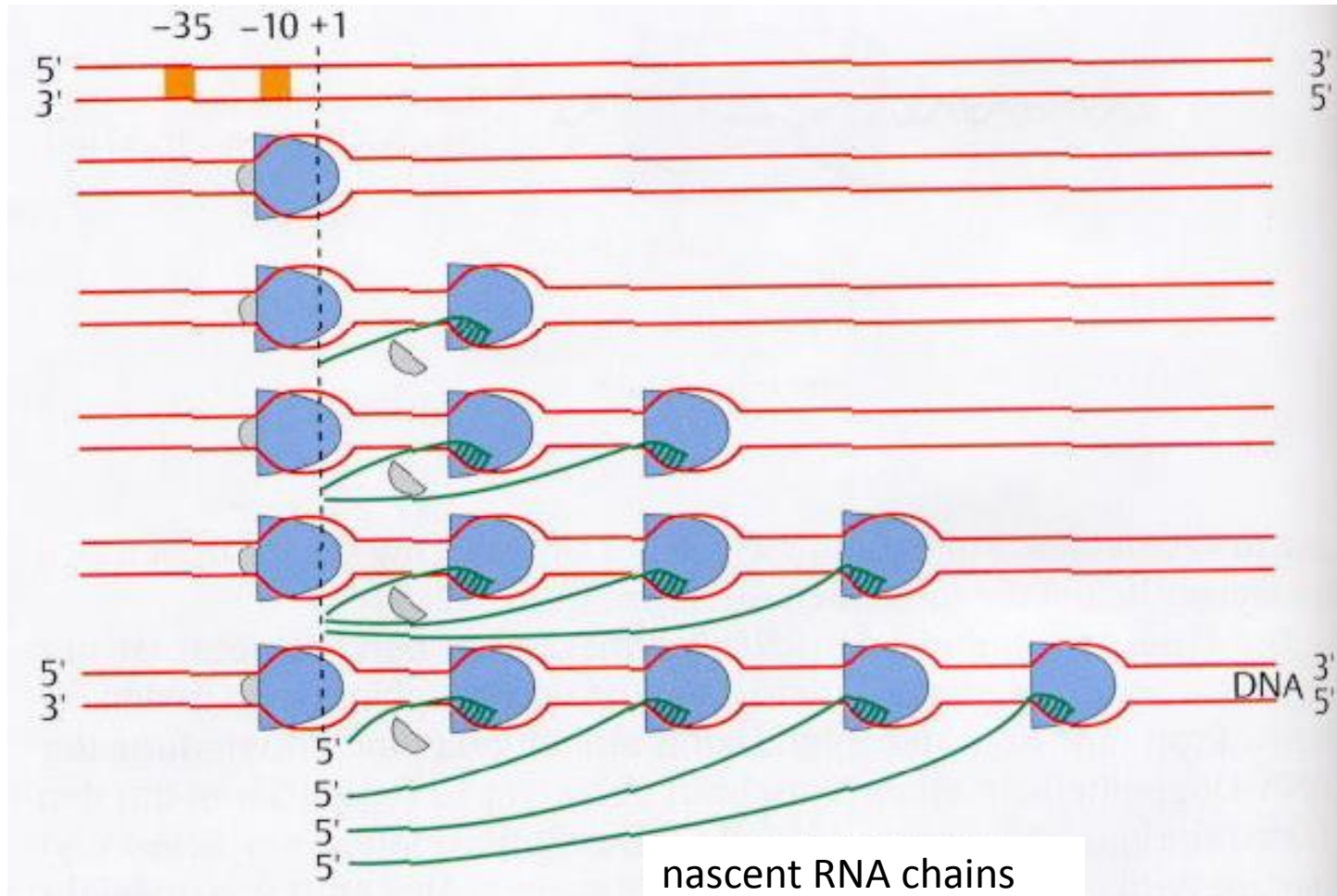
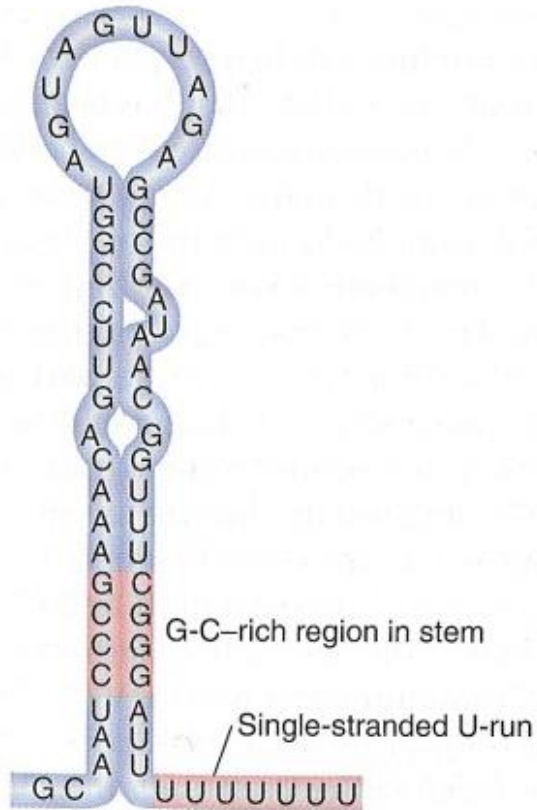


Fig. 19.24: Sigma factor and core enzyme recycle at different points in transcription.

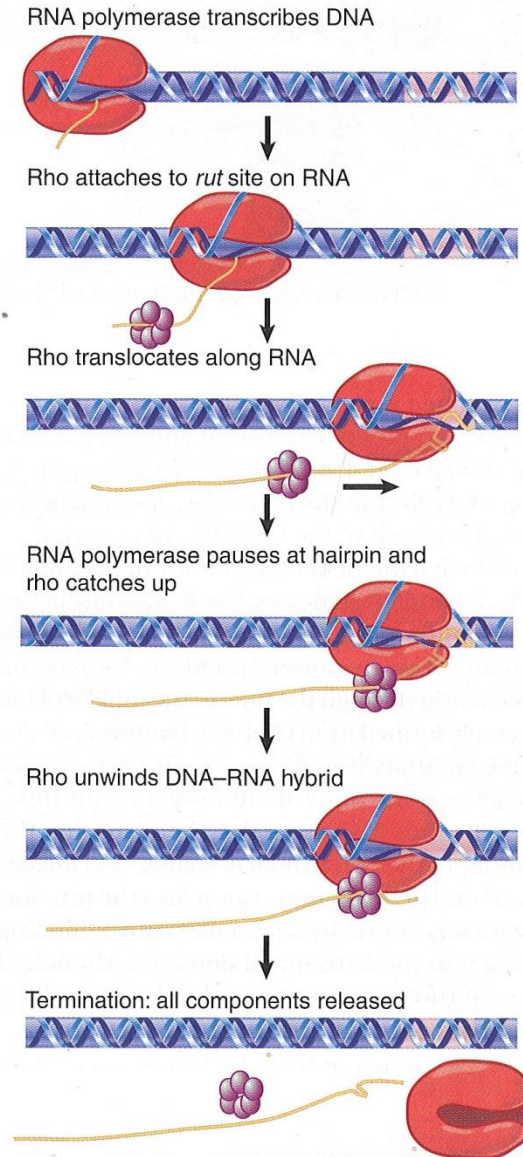


**Scheme of transcription.** Transcription starts with an „open“ promoter complex. The holoenzyme causes the unwinding of the region close to the transcription start point (+1). After a few polymerisation steps, the sigma factor leaves the core enzyme which continues its way along the transcribed DNA strand. The released promoter is re-occupied. In parallel, several RNA polymerases are occupied with transcription

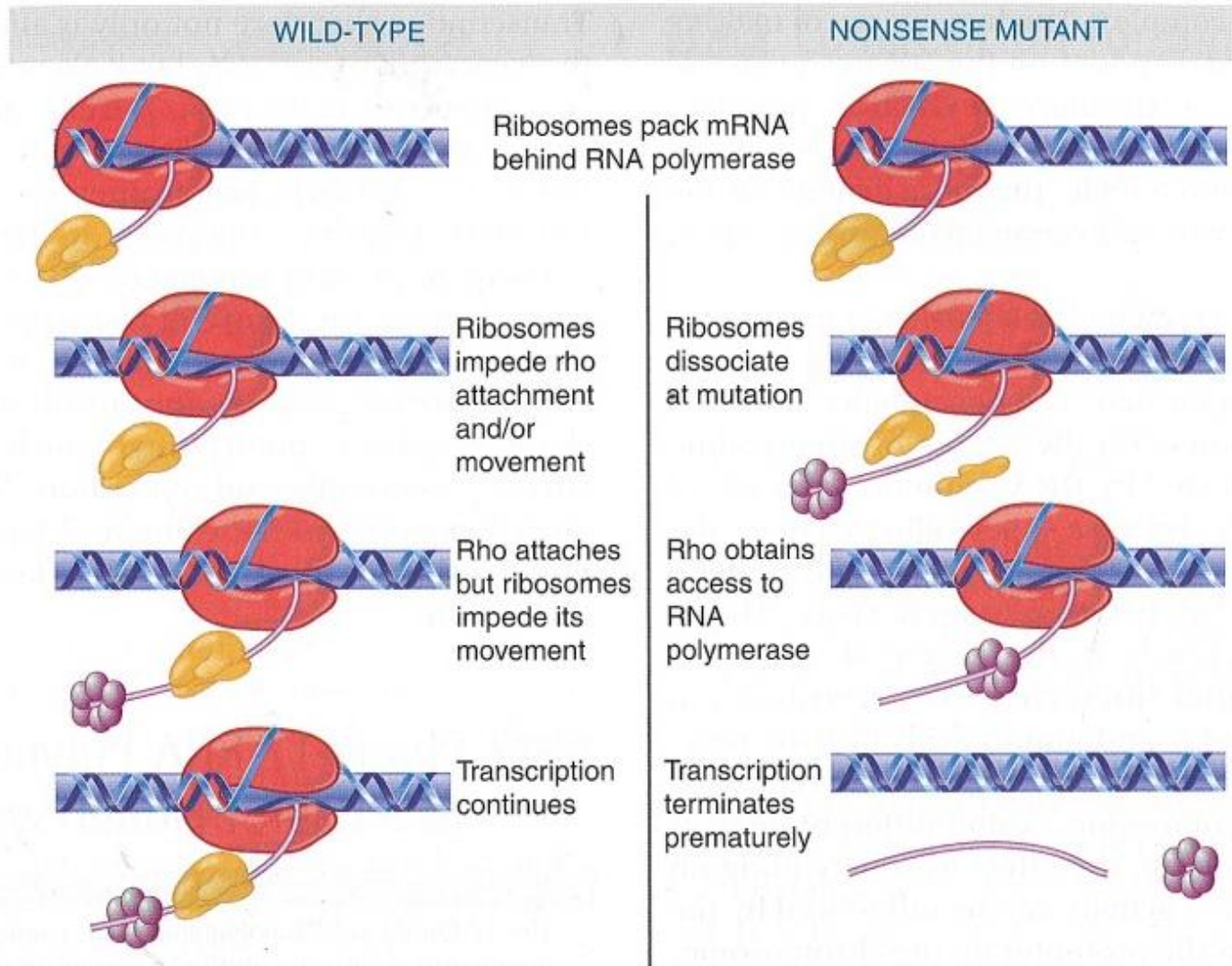
11.11.14



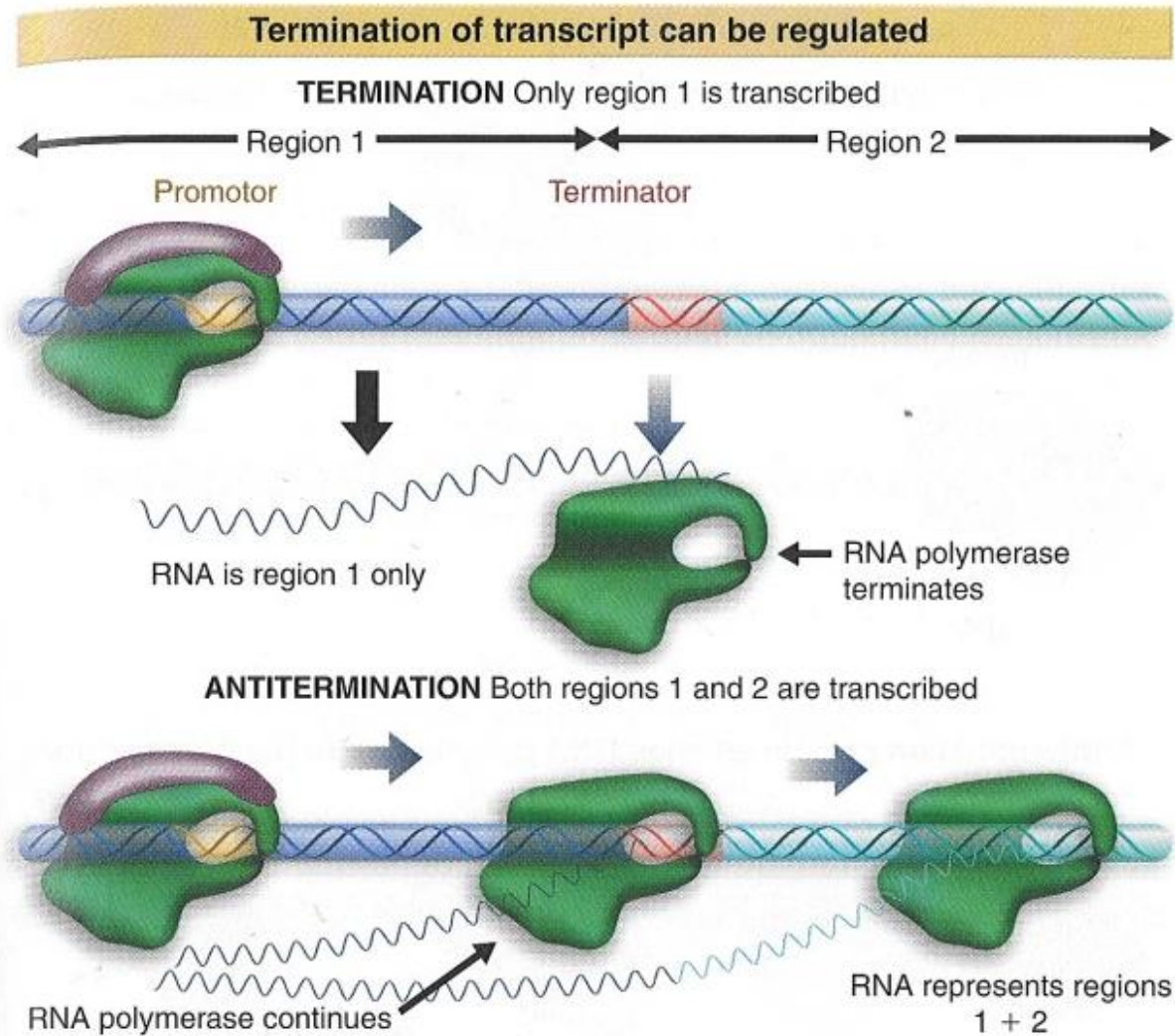
**FIGURE 19.29** Intrinsic terminators include palindromic regions that form hairpins varying in length from 7 to 20 bp. The stem-loop structure includes a G-C-rich region and is followed by a run of U residues.



**FIGURE 19.30** Rho factor binds to RNA at a *rut* site and translocates along RNA until it reaches the RNA-DNA hybrid in RNA polymerase, where it releases the RNA from the DNA.

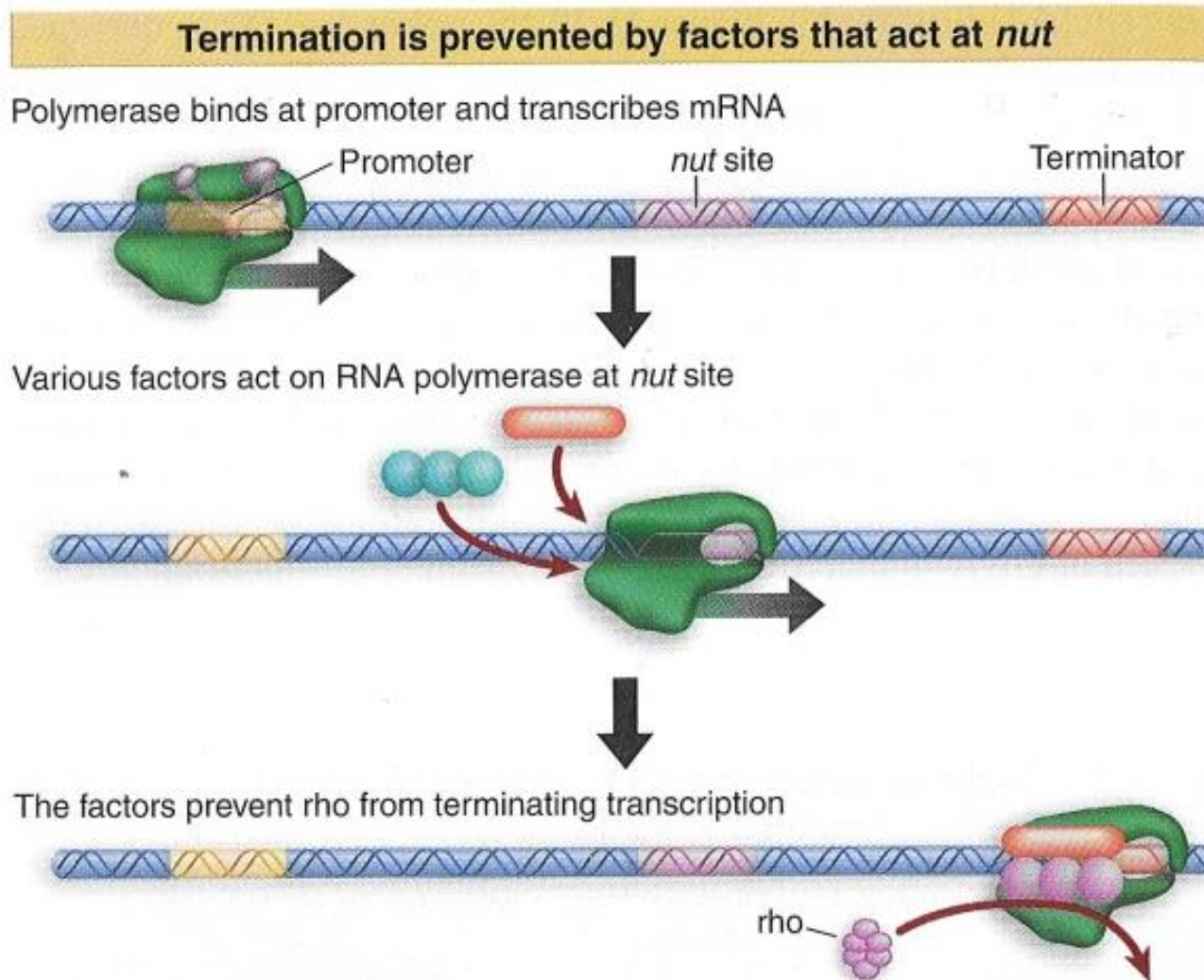


**FIGURE 19.33** The action of rho factor may create a link between transcription and translation when a rho-dependent terminator lies soon after a nonsense mutation.

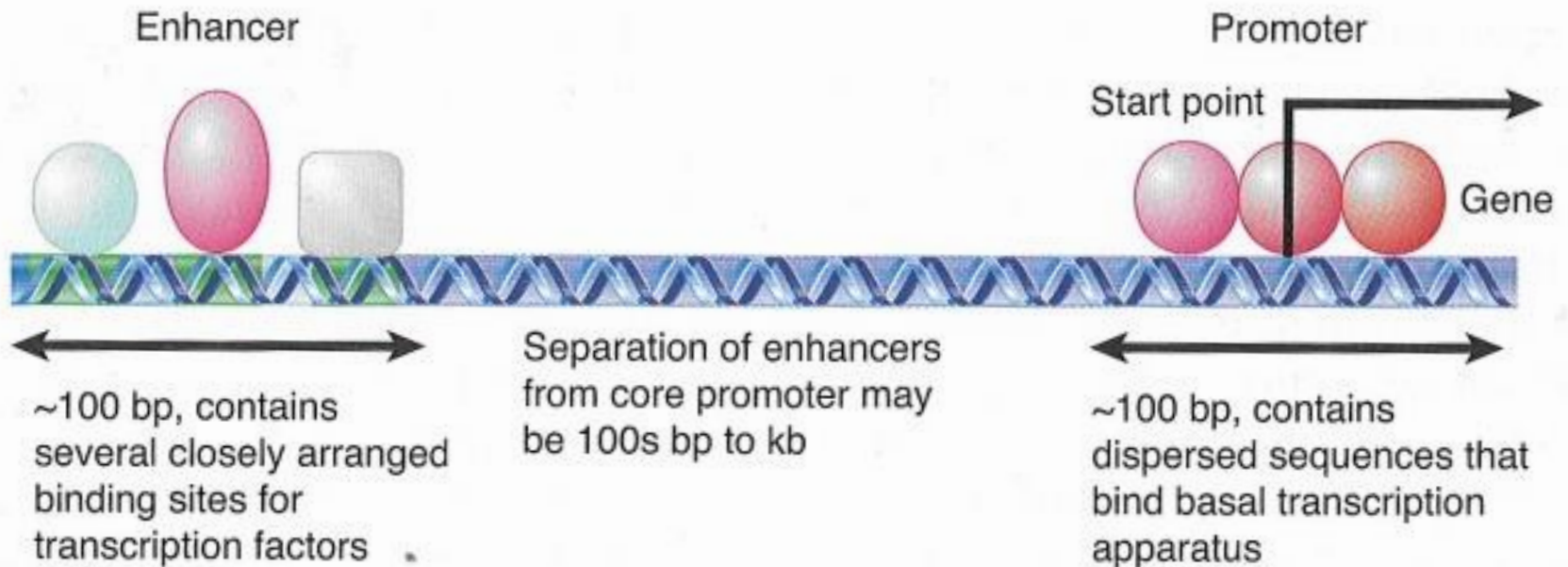


**Figure 11.36** Antitermination can control transcription by determining whether RNA polymerase terminates or reads through a particular terminator into the following region.



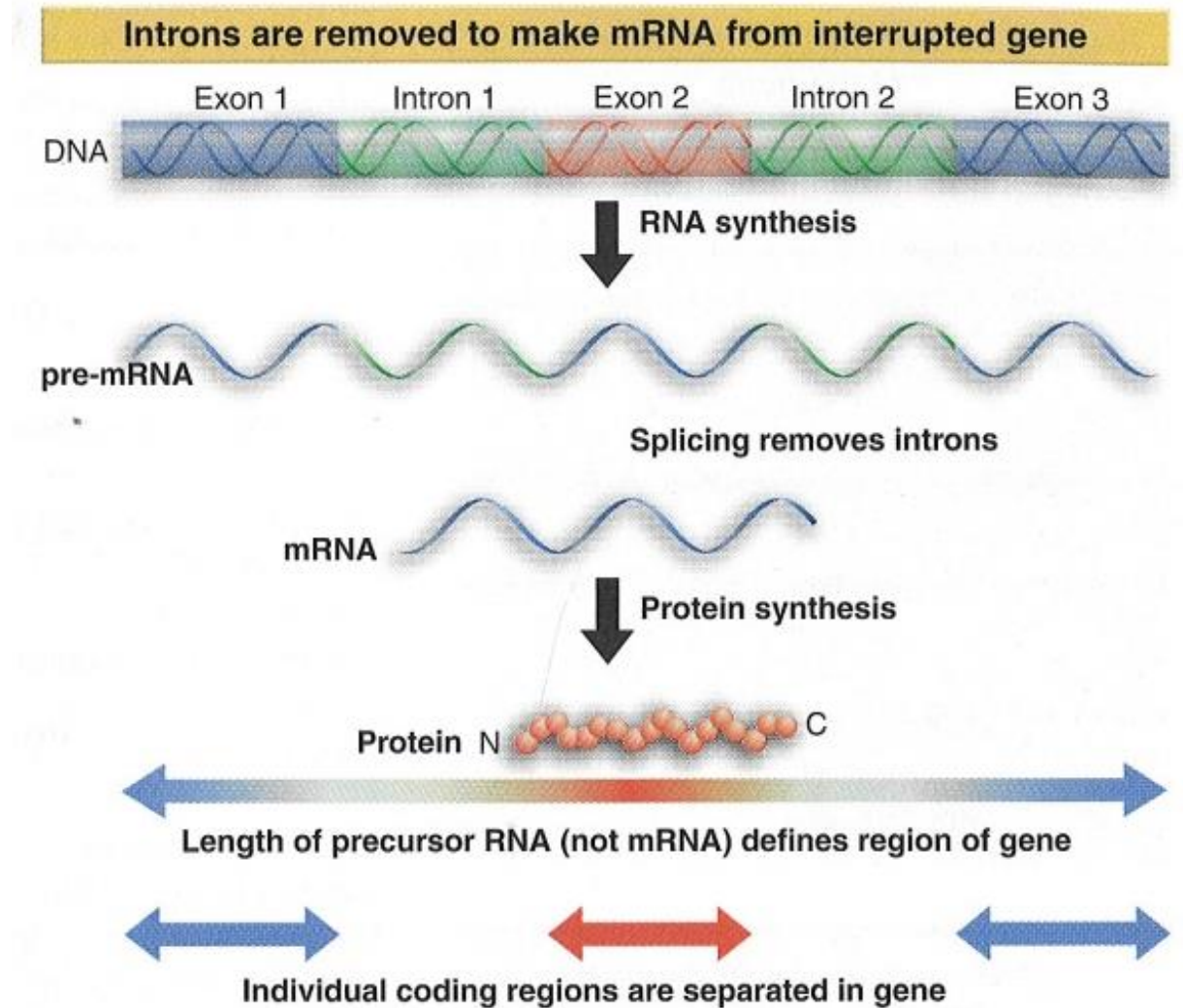


**Figure 11.39** Ancillary factors bind to RNA polymerase as it passes the *nut* site. They prevent rho from causing termination when the polymerase reaches the terminator.



**FIGURE 20.1** A typical gene transcribed by RNA polymerase II has a promoter that extends upstream from the site where transcription is initiated. The promoter contains several short-sequence ( $\sim 10$  bp) elements that bind transcription factors, dispersed over  $\sim 100$  bp. An enhancer containing a more closely packed array of elements that also bind transcription factors may be located several hundred bp to several kb distant. (DNA may be coiled or otherwise rearranged so that transcription factors at the promoter and at the enhancer interact to form a large protein complex.)

## Gene Expression in Eukaryotes -- Introns



**Figure 3.1** Interrupted genes are expressed via a precursor RNA. Introns are removed when the exons are spliced together. The mRNA has only the sequences of the exons.

**Table 9–1** The Size of Some Human Genes in Thousands of Nucleotides

	<b>kb</b> <b>Gene Size</b>	<b>kb</b> <b>mRNA Size</b>	<b>Number of</b> <b>Introns</b>
$\beta$ -Globin	1.5	0.6	2
Insulin	1.7	0.4	2
Protein kinase C	11	1.4	7
Albumin	25	2.1	14
Catalase	34	1.6	12
LDL receptor	45	5.5	17
Factor VIII	186	9	25
Thyroglobulin	300	8.7	36
Dystrophin*	more than 2000	17	more than 50

The size specified here for a gene includes both its transcribed portion and nearby regulatory DNA sequences. (Compiled from data supplied by Victor McKusick.)

\*An altered form of this gene causes Duchenne muscular dystrophy.

## mRNA Synthesis in Eukaryote is a complex Process

Transcription Initiation

Transcription Elongation

5' Transcript Processing (CAP)

Transcription Termination

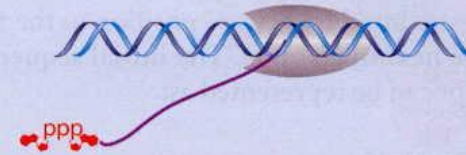
3' Transcript Processing

Intron Splicing

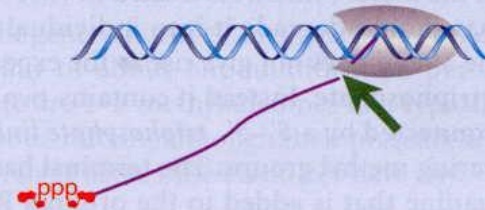
Transport into Cytoplasm

**Figure 5.16** Overview: expression of mRNA in animal cells requires transcription, modification, processing, nucleocytoplasmic transport, and translation.

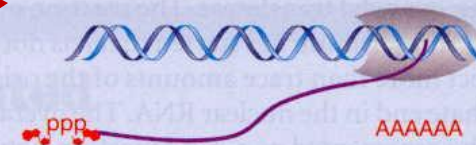
<1 min Transcription starts: 5' end is modified



6 min 3' end of mRNA is released by cleavage



20 min 3' end is polyadenylated



25 min mRNA is transported to cytoplasm



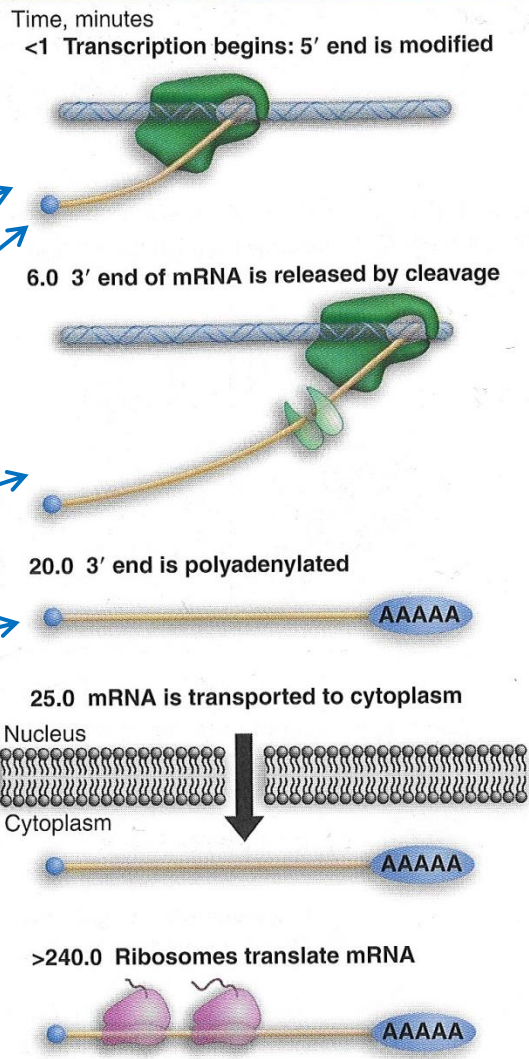
> 4 hr Ribosomes translate mRNA



# mRNA Synthesis in Eukaryote is a complex Process

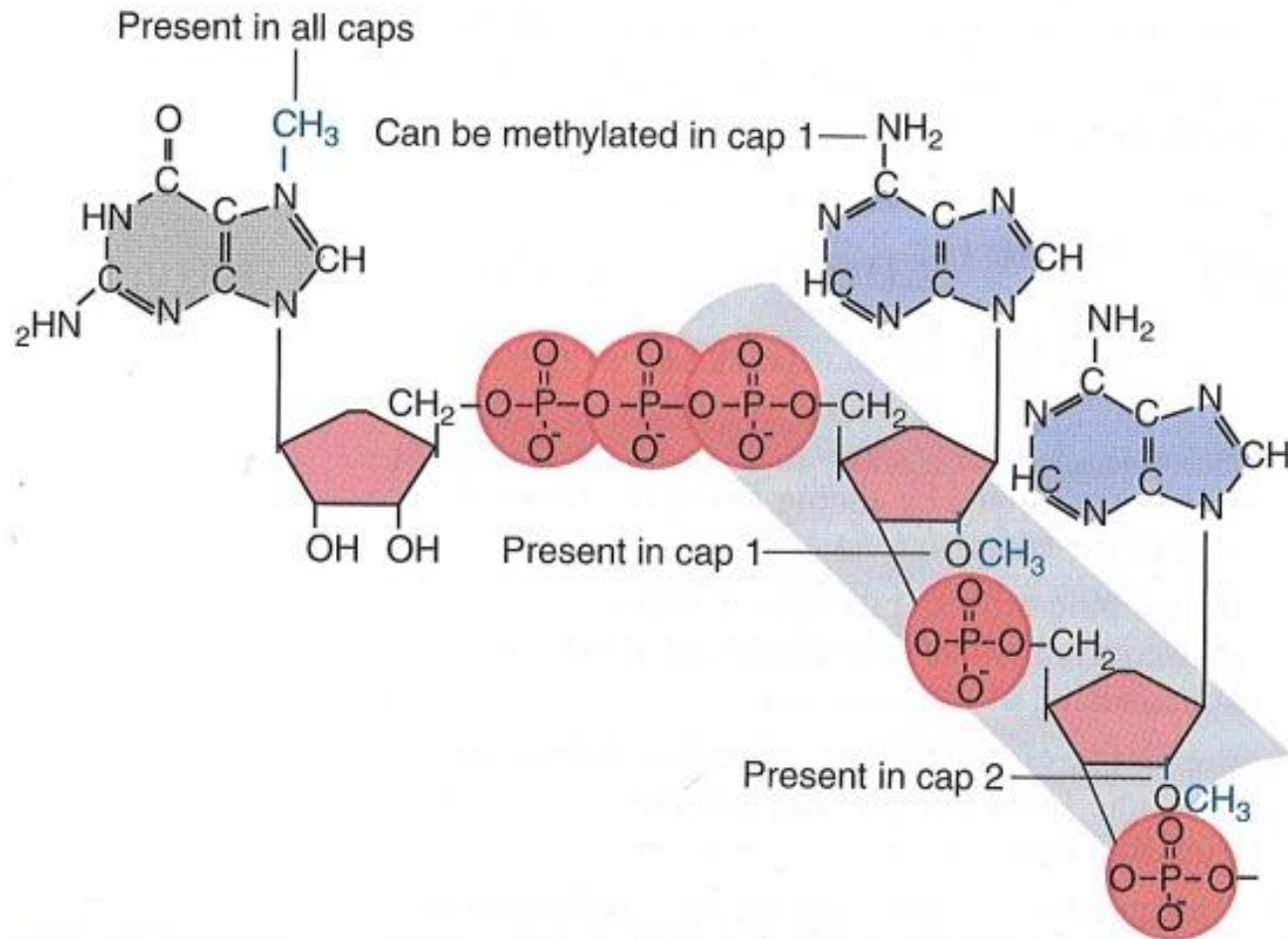
- Transcription Initiation
- Transcription Elongation
- 5' Transcript Processing (CAP)
- Transcription Termination
- 3' Transcript Processing
- Intron Splicing
- Transport into Cytoplasm

## Eukaryotic mRNA is modified and exported

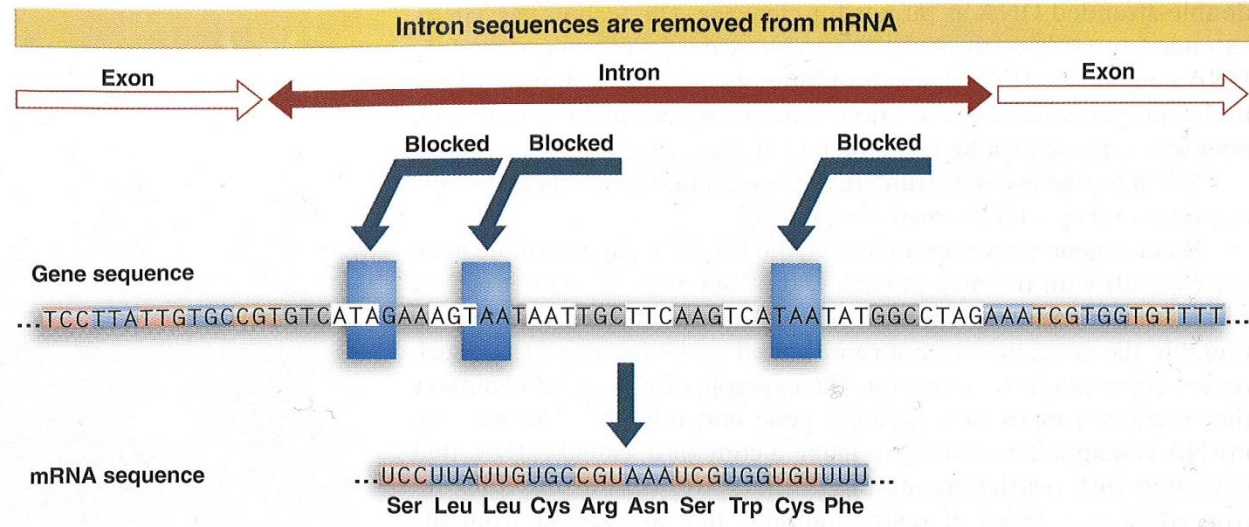


**Figure 7.17** Overview: expression of mRNA in animal cells requires transcription, modification, processing, nucleocytoplasmic transport, and translation.

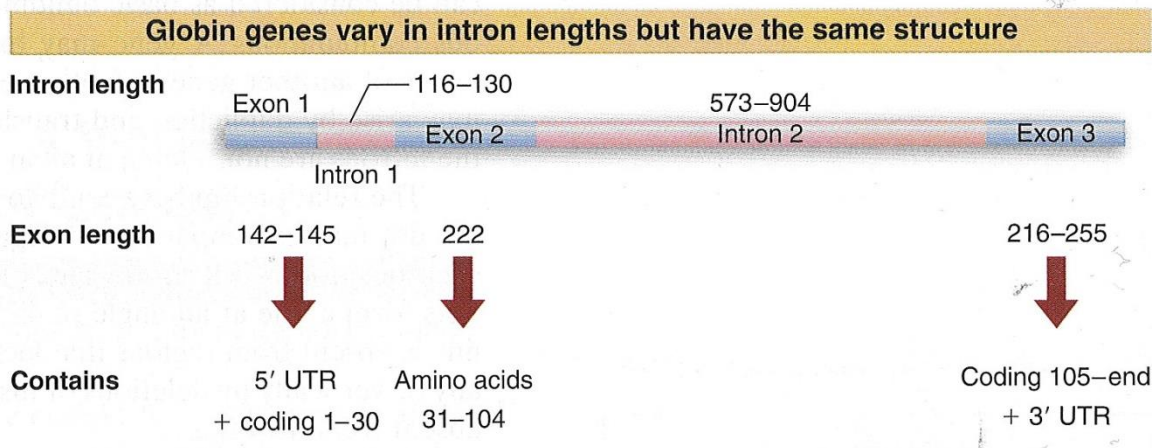
## CAP structure at 5' end of mRNA



**FIGURE 21.2** The cap blocks the 5' end of mRNA and can be methylated at several positions.

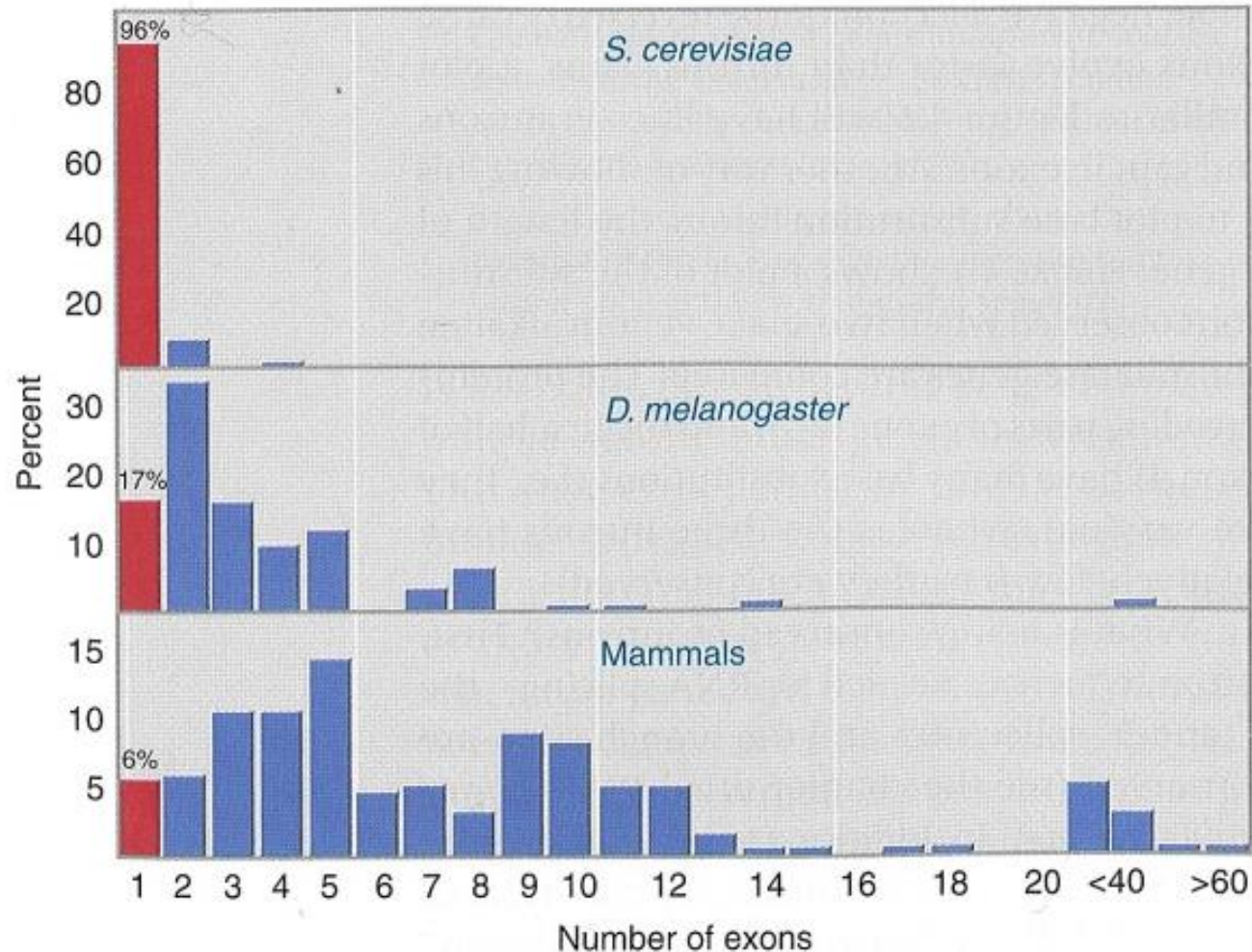


**Figure 3.6** An intron is a sequence present in the gene but absent from the mRNA (here shown in terms of the cDNA sequence). All three possible reading frames are blocked by termination codons in the intron.



**Figure 3.8** All functional globin genes have an interrupted structure with three exons. The lengths indicated in the figure are those of the mammalian  $\beta$ -globin genes.



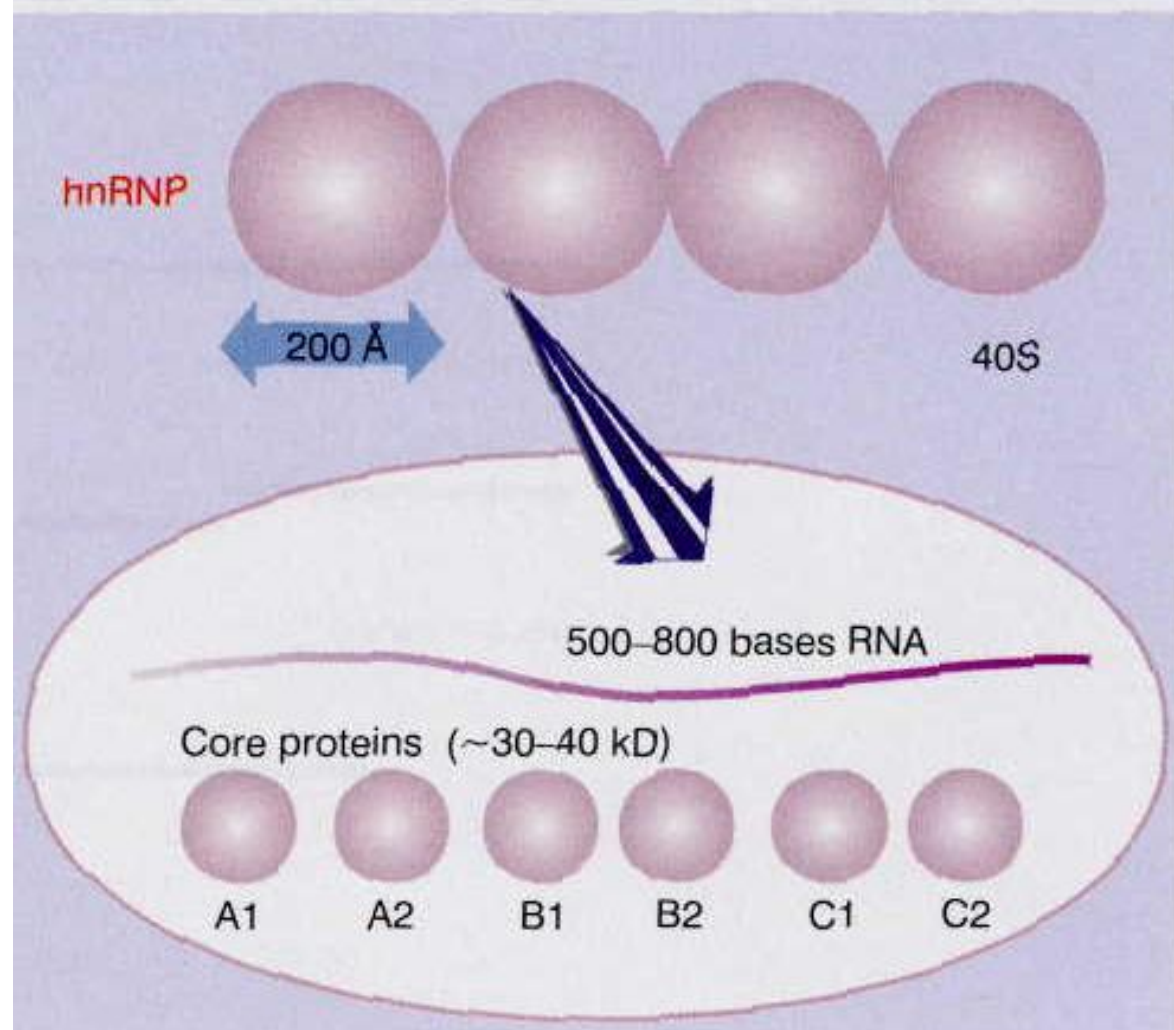


**FIGURE 4.8** Most genes are uninterrupted in yeast, but most genes are interrupted in flies and mammals. (Uninterrupted genes have only one exon and are totaled in the leftmost column in red.)

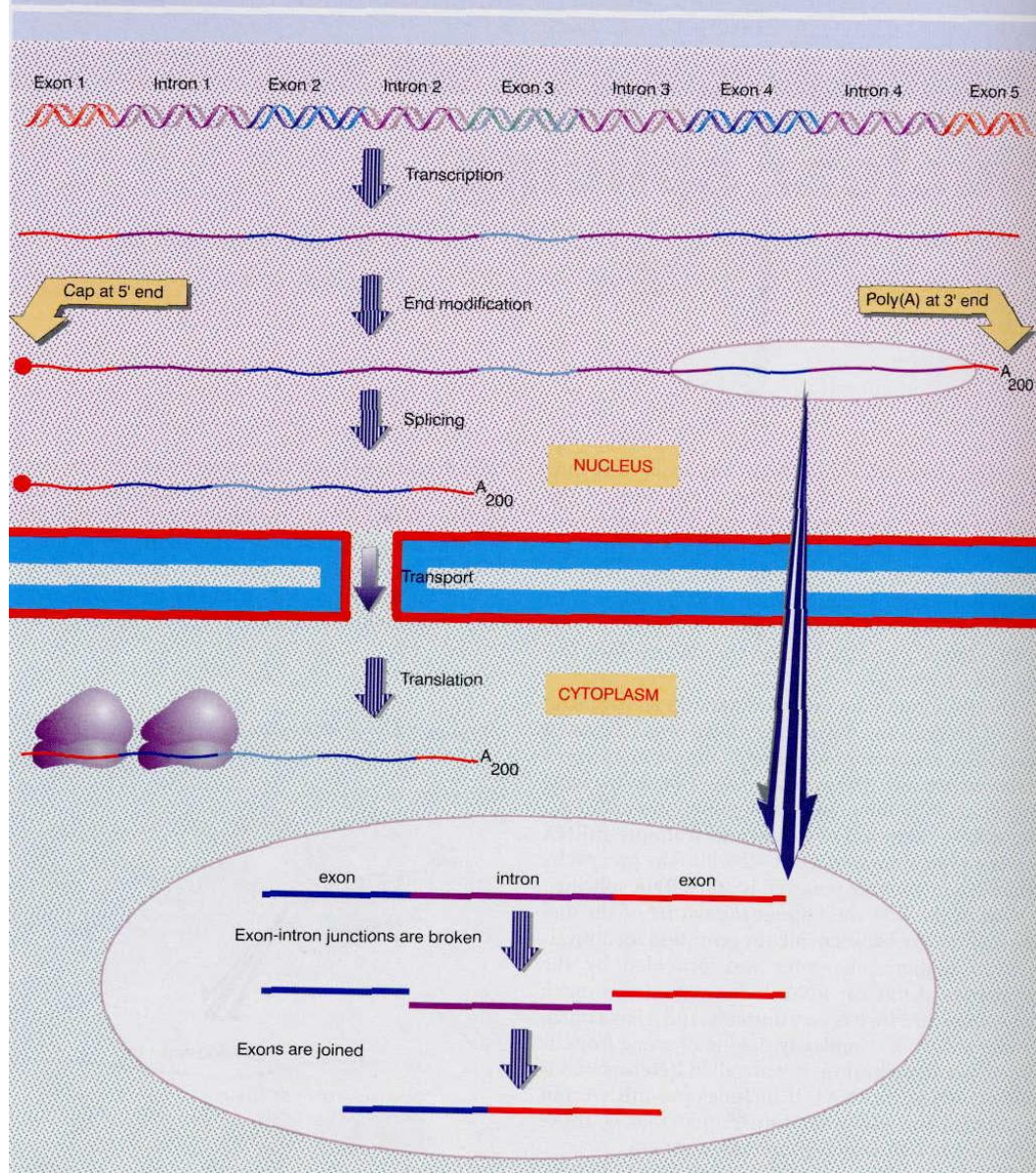
hnRNA: heterogeneous  
nuclear RNA

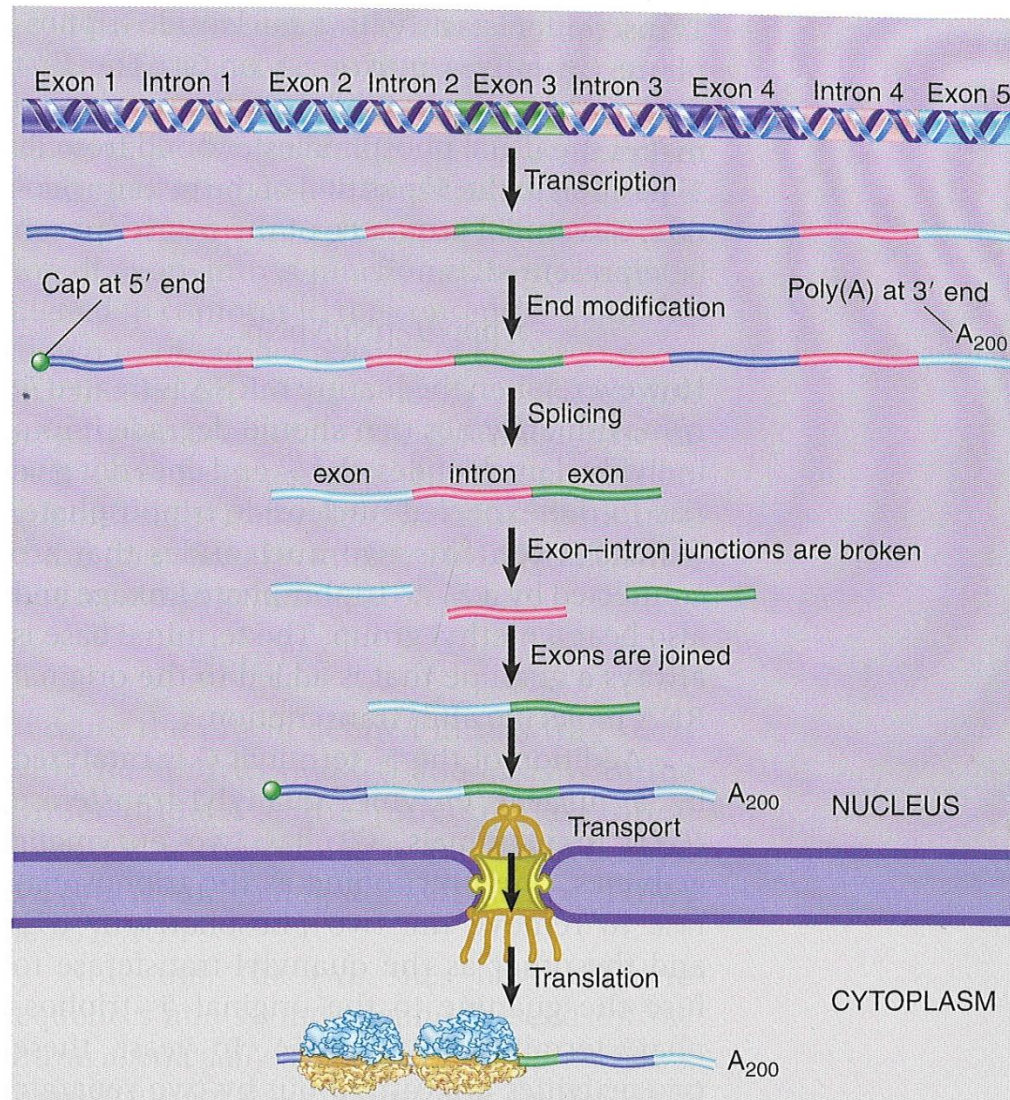
hnRNP: heterogeneous  
nuclear Ribonucleoprotein

**Figure 22.1** hnRNA exists as a ribonucleo-protein particle organized as a series of beads.



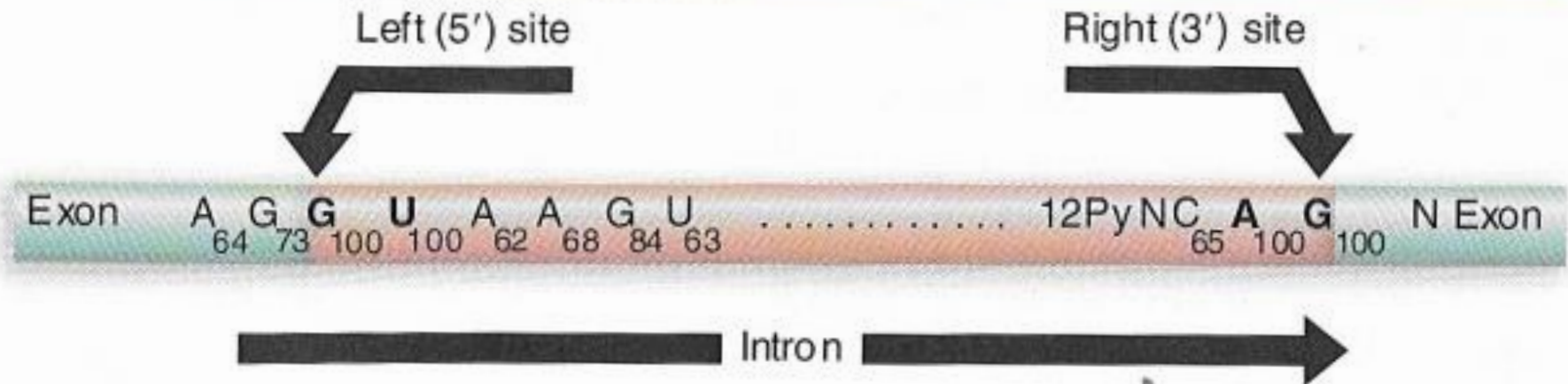
**Figure 22.2** Overview: RNA is modified in the nucleus by additions to the 5' and 3' ends and by splicing to remove the introns. The splicing event requires breakage of the exon-intron junctions and joining of the ends of the exons; the expanded illustration shows the principle schematically, but not the actual order of events. Mature mRNA is transported through nuclear pores to the cytoplasm, where it is translated.



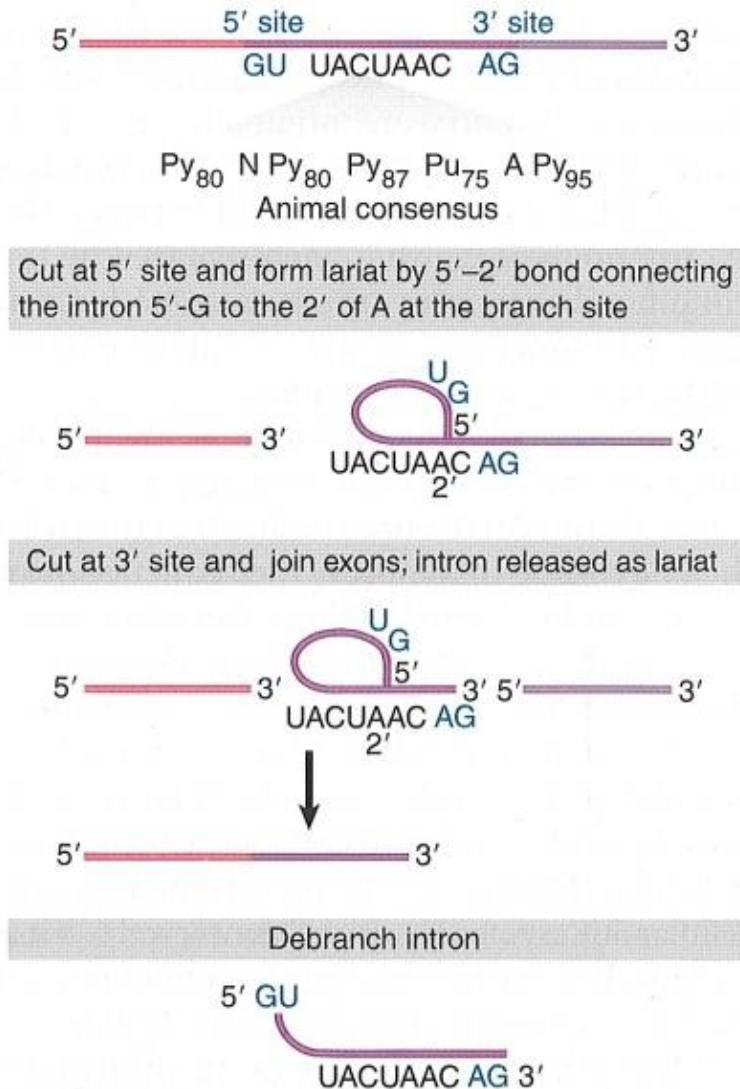


**FIGURE 21.1** RNA is modified in the nucleus by additions to the 5' and 3' ends and by splicing to remove the introns. The splicing event requires breakage of the exon-intron junctions and joining of the ends of the exons. Mature mRNA is transported through nuclear pores to the cytoplasm, where it is translated.

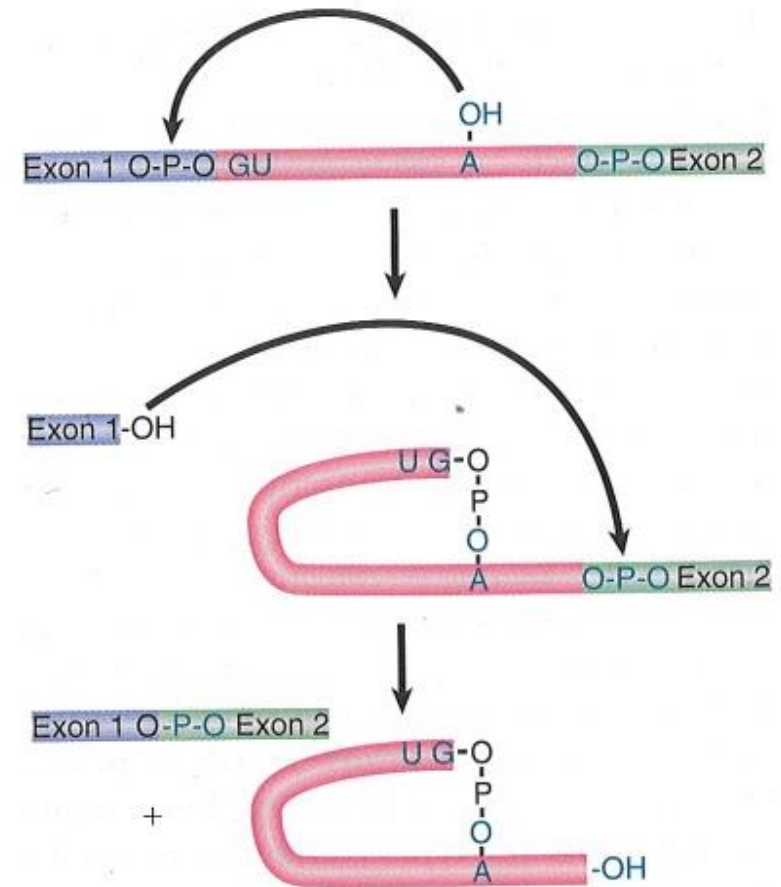
## Intron–exon boundaries have short consensus sequences in the intron



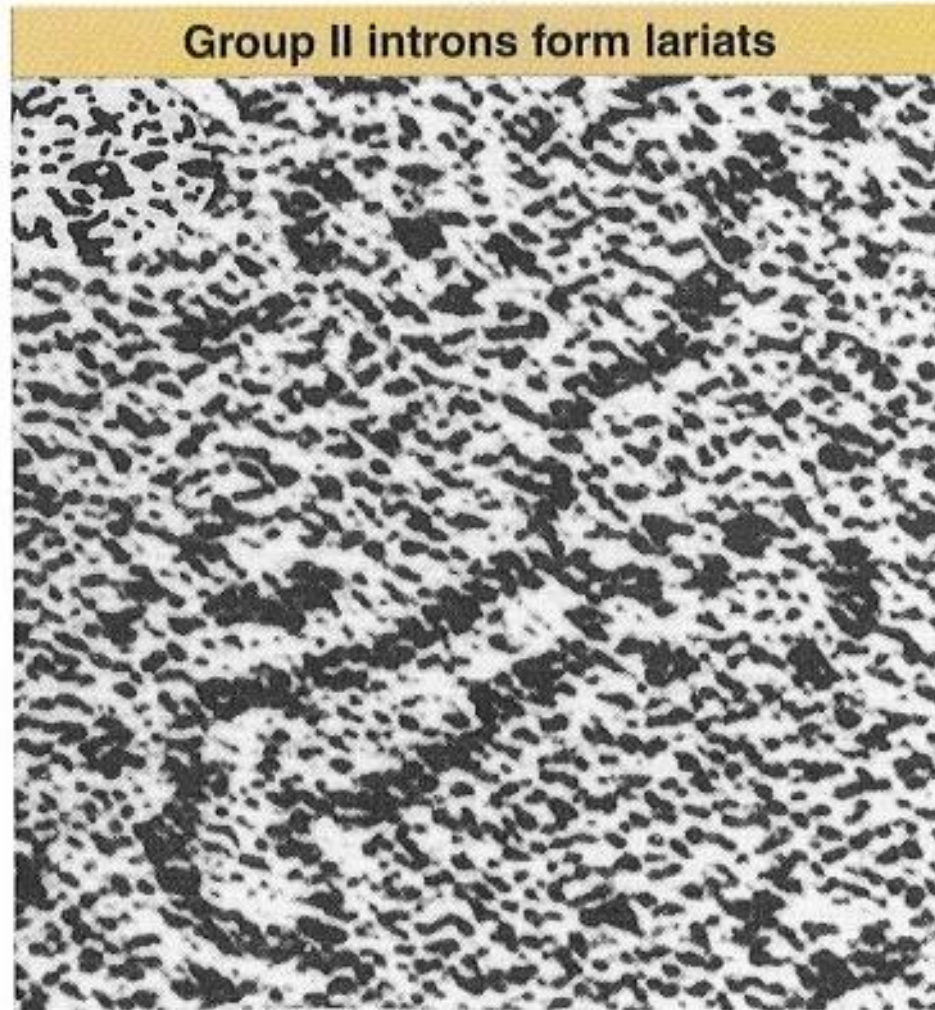
**Figure 26.2** The ends of nuclear introns are defined by the GU–AG rule.



**FIGURE 21.5** Splicing occurs in two stages. First the 5' exon is cleaved off, and then it is joined to the 3' exon.

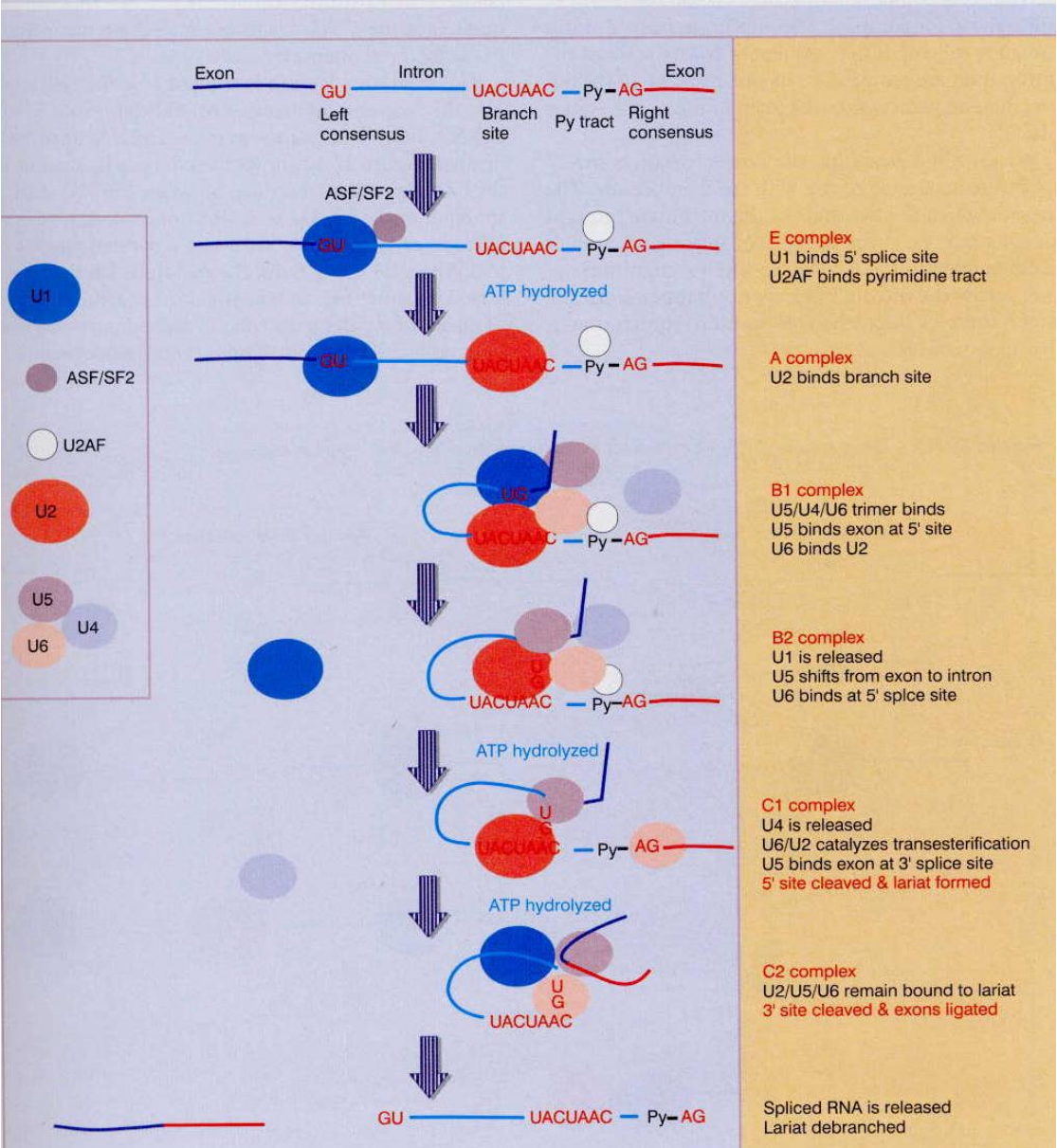


**FIGURE 21.6** Nuclear splicing occurs by two transesterification reactions, in which an -OH group attacks a phosphodiester bond.

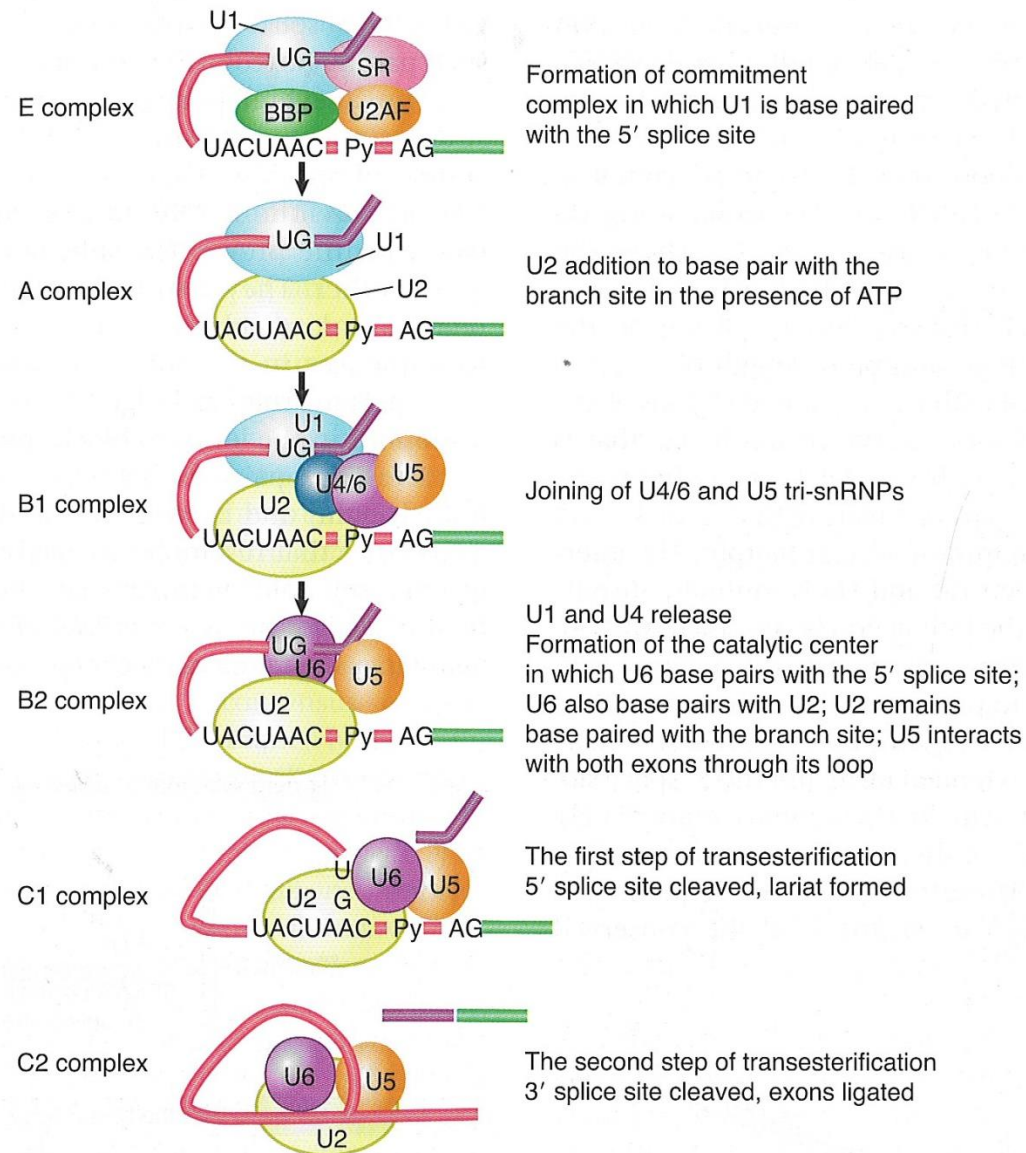


**Figure 26.18** Splicing releases a mitochondrial group II intron in the form of a stable lariat. Photograph kindly provided by Leslie Grivell and Annika Arnberg.

**Figure 22.10** The splicing reaction proceeds through discrete stages in which spliceosome formation involves the interaction of components that recognize the consensus sequences.







**FIGURE 21.12** The splicing reaction proceeds through discrete stages in which spliceosome formation involves the interaction of components that recognize the consensus sequences.

A spliceosome forms through several discrete complexes

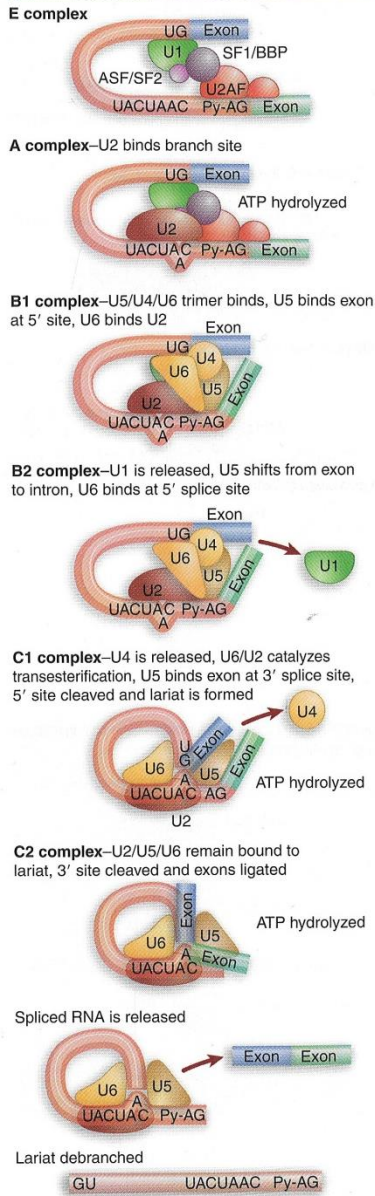
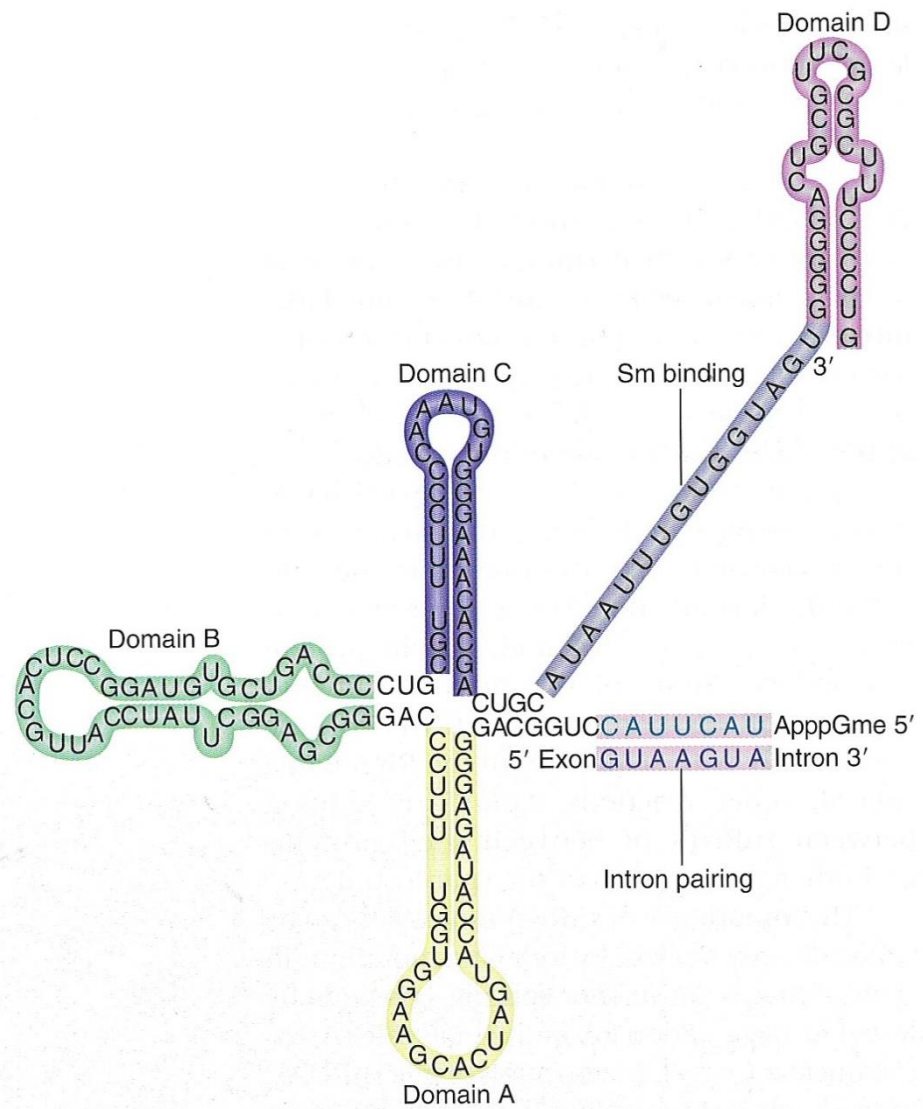


Figure 26.12 The splicing reaction proceeds through discrete stages in which spliceosome formation involves the interaction of components that recognize the consensus sequences.



snRNA: small nuclear RNA

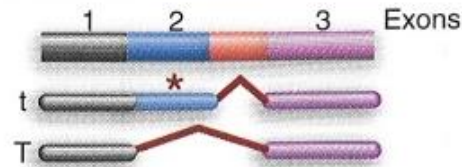
snRNP: small nuclear Ribonucleoparticle „snurps“

**FIGURE 21.8** U1 snRNA has a base-paired structure that creates several domains. The 5' end remains single stranded and can base pair with the 5' splice site.

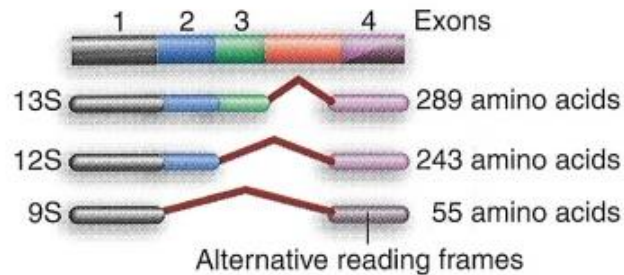
Taken from: J.E. Krebs, E.S. Goldstein, S.T. Kilpatrick; „Lewin’s Genes XI“; Jones&Bartlett Learning

### Alternative splicing generates multiple RNAs

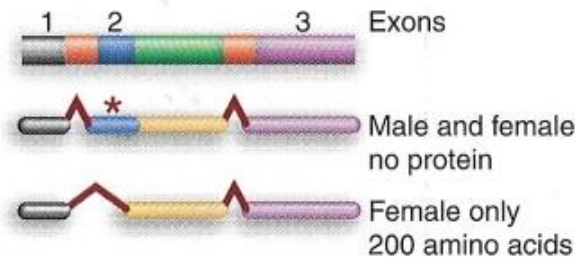
SV40 T/t antigens splice two 5' sites to a common 3' site



Adenovirus E1A splices variable 5' sites to a common 3' site



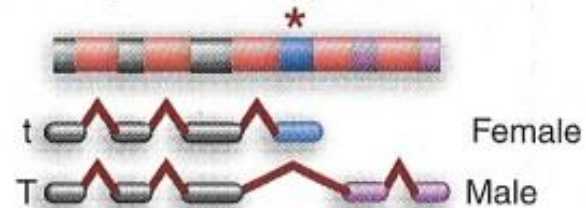
*D. melanogaster tra* splices 5' site to alternative 3' sites



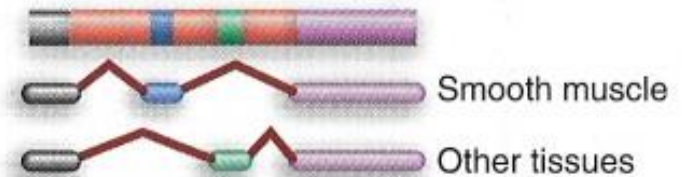
**Figure 26.20** Alternative forms of splicing may generate a variety of protein products from an individual gene. Changing the splice sites may introduce termination codons (shown by asterisks) or change reading frames.

### Alternative splicing may substitute exons

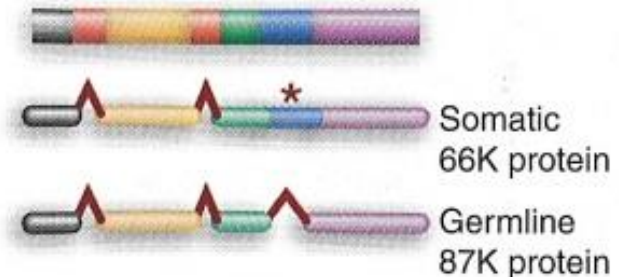
*D. melanogaster dsx* skips an exon



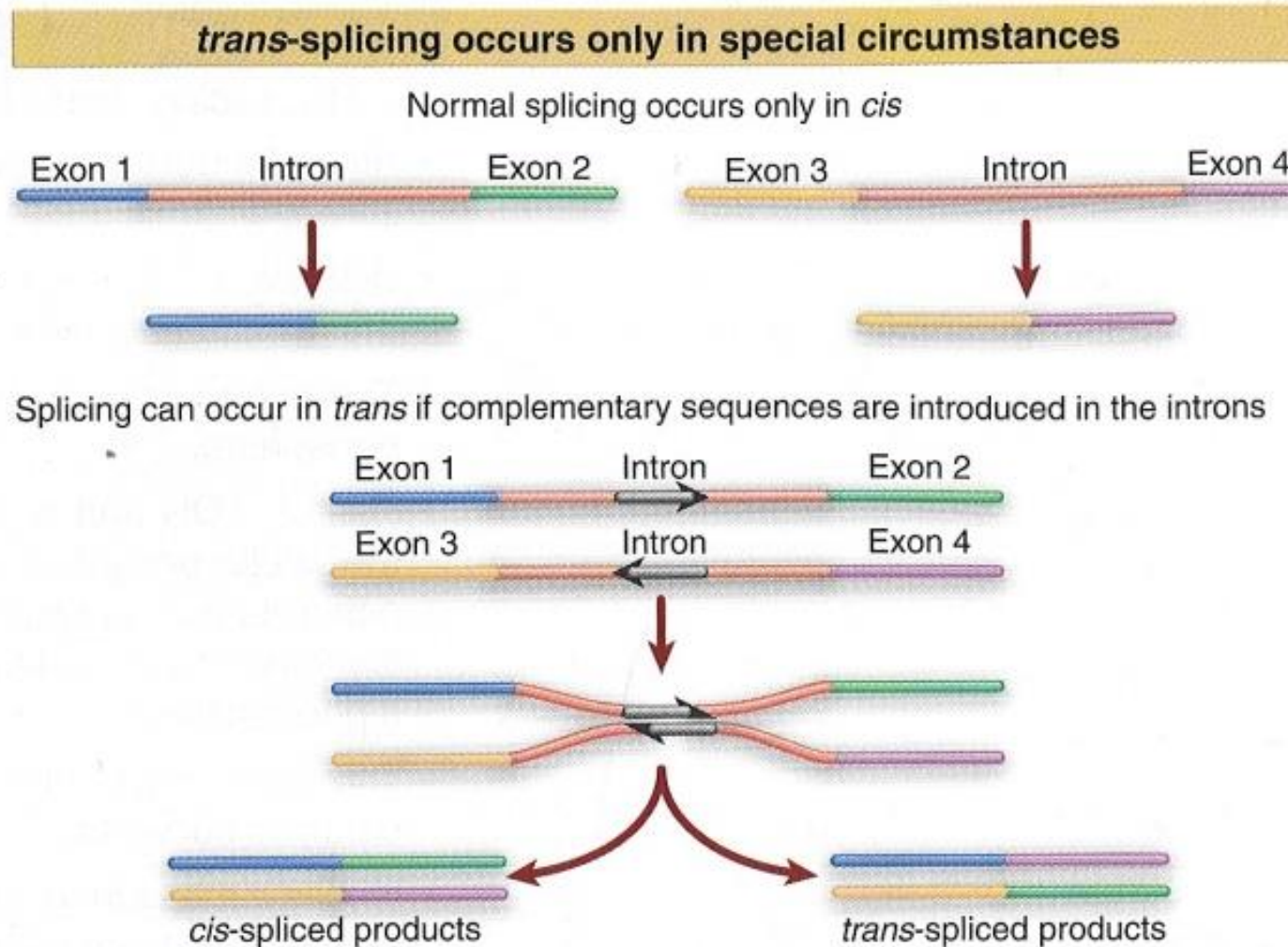
$\alpha$ -tropomyosin splices alternative exons



*P* elements splice out an extra intron



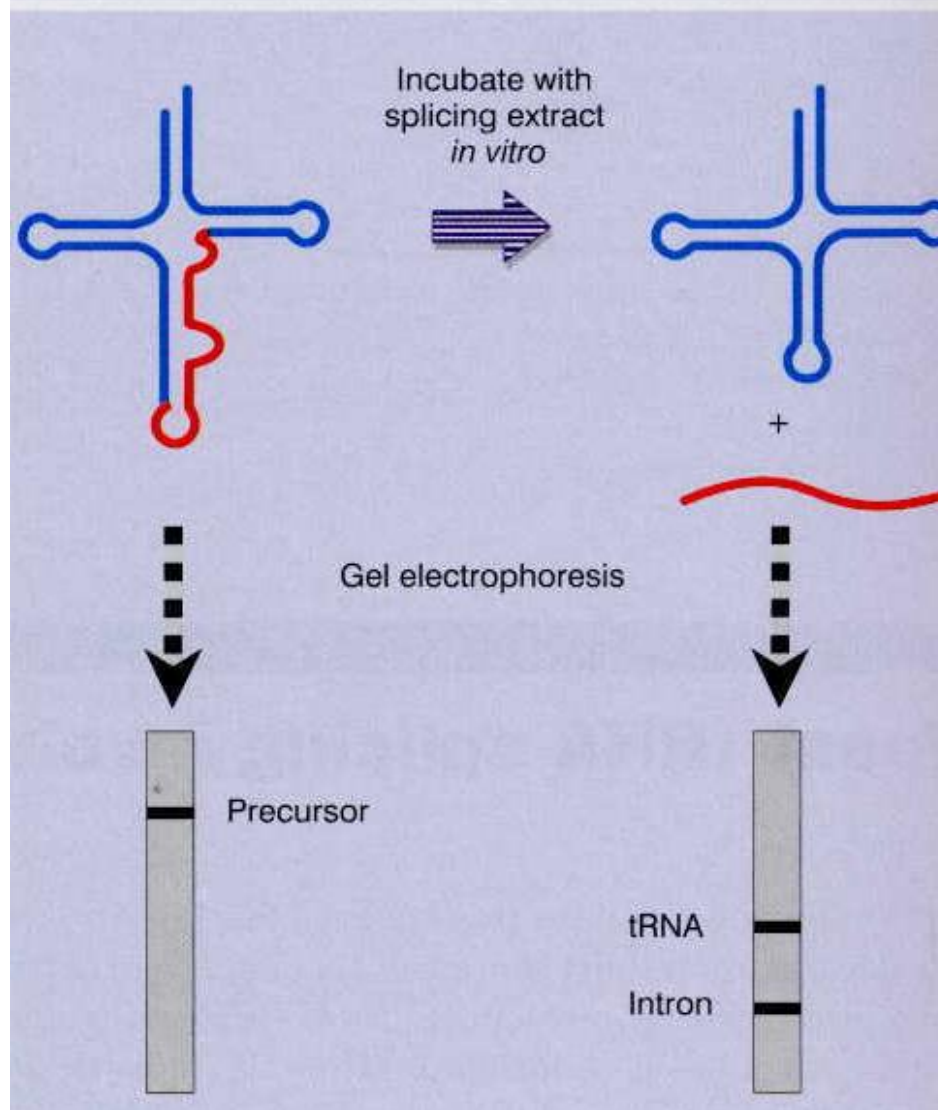
**Figure 26.21** Alternative splicing events may cause exons to be added or substituted.



**Figure 26.22** Splicing usually occurs only in *cis* between exons carried on the same physical RNA molecule, but *trans* splicing can occur when special constructs are made that support base pairing between introns.

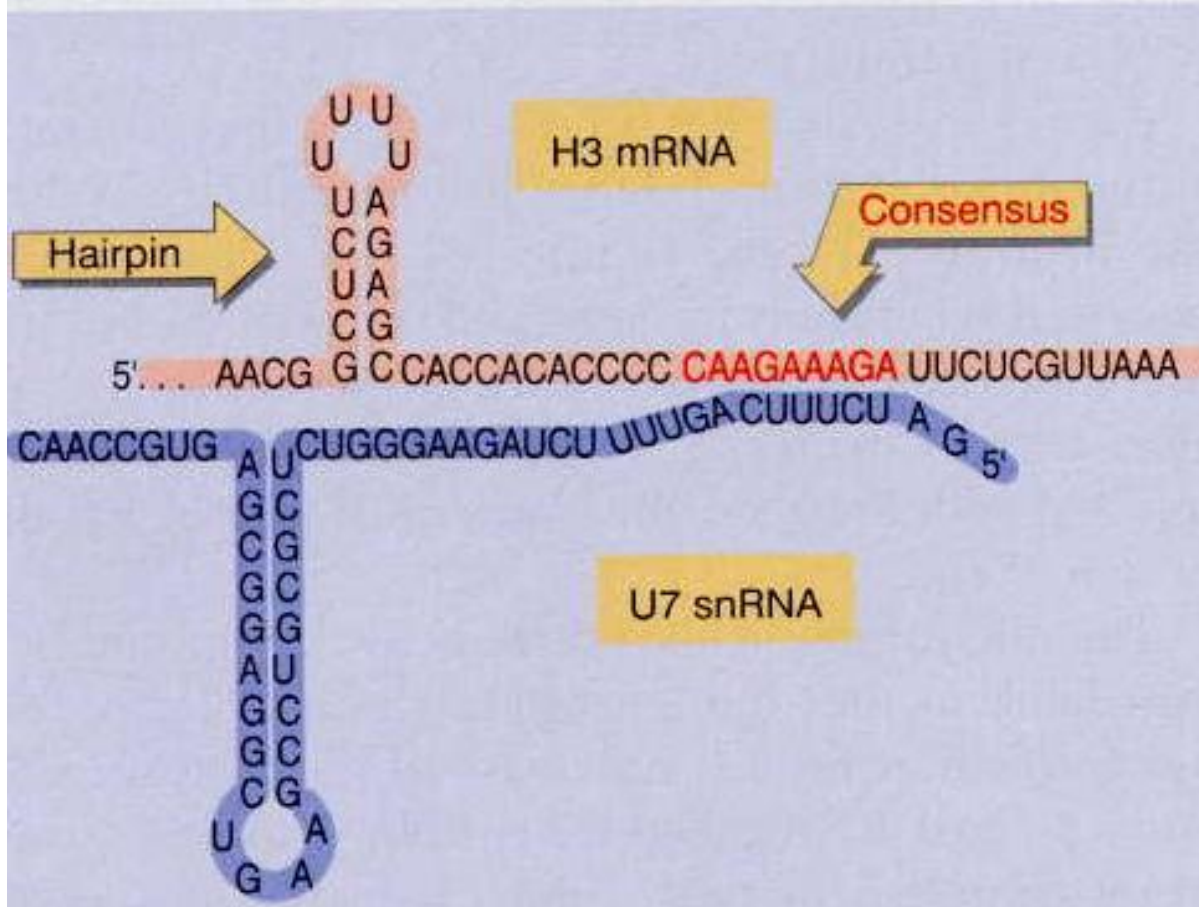
17.11.15

**Figure 22.24** Splicing of yeast tRNA *in vitro* can be followed by assaying the RNA precursor and products by gel electrophoresis.



## 3'-end processing

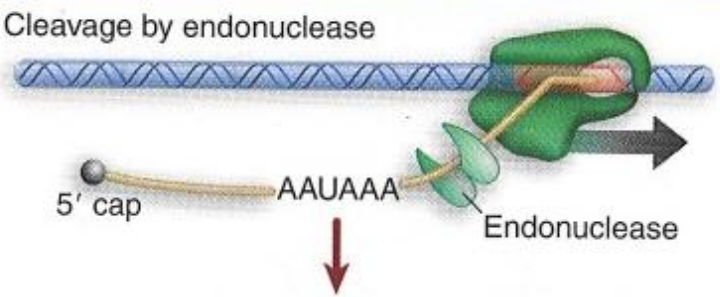
**Figure 22.30** Generation of the 3' end of histone H3 mRNA depends on a conserved hairpin and a sequence that base pairs with U7 snRNA.



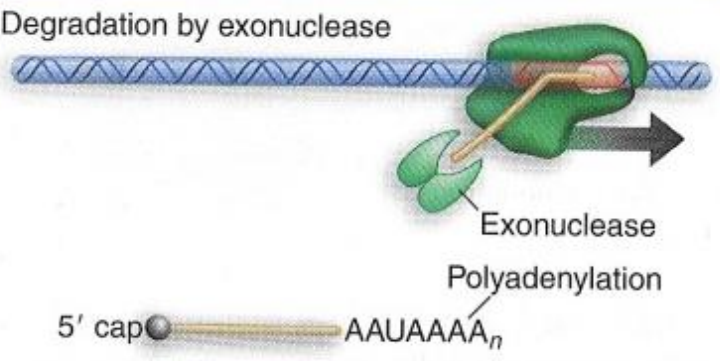


# 3'-end processing

The 3' end of mRNA is generated by cleavage

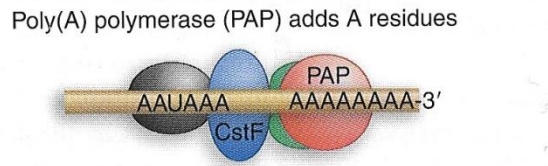
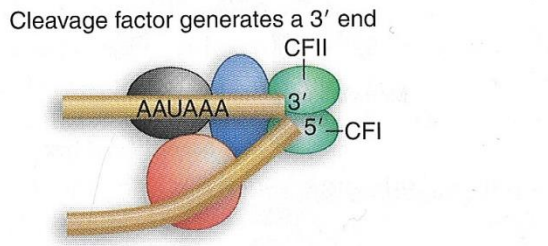
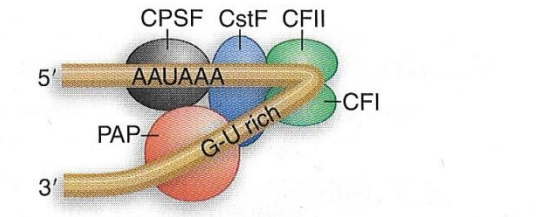


mRNA is stabilized by polyadenylation

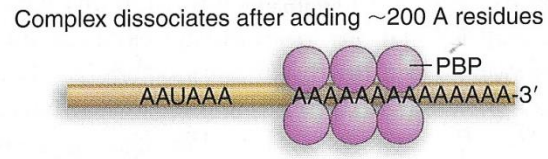
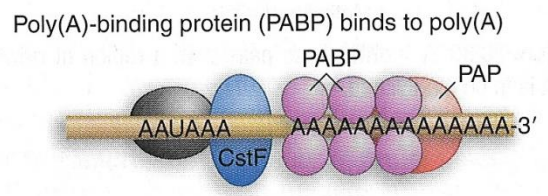


**Figure 26.28** The sequence AAUAAA is necessary for cleavage to generate a 3' end for polyadenylation.

There is a single 3' end-processing complex



PAP: PolyA-Polymerase



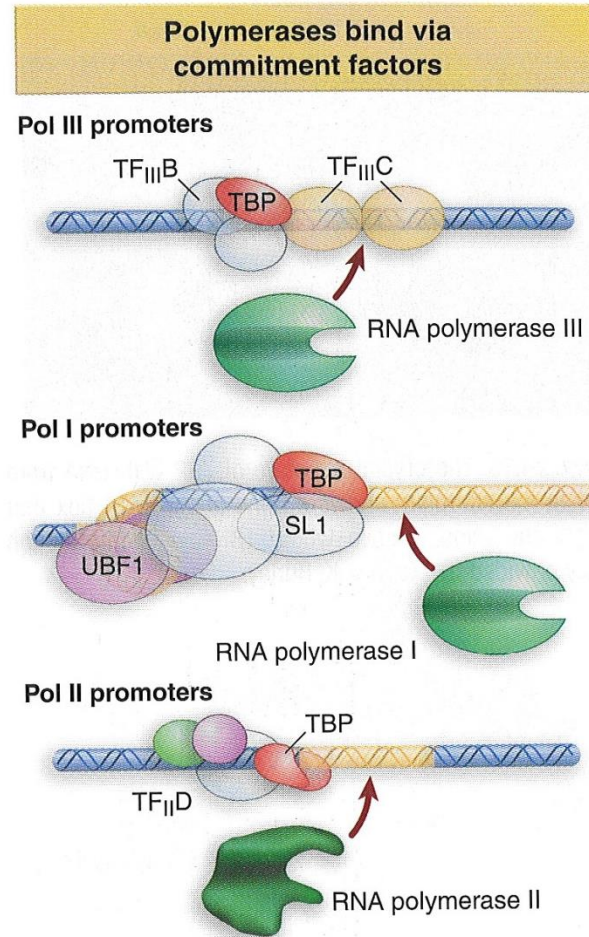
**Figure 26.29** The 3' processing complex consists of several activities. CPSF and CstF each consist of several subunits; the other components are monomeric. The total mass is >900 kD.

## Transcription in Eukayotes

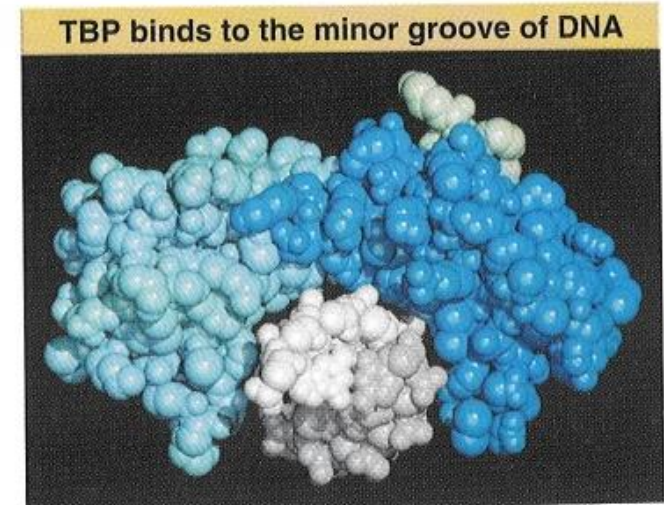
Pol III:  
Transfer RNA  
5S rRNA  
Small nuclear RNA U6  
Repeated DNA sequ.  
(e.g. Alu)

Pol I:  
Ribosomal RNAs

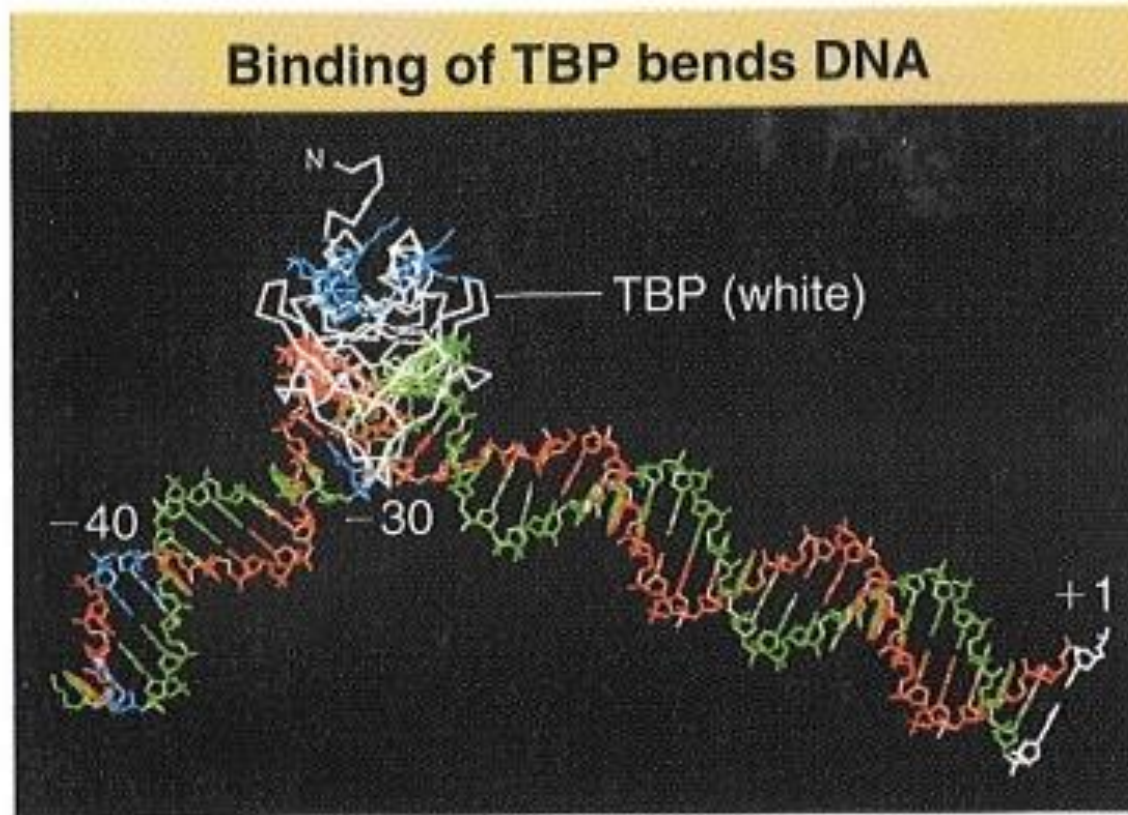
Pol II:  
All coding genes  
Small nuclear RNAs



**Figure 24.8** RNA polymerases are positioned at all promoters by a factor that contains TBP.

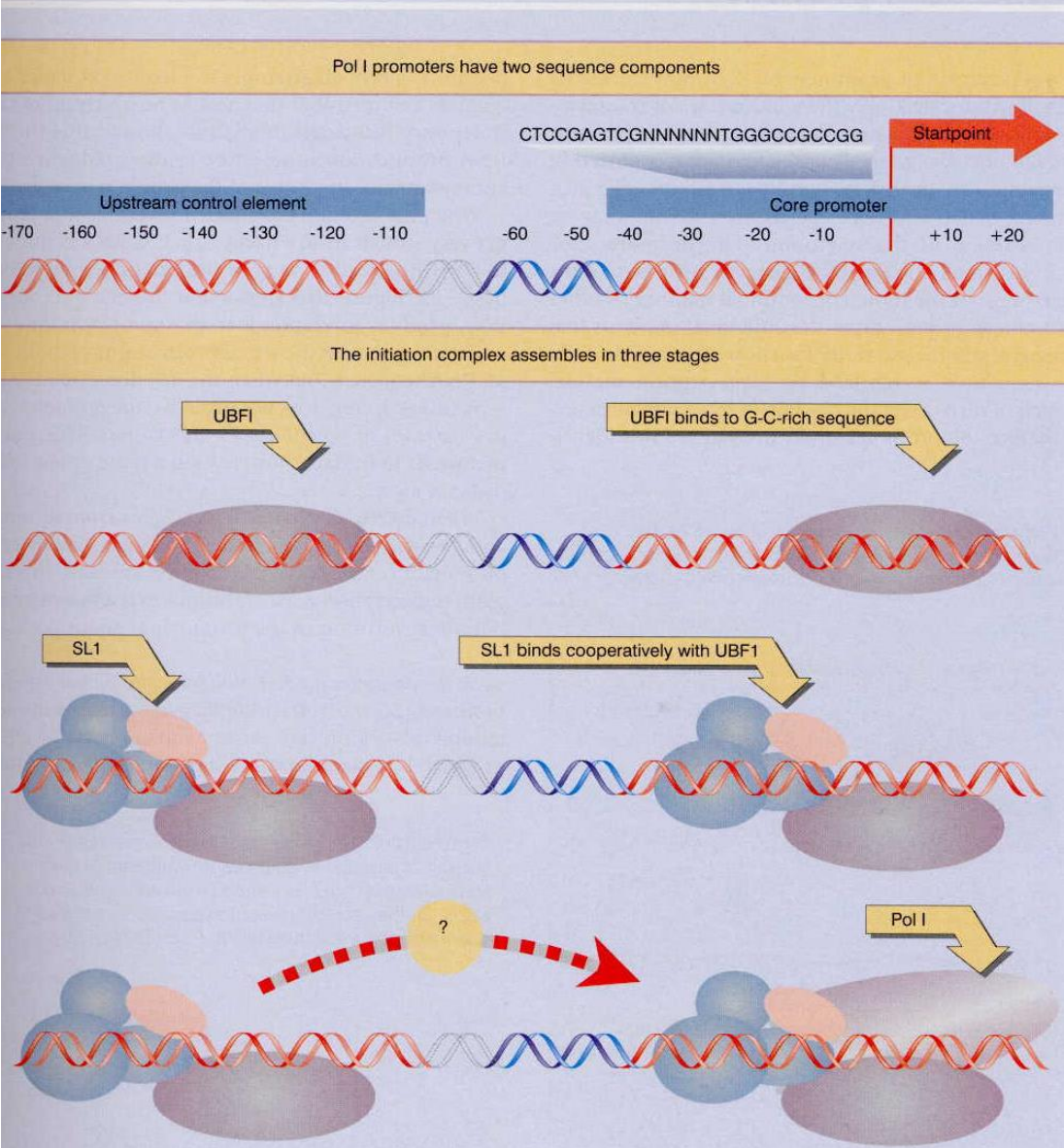


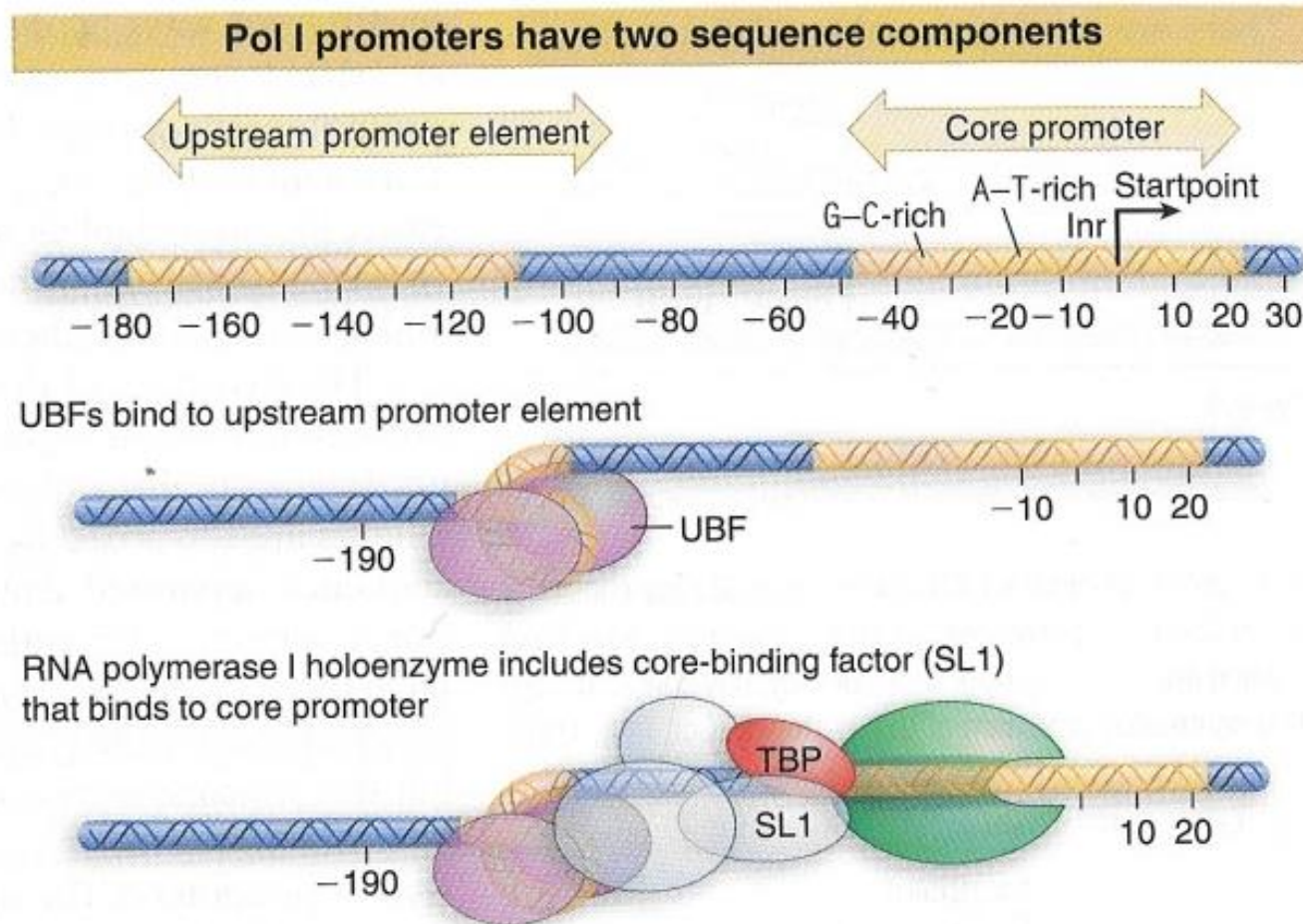
**Figure 24.9** A view in cross section shows that TBP surrounds DNA from the side of the minor groove. TBP consists of two related (40% identical) conserved domains, which are shown in light and dark blue. The N-terminal region varies extensively and is shown in green. The two strands of the DNA double helix are in light and dark grey. Photograph kindly provided by Stephen K. Burley.



**Figure 24.10** The crystal structure of TBP with DNA from -40 to the startpoint shows a bend at the TATA box that widens the minor groove where TBP binds. Photograph kindly provided by Stephen K. Burley.

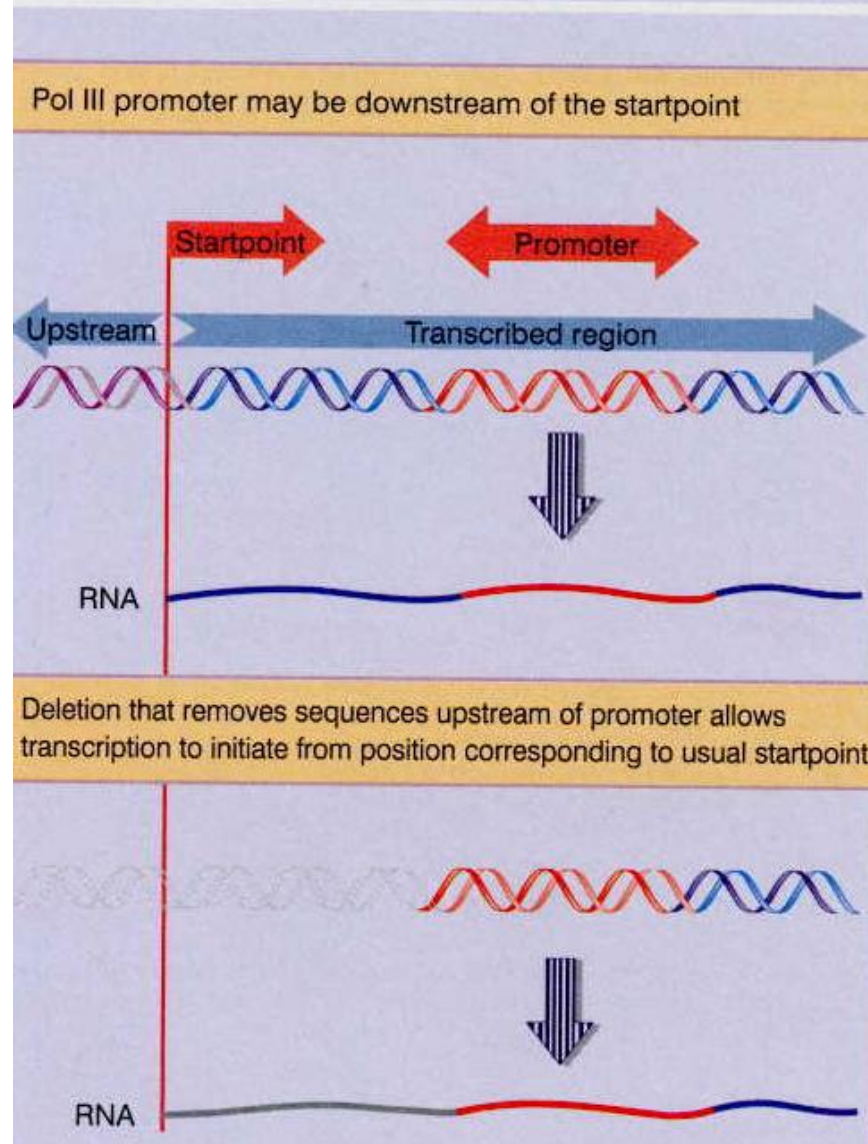
**Figure 20.4** Transcription units for RNA polymerase I have a core promoter separated by ~70 bp from the upstream control element. UBF1 binds to both regions, after which SL1 can bind. RNA polymerase I then binds to the core promoter. The nature of the interaction between the factors bound at the upstream control element and those at the core promoter is not known.



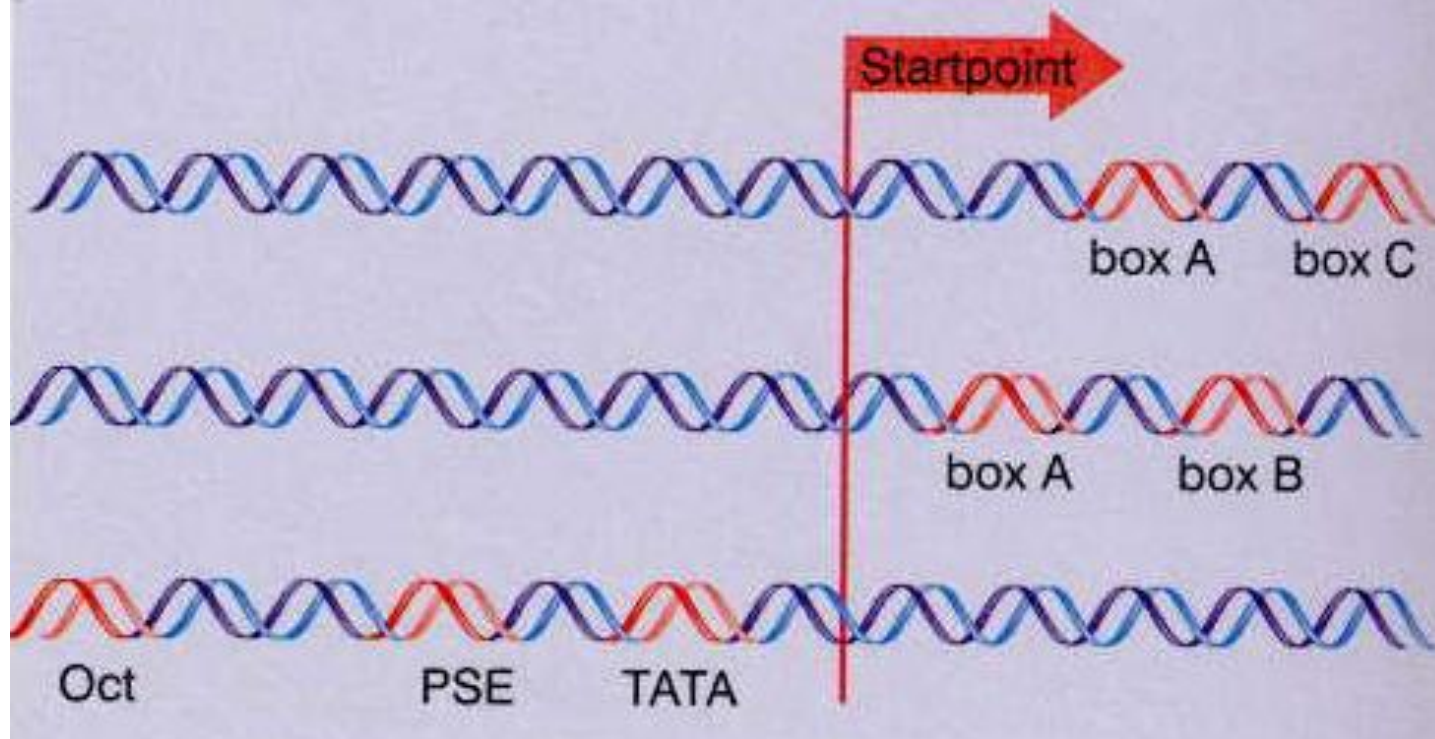


**Figure 24.3** Transcription units for RNA polymerase I have a core promoter separated by ~70 bp from the upstream promoter element. UBF binding to the UPE increases the ability of core-binding factor (SL1) to bind to the core promoter.

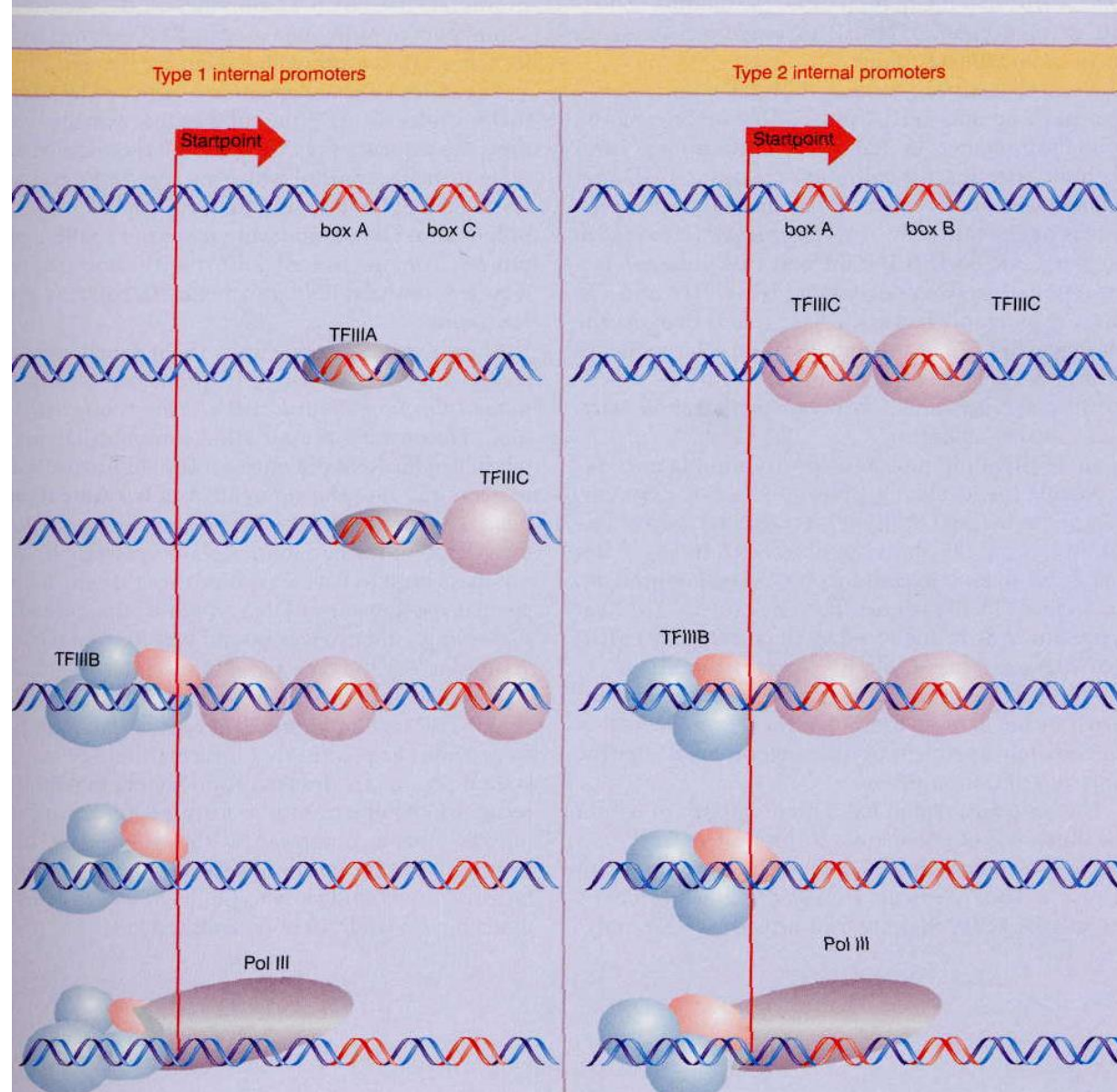
**Figure 20.5** Deletion analysis shows that the promoter for 5S RNA genes is internal; initiation occurs a fixed distance (~55 bp) upstream of the promoter.



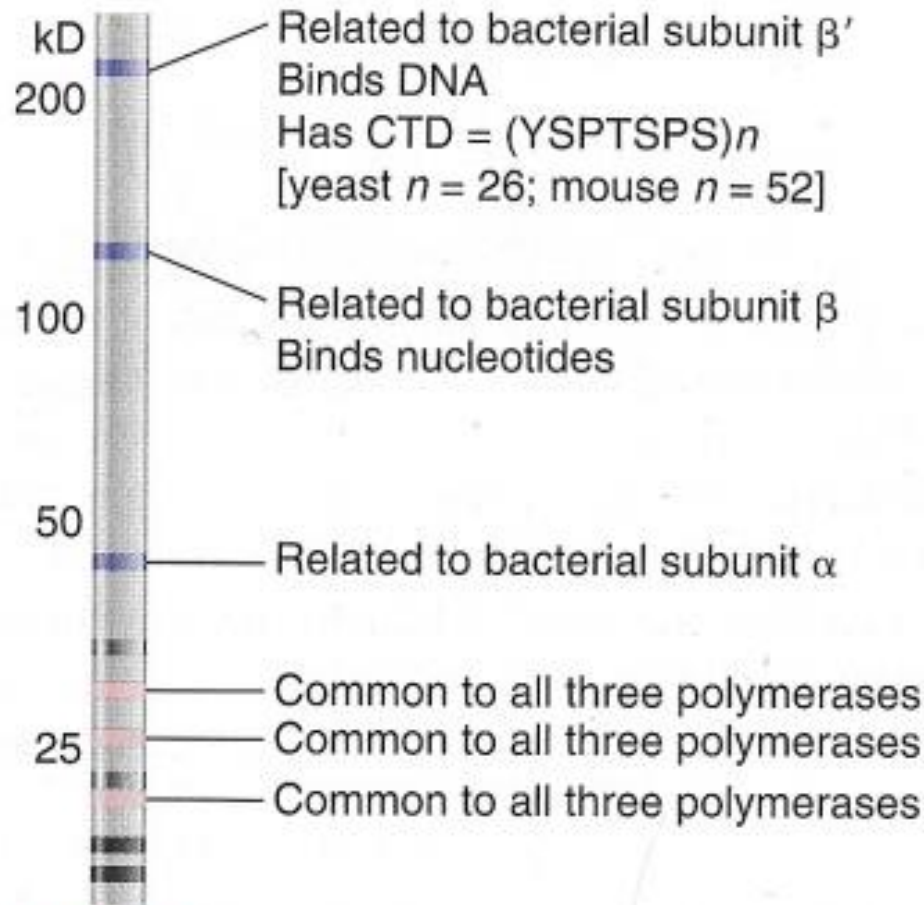
**Figure 20.6** Promoters for RNA polymerase III may consist of bipartite sequences downstream of the startpoint, with boxA separated from either boxC or boxB. Or they may consist of separated sequences upstream of the startpoint (Oct, PSE, TATA).



**Figure 20.7** Initiation via the internal pol III promoters involves the assembly factors TFIIIA and TFIIIC, the initiation factor TFIIIB, and RNA polymerase III.



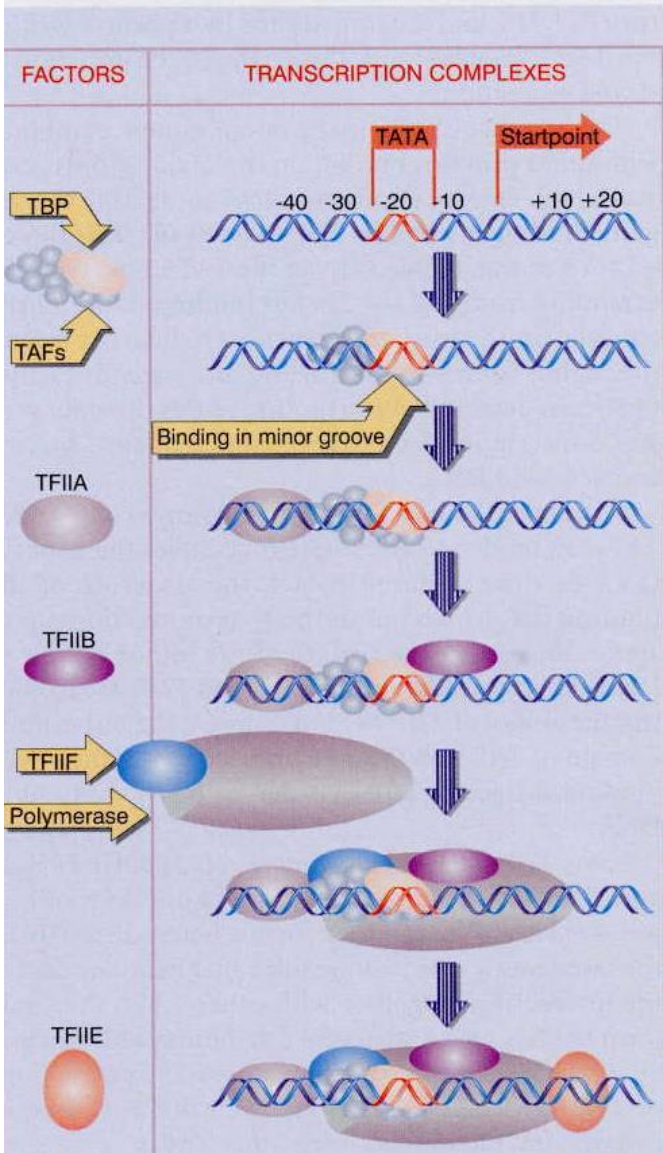




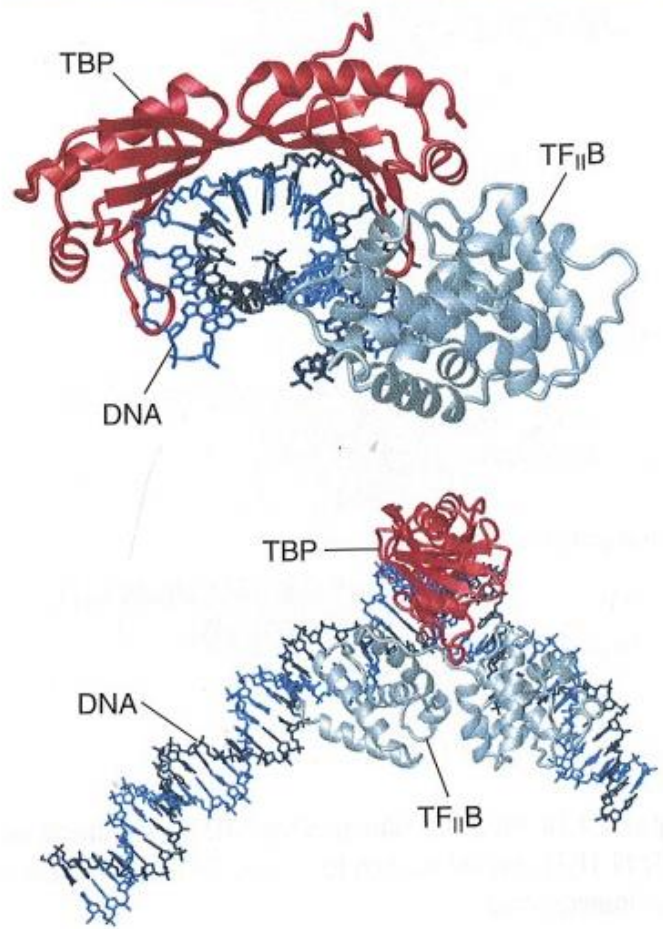
Pol II Promoters

**FIGURE 20.2** Some subunits are common to all classes of eukaryotic RNA polymerases and some are related to bacterial RNA polymerase. This drawing is a simulation of purified yeast RNA polymerase II run on an SDS gel to separate the subunits by size.

**Figure 20.11** An initiation complex assembles at promoters for RNA polymerase II by an ordered sequence of association with transcription factors.



**TF<sub>II</sub>B binds to bent DNA downstream from TBP**



**Figure 24.12** Two views of the ternary complex of TF<sub>II</sub>B-TBP-DNA show that TF<sub>II</sub>B binds along the bent face of DNA. Photograph kindly provided by Stephen K. Burley.

Taken from: B. Lewin, *Essential Genes*, Pearson Ed. International

The  $\beta$ -globin promoter has three short sequence elements

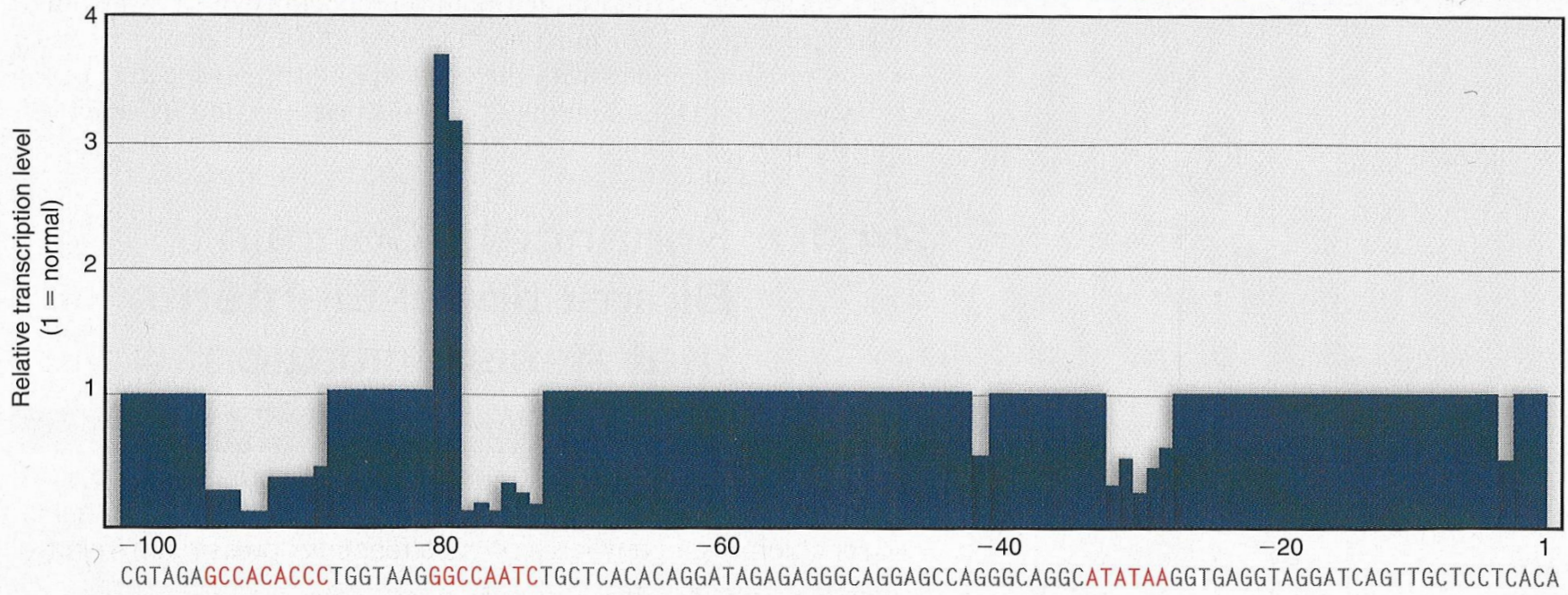
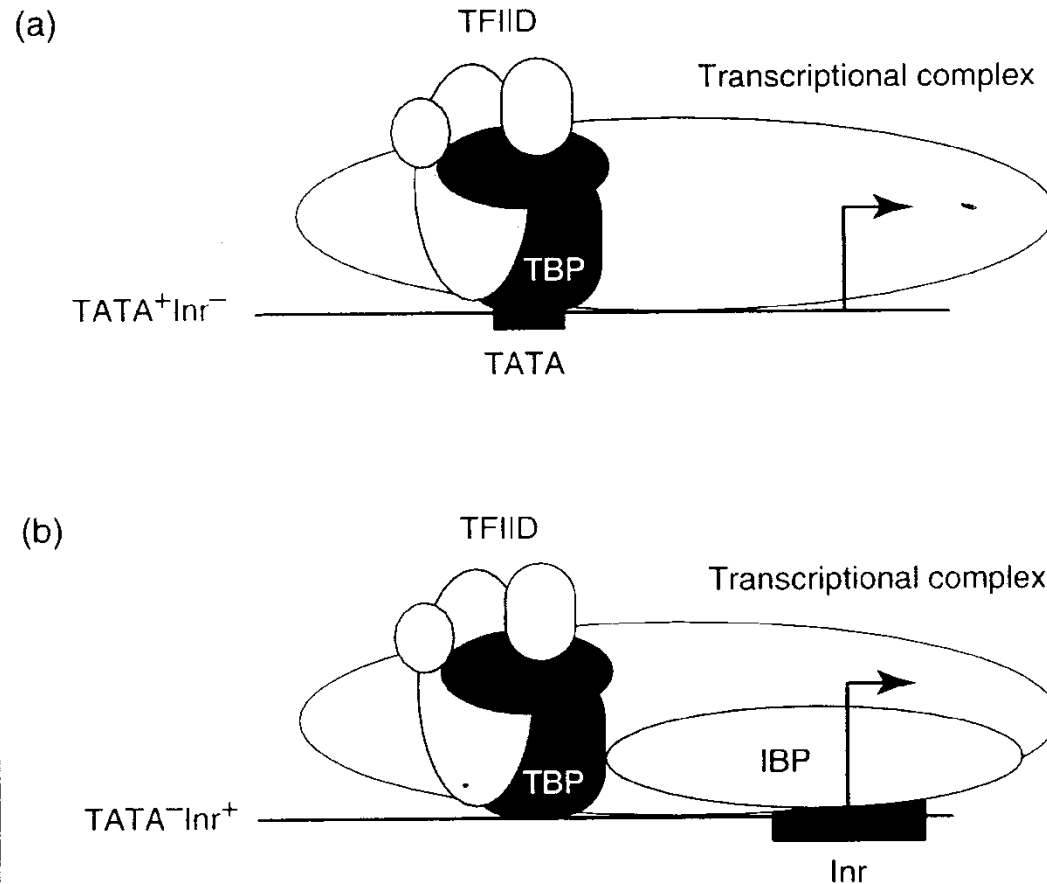
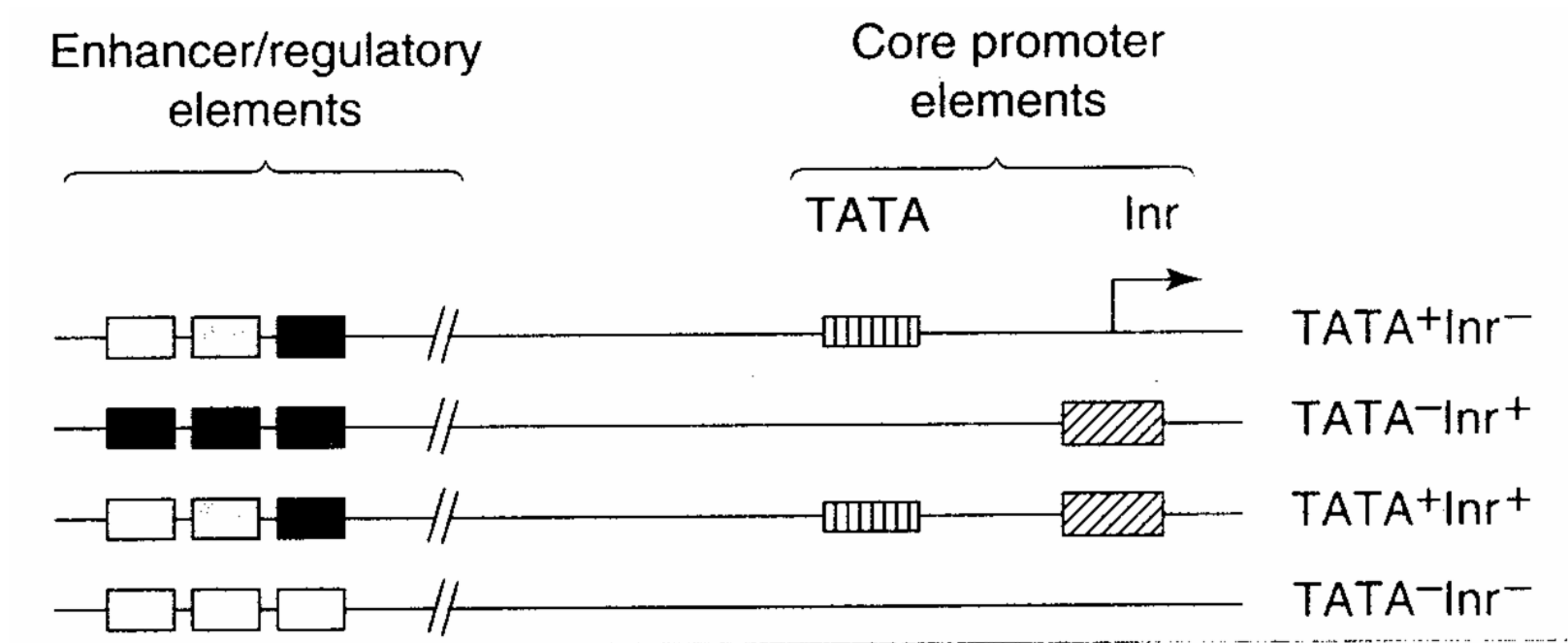


Figure 24.17 Saturation mutagenesis of the upstream region of the  $\beta$ -globin promoter identifies three short regions (centered at -30, -75, and -90) that are needed to initiate transcription. These correspond to the TATA, CAAT, and GC boxes.



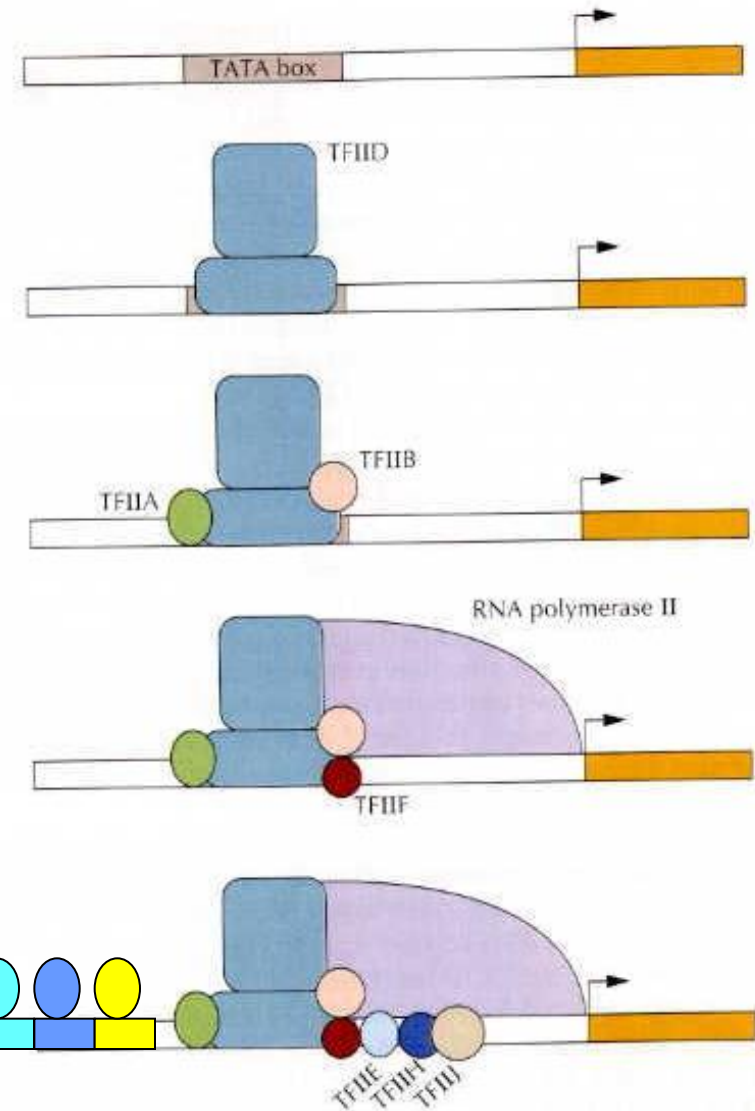
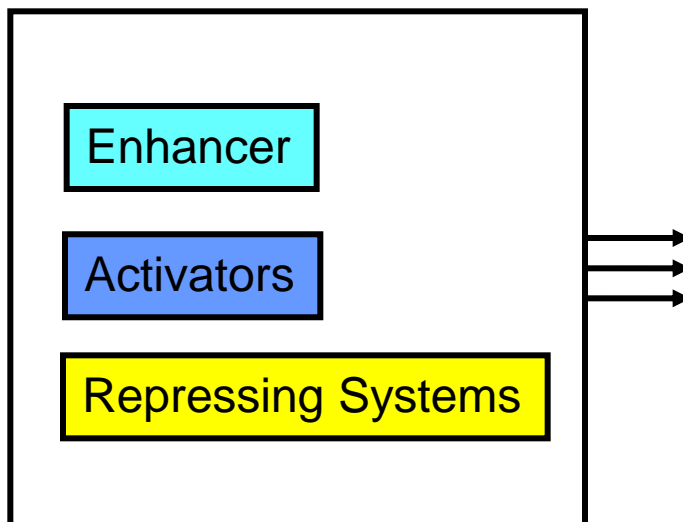
**FIGURE 2.** Formation of the transcription complex at (a) the TATA box or (b) an initiator (Inr) element. In TATA<sup>+</sup> Inr<sup>-</sup> promoters, the initial recognition step is the binding of TFIIID to the TATA box via its DNA-binding subunit, the TATA-binding protein (TBP). Following this event, other general transcription factors (GTFs) might enter into the complex, either in a stepwise fashion or as a holoenzyme complex, giving rise to a transcriptionally competent complex. In TATA<sup>-</sup> Inr<sup>+</sup> promoters, the initial recognition step is the interaction of an Inr-binding protein (IBP). An IBP could be a distinct factor, a GTF, or a subunit of the TFIIID complex (TAF, TBP associated factor). TFIIID is next recruited to the promoter, potentially via protein–protein interactions, and finally other components enter the transcription complex.



**FIGURE 1.** Architecture of different classes of eukaryotic RNA polymerase II (Pol II) promoters. The core promoter region may contain either a TATA box (TATA<sup>+</sup>Inr<sup>-</sup>) or an initiator (Inr) element (TATA<sup>-</sup>Inr<sup>+</sup>). Some promoters might contain both core elements (TATA<sup>+</sup>Inr<sup>+</sup>) and others none (TATA<sup>-</sup>Inr<sup>-</sup>). The transcription start site (+1) is indicated by the arrow. Each promoter may have co-evolved with its associated enhancer region, thereby maintaining specificity of gene expression, especially *in vivo*.

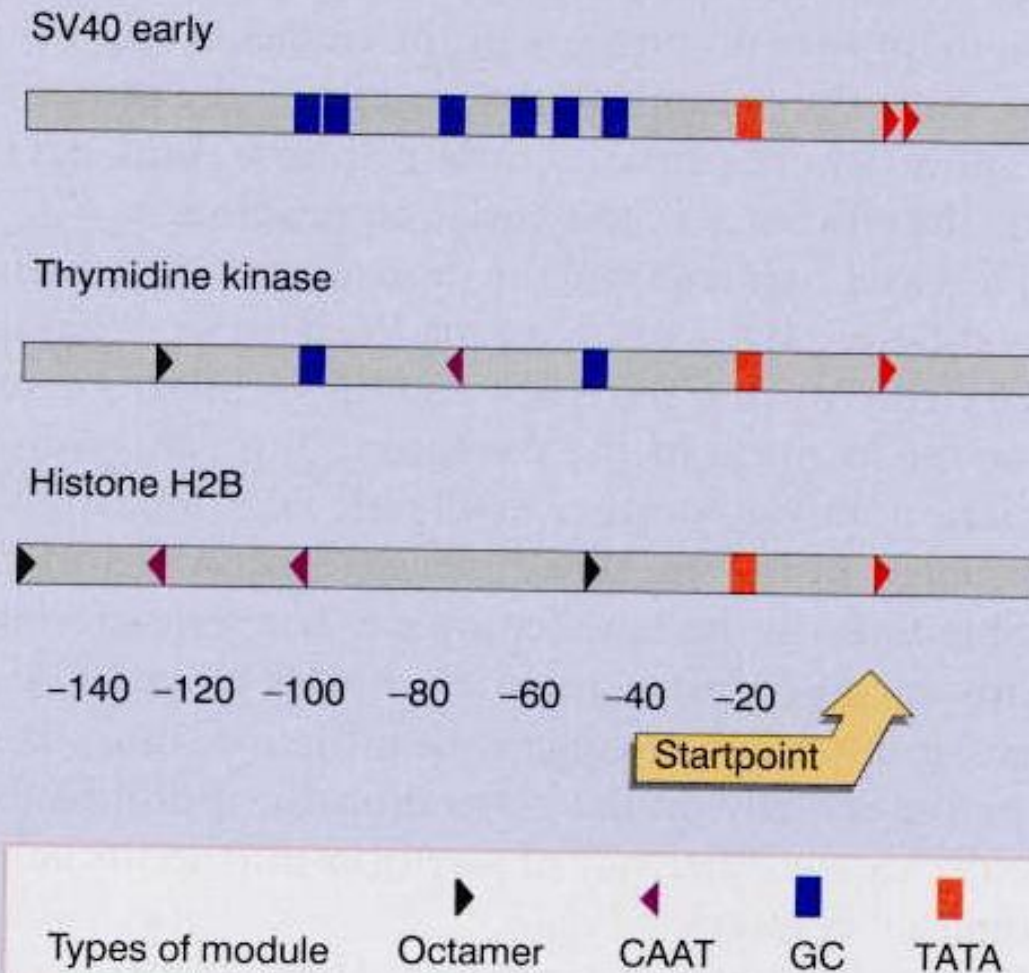
## Regulated Expression in Eukaryotes

### Complex Initiation System



*Figure 3.24* Formation of an RNA polymerase II transcription initiation complex at a TATA box. Transcription factor TFIID binds to a TATA box, and, in sequence, other transcription factors and RNA polymerase II bind to form a protein aggregate that is responsible for initiating transcription. The right-angled arrow designates the site of initiation and direction of transcription.

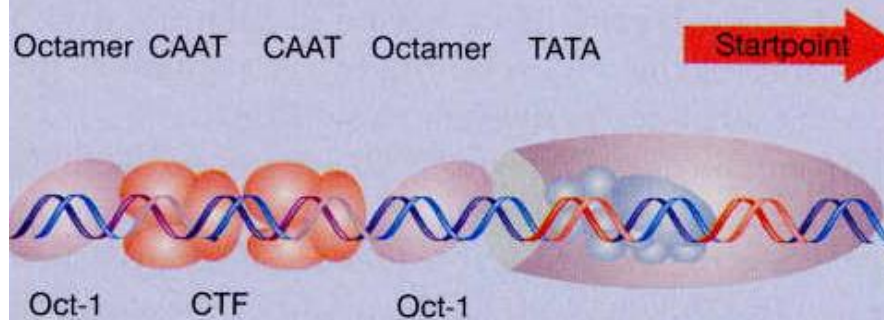
**Figure 20.17** Promoters contain different combinations of TATA boxes, CAAT boxes, GC boxes, and other elements.



**Figure 20.18** A transcription complex involves recognition of several elements in the sea urchin H2B promoter in testis. Binding of the CAAT displacement factor in embryo prevents the CAAT-binding factor from binding, so an active complex cannot form.

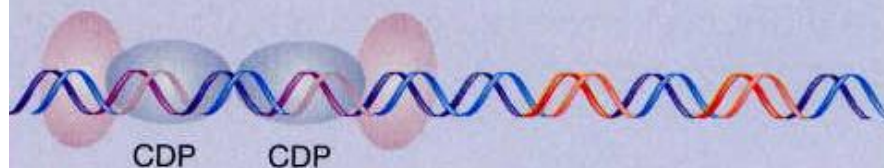
### Active

Transcription complex assembles in testis  
(not all components of basal apparatus are shown)



### Inactive

CDP prevents other factors from binding to CAAT box,  
and basal factors cannot bind





Essential elements are more concentrated in enhancers

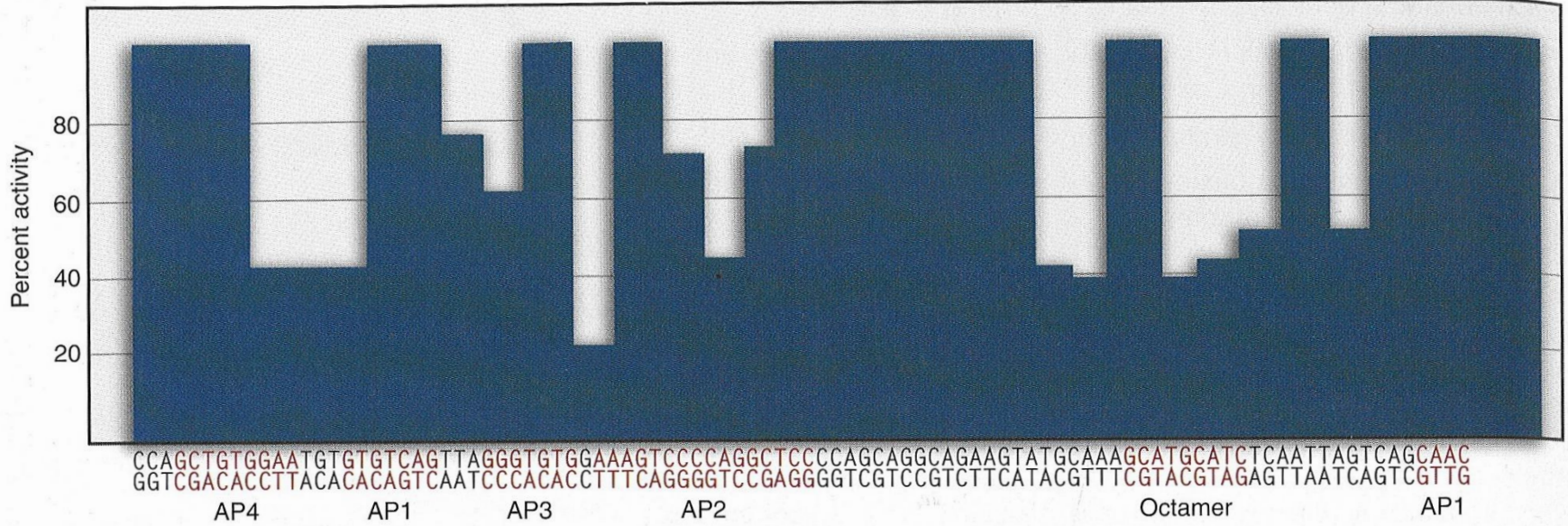


Figure 24.20 An enhancer contains several structural motifs. The histogram plots the effect of all mutations that reduce enhancer function to <75% of wild type. Binding sites for proteins are indicated below the histogram.