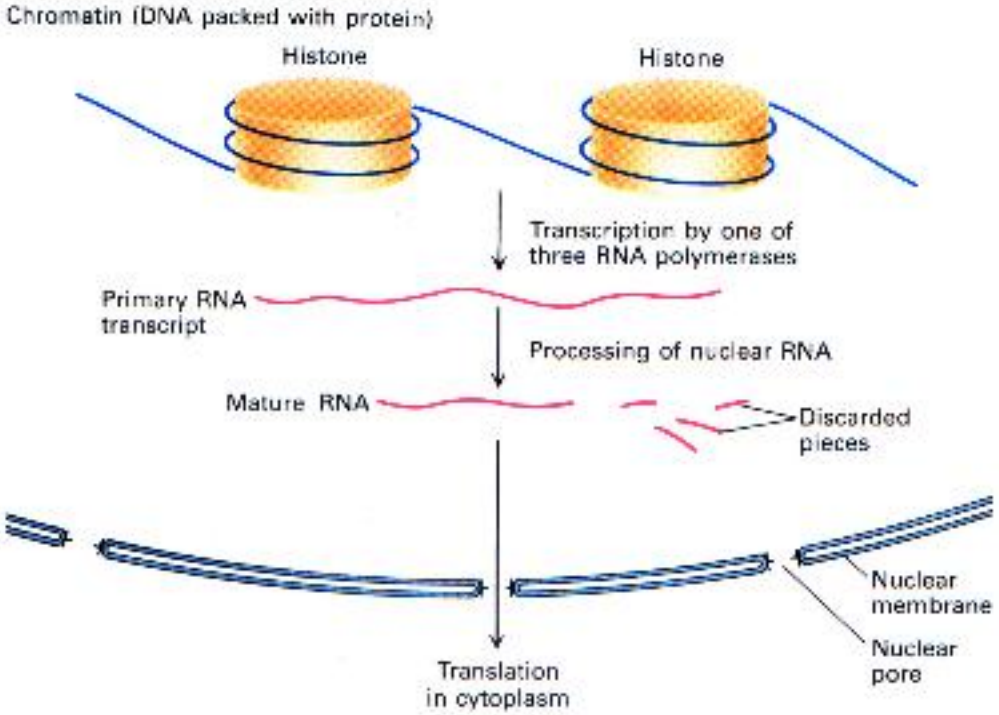
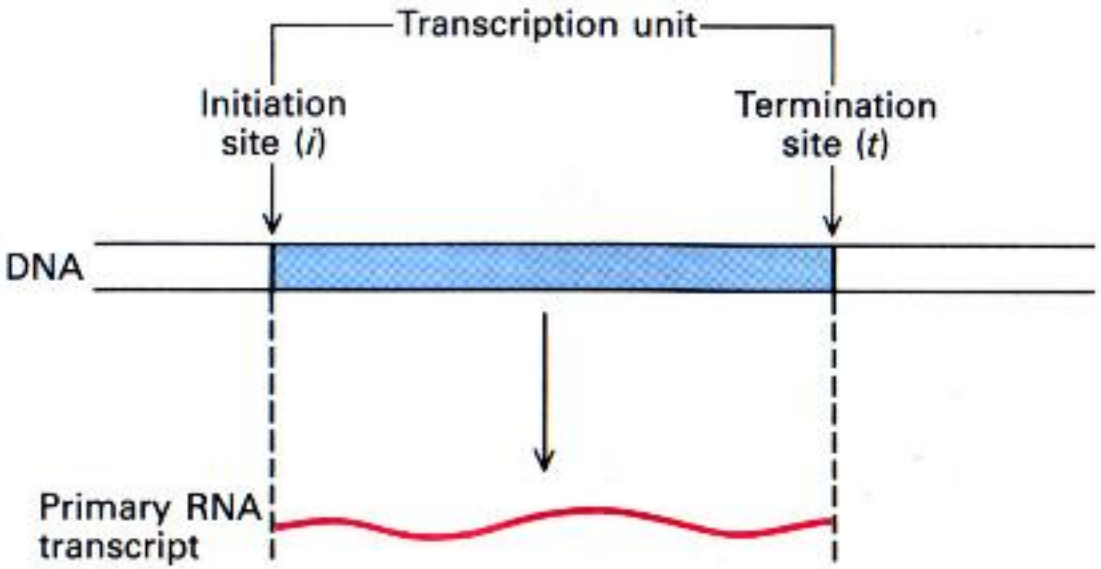
Синтез РНК и процессинг

Н.Н. Колесников

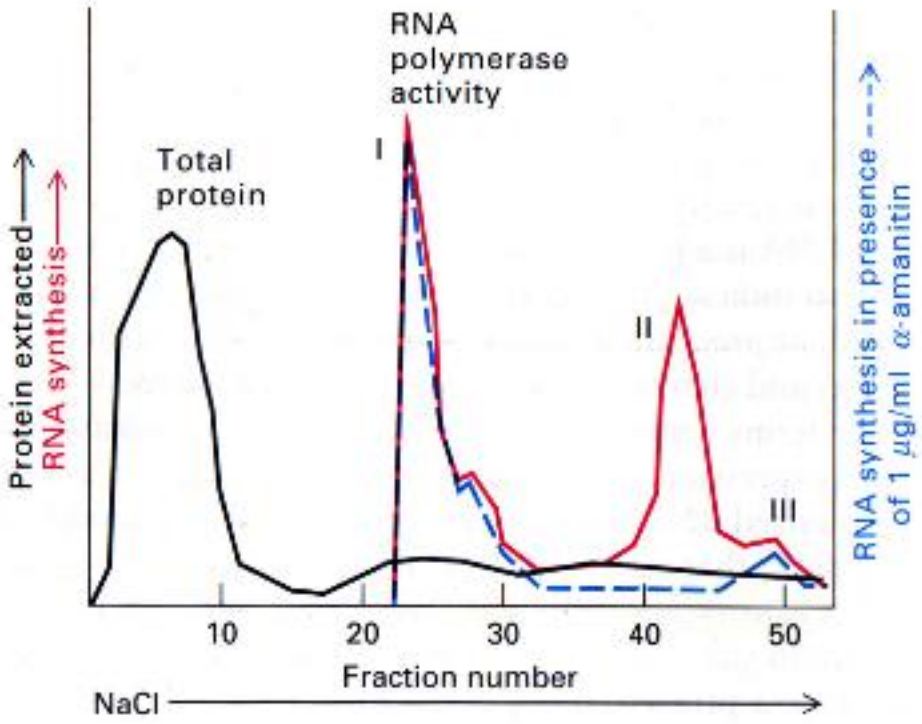
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia



◀ **Figure 8-1** Four features distinguish RNA production in eukaryotes from that in prokaryotes: (1) The DNA to be transcribed is wound around a histone core. (2) It is transcribed by one of three RNA polymerases. (3) The primary RNA transcript is processed within the nucleus into the finished RNA. (4) Only after transport of the RNA—presumably through a nuclear pore—does translation occur in the cytoplasm. In bacteria, all precursor rRNA and tRNA are processed but mRNA only rarely requires processing.



▲ **Figure 8-5** Transcription units in DNA have initiation sites and termination sites. Primary RNA transcripts are the initial unmodified RNA products.



▲ **Figure 8-6** The separation and identification of the three eukaryotic RNA polymerases by chromatographic analysis. A protein extract from the nuclei of cultured frog cells was passed through a DEAE Sephadex column to which charged proteins adsorb differentially, and adsorbed proteins were eluted (black curve) with NaCl solutions of increasing concentration. Successive fractions of the eluted proteins were assayed for the ability to transcribe DNA (red curve) in the presence of the four nucleoside triphosphates (including radioactive UTP). Most of the proteins did not bind to the column, but the enzymes did. The synthesis of RNA by each fraction in the presence of 1 μg/ml of α-amanitin also was measured in (blue curve). Polymerases I and III are insensitive to the compound at that concentration, whereas polymerase II is sensitive, that is, it ceases RNA synthesis. (Polymerase III is sensitive to 10 μg/ml of α-amanitin, however, although polymerase I is unaffected even by this higher concentration.) [See R. G. Roeder, 1974, *J. Biol. Chem.* **249:241.**]

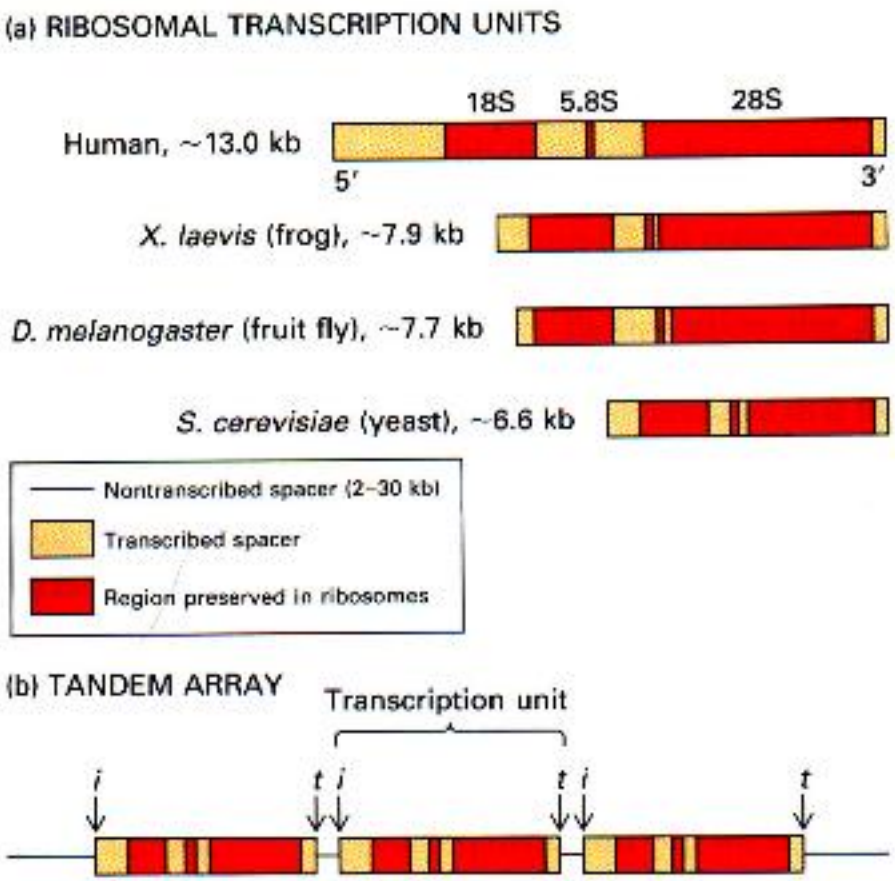


Figure 8-12 (a) The ribosomal transcription units of four eukaryotes. Variations in the lengths of the transcribed spacer regions (orange bars) account for the major differences in the lengths of the transcription units. The regions that appear in ribosomes are always 18S, 5.8S, and 26 to 28S. (b) The genomes of all animals contain multiple tandem copies of the rRNA transcription units. The nontranscribed spacer regions between transcription units can vary greatly, e.g., ~2 kb in frogs to ~30 kb in humans.

Table 8-1 Lengths of primary rRNA transcripts and cytoplasmic rRNAs in various species

	Primary transcript*		Ribosomal RNA length*		
	S value	Length (kb)	26S–28S (kb)	16S–18S (kb)	Percentage of precursor preserved
Escherichia coli (prokaryote)	30	6.0	3.0	1.5	75
Saccharomyces cerevisiae (yeast)	37	6.6	3.8	1.7	77
Dictyostelium discoideum (slime mold)	37	7.4	4.1	1.8	80
Drosophila melanogaster (fruit fly)	34	7.7	4.1	1.8	76
Xenopus laevis (frog)	40	7.9	4.5	1.9	81
Gallus domesticus (chicken)	45	11.2	4.6	1.8	57
Mus musculus (mouse)	45	13.7	5.1	1.9	51
Homo sapiens (human)	45	13.7	5.1	1.9	51

*The lengths of the various RNA molecules are estimates based on gel electrophoresis and direct measurements of electron micrographs. The size is of the first major product with definite 5′ and 3′ ends.

SOURCE: B. Lewin, 1980, *Gene Expression*, vol. 3, Wiley, p. 867.

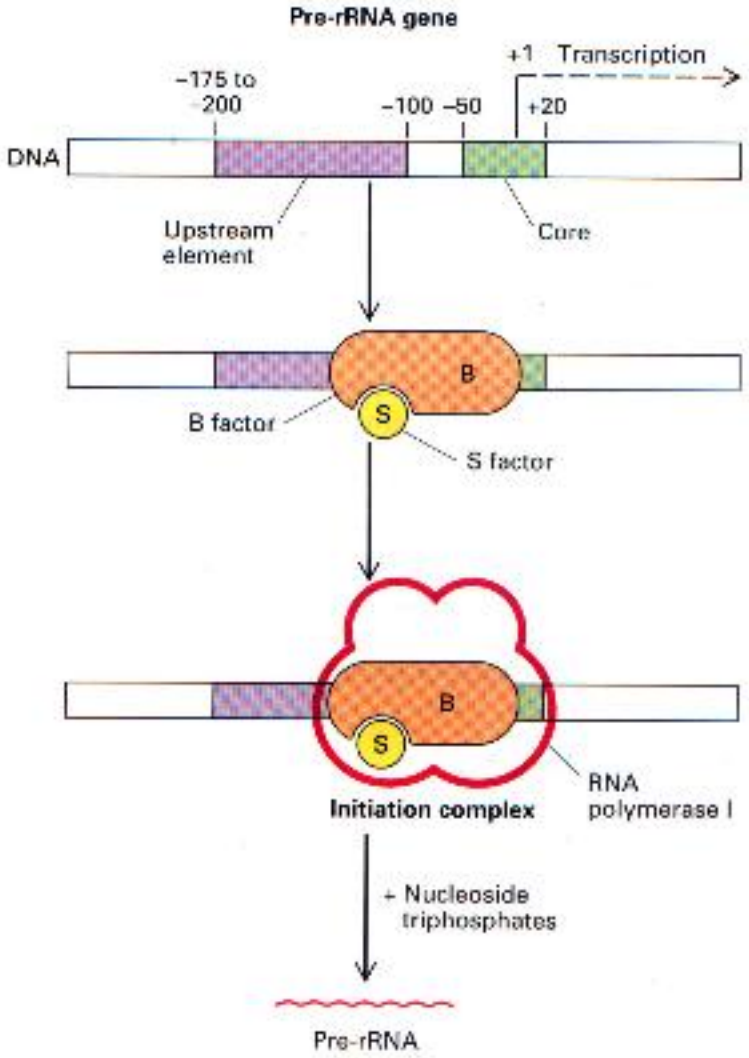
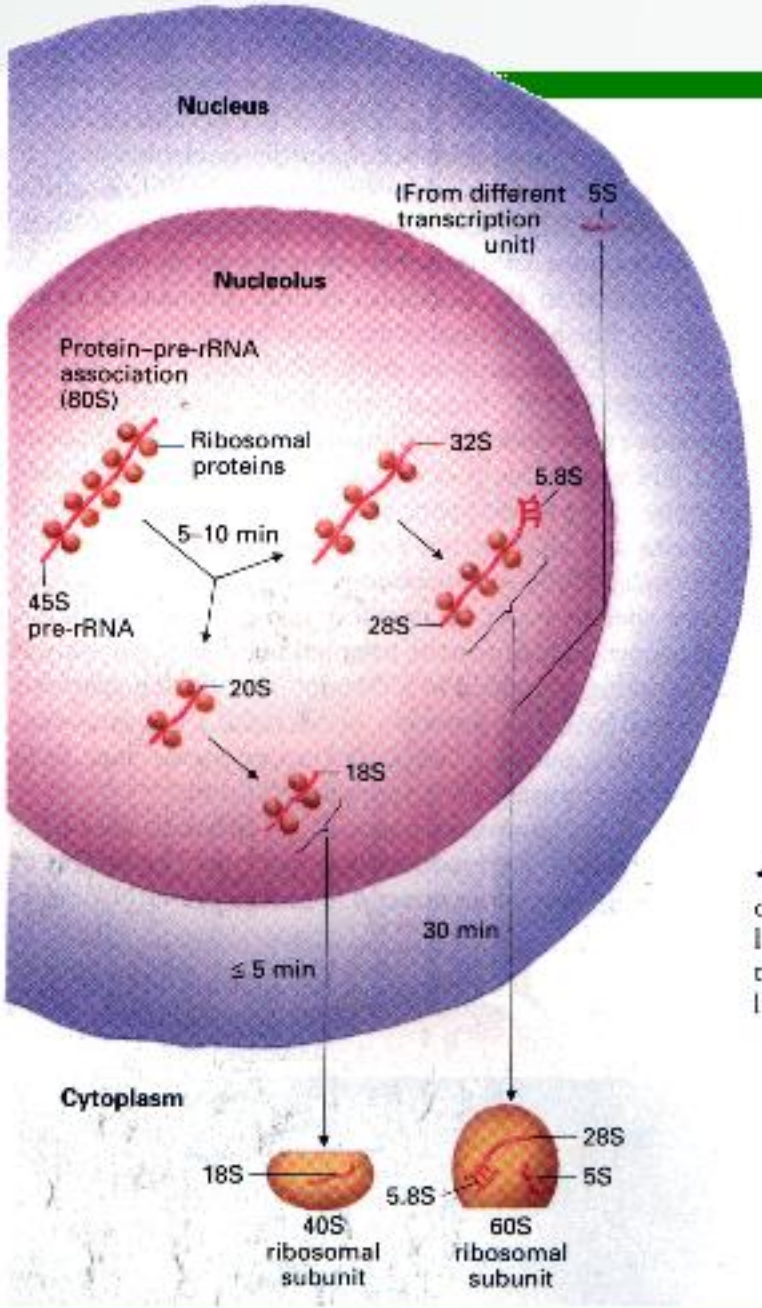
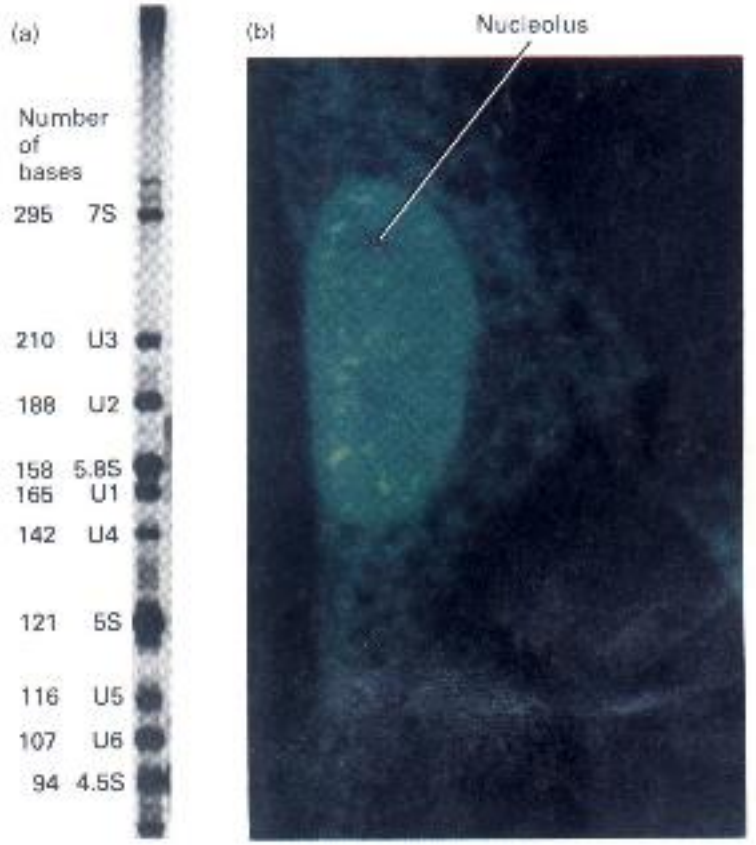


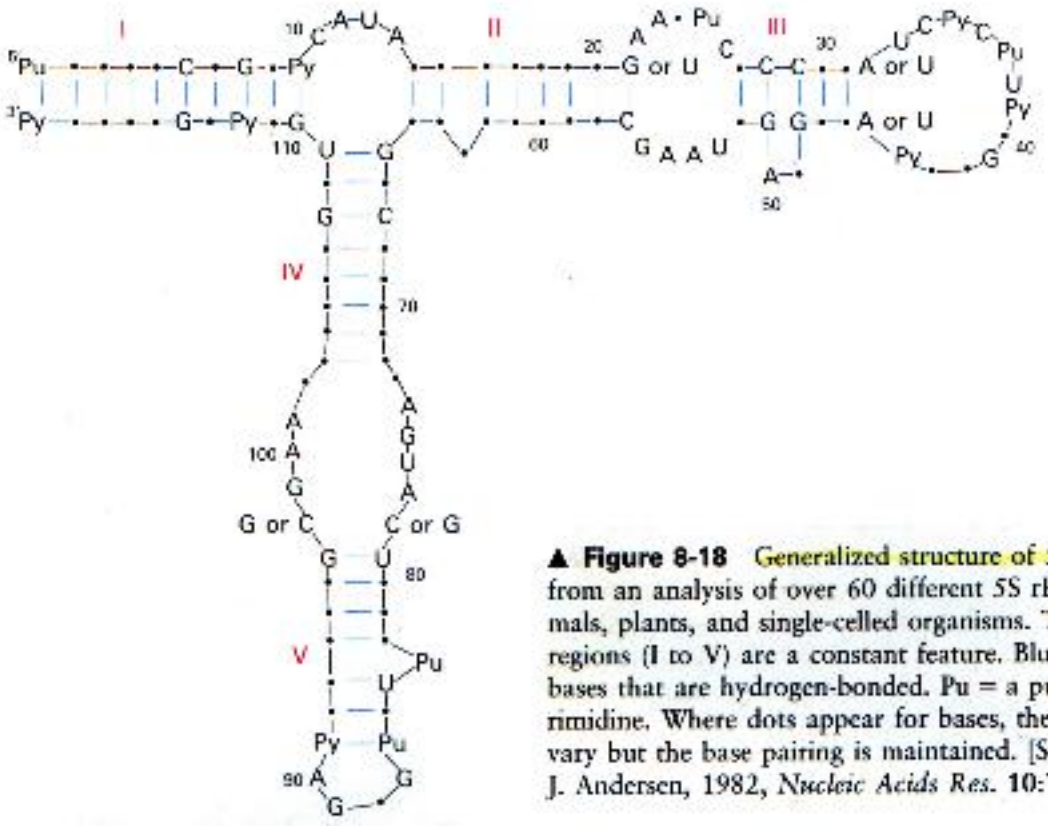
Figure 8-13 Formation of initiation complex for transcription of rDNA by RNA polymerase I. The DNA and B and S factors must be from the same species (e.g., mouse or human), although the polymerase can be from a different species. Both mouse and human pre-rRNA transcription units have a core promoter sequence of about 70 bases and a necessary upstream DNA sequence or element. Removal of the core completely abolishes transcription; removal of the upstream element decreases transcription 10- to 20-fold. These sequences are not the same in the two DNAs except for an identical sequence in the core region extending from the start site (+1) to about +20. Thus there are species-specific DNA sequences and species-specific binding factors. The S factor increases the affinity of the B factor for DNA but does not bind directly to DNA. [See S. Bell et al., 1988, *Science* **241**:1192.]



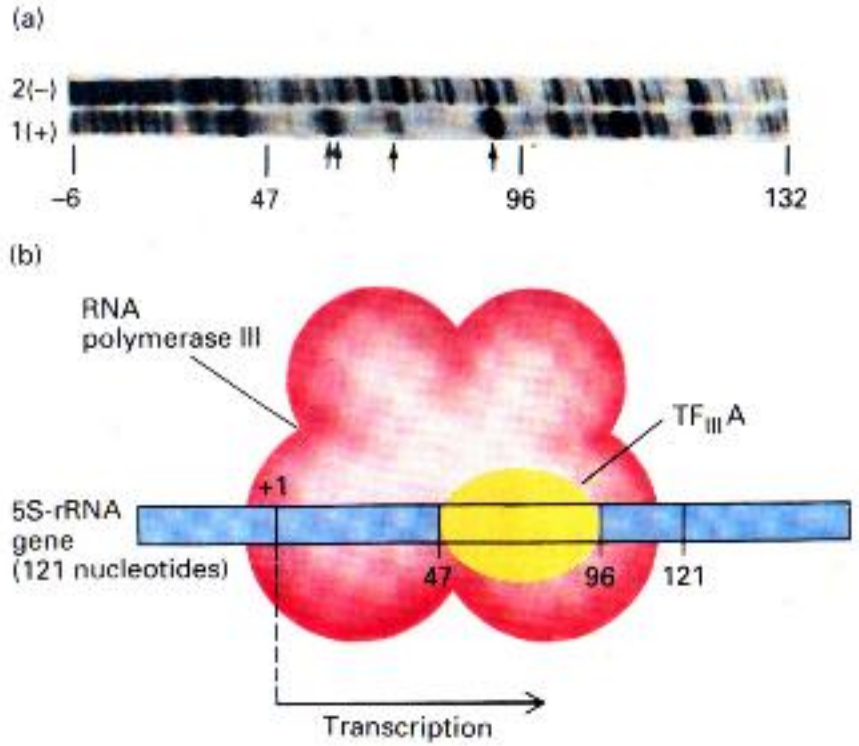
◀ **Figure 8-15** Cleavages and times required in processing of pre-rRNA into ribosomal subunits. Ribosomal and nucleolar proteins associate with 45S pre-rRNA soon after it is synthesized. Synthesis of 5S rRNA occurs outside of the nucleolus.



▲ **Figure 8-17** (a) The small RNA species from cultured Chinese hamster cells. RNA labeled with $^{32}PO_4$ was purified from the nuclei of cultured cells and subjected to gel electrophoresis and autoradiography. Almost all of the RNAs identified here have been sequenced. The species labeled U1 to U6 are rich in uridylates. All except U3 are involved in mRNA splicing; U3 is found mainly in the nucleolus. The 5.8S and 5S species are ribosomal RNAs; the 7S RNA belongs to a particle that assists in transporting proteins through membranes. These last three RNAs are present mainly in the cell cytoplasm. (b) Antibody stain of nuclear ribonucleoproteins that participate in splicing the antigen termed Sm is present in particles containing U1, U2, U4, U5, and U6 small RNAs. Antigen is not present in U3, which is evidenced here by an absence of staining in the nucleolus. *Part (a) courtesy of W. Jelinek and S. Haynes; part (b) courtesy of M. Lerner.*

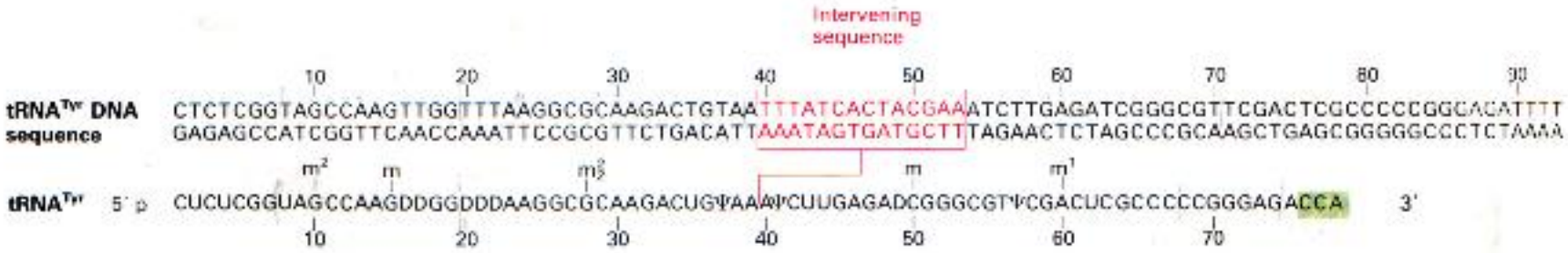


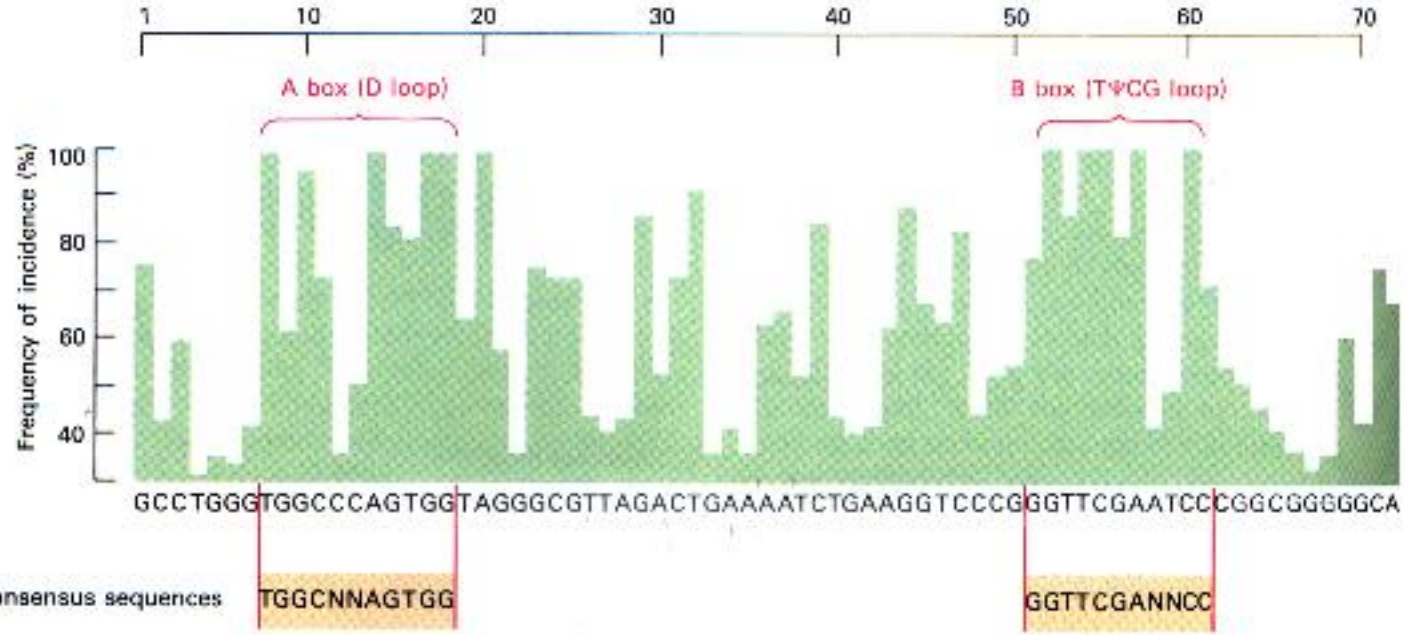
▲ **Figure 8-18** Generalized structure of 5S rRNA obtained from an analysis of over 60 different 5S rRNAs from animals, plants, and single-celled organisms. The five stem-loop regions (I to V) are a constant feature. Blue lines connect bases that are hydrogen-bonded. Pu = a purine; Py = a pyrimidine. Where dots appear for bases, the sequence may vary but the base pairing is maintained. [See N. Delihas and J. Andersen, 1982, *Nucleic Acids Res.* **10**:7323.]



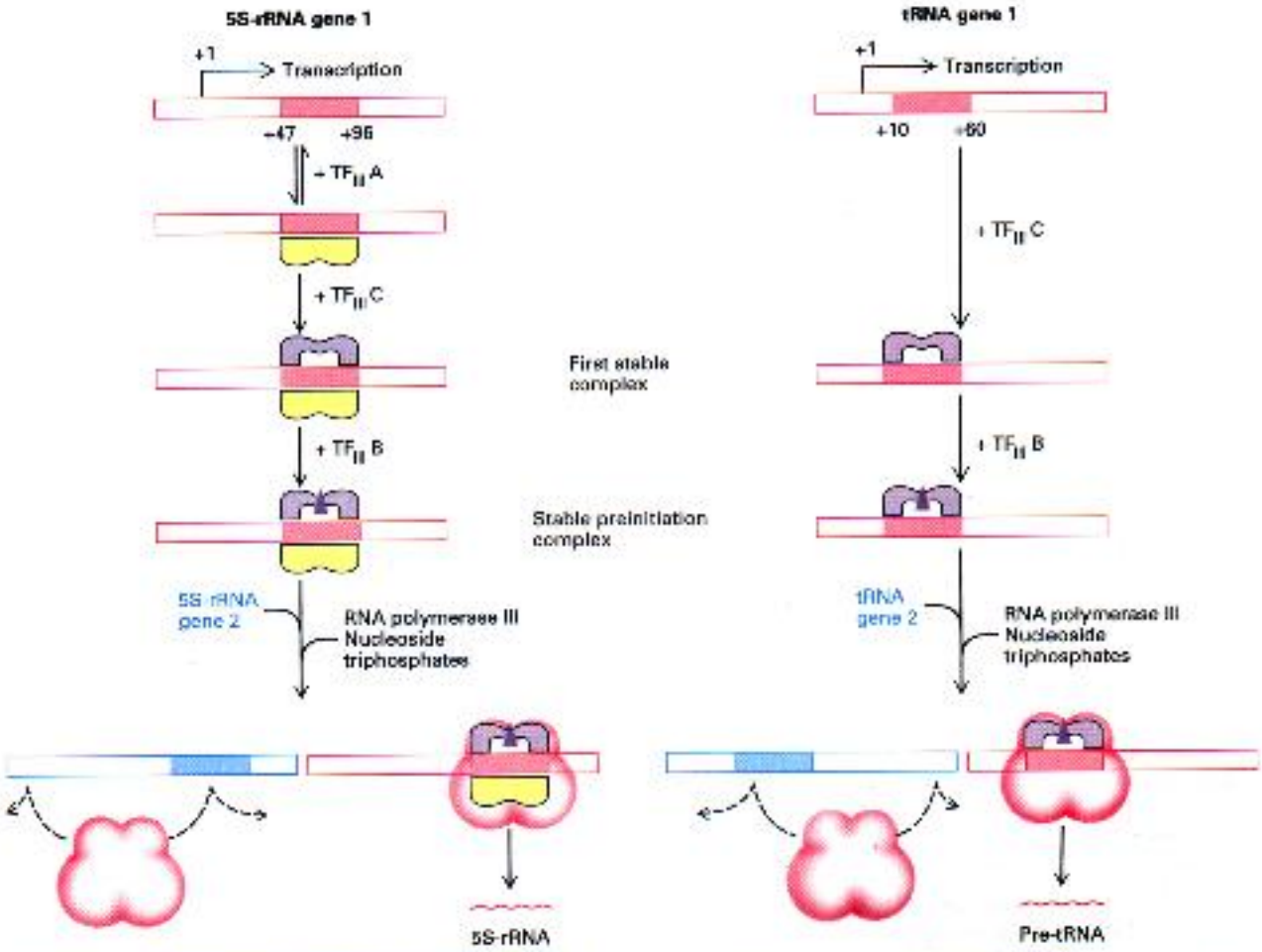
▶ **Figure 8-20** Locating $TF_{III}A$-binding region on 5S-rRNA gene by footprint experiment. Purified $TF_{III}A$ was mixed with 5′ end-labeled 5S rDNA, and the mixture, as well as a control without the transcription factor, was digested briefly with DNase, producing a series of fragments stretching from the 5′ end to the cleavage site. (a) Autoradiograph of DNase-treated mixture containing $TF_{III}A$ *(lane 1)* and without factor *(lane 2)*. The lack of bands in the region from base 47 to base 96 on the 5S rDNA in lane 1 indicates that these bases were protected from digestion by bound protein. The four bands that appear in lane 1 *(small arrows)* but not in lane 2 represent DNA regions that become "hypersensitive" to DNase attack as a result of binding the transcription factor. (b) Diagram of 5S-rRNA gene showing regions where $TF_{III}A$ and RNA polymerase III bind. [See R. G. Roeder, 1980, *Cell* 19:717.] *Photograph courtesy of R. G. Roeder.*

▼ **Figure 8-21** The sequences of a tyrosine tRNA gene from yeast and of corresponding $tRNA^{Tyr}$ found in the cytoplasm of yeast cells. The DNA contains a 14-base intervening sequence that does not appear in the tRNA. (Note that the CCA 3′ terminus of the mature tRNA is not encoded in the DNA.) The modified bases in the tRNA are as follows: D = dihydrouridine, Ψ = pseudouridine, m = methyl group on ribose, m^2 = methyl on position 2 of base, m_2^2 = dimethyl on position 2 of base, and m^1 = methyl on position 1 of base. [See H. M. Goodman, M. V. Olson, and B. D. Hall, 1977, *Proc. Nat'l Acad. Sci. USA* 74:5453.]





▲ **Figure 8-23** Frequency of occurrence (%) of most common nucleotides in 80 different eukaryotic tRNA genes. Two consensus sequences, called the A box and the B box, are present. The A box encodes the D loop of tRNA, and the B box encodes the TΨCG loop; both loops are constant features of tRNA. The A and B boxes also serve as parts of the promoter in tRNA genes. [See D. H. Gauss and M. Sprinzl, 1981, *Nucleic Acids Res.* 9:r1; and G. Galli, H. Holstetter, and M. L. Birnstiel, 1981, *Nature* 294:626.]



▲ **Figure 8-24** Formation of preinitiation complexes for genes transcribed by RNA polymerase III. 5S genes require factors TF$_{III}$A, B, and C for transcription by RNA polymerase III, whereas tRNA genes require only factors B and C. In both cases, after factor addition to a mixture containing excess template DNA, stable complexes are formed. At this stage, addition of any other template DNAs (e.g., 5S gene 2 or tRNA gene 2 [blue]) does not displace factors already bound to first template complexes and thus does not affect their transcription. Second templates are not transcribed. [See A. B. Lasser, P. L. Markin, and R. G. Roeder, 1983, *Science* 222:740.]

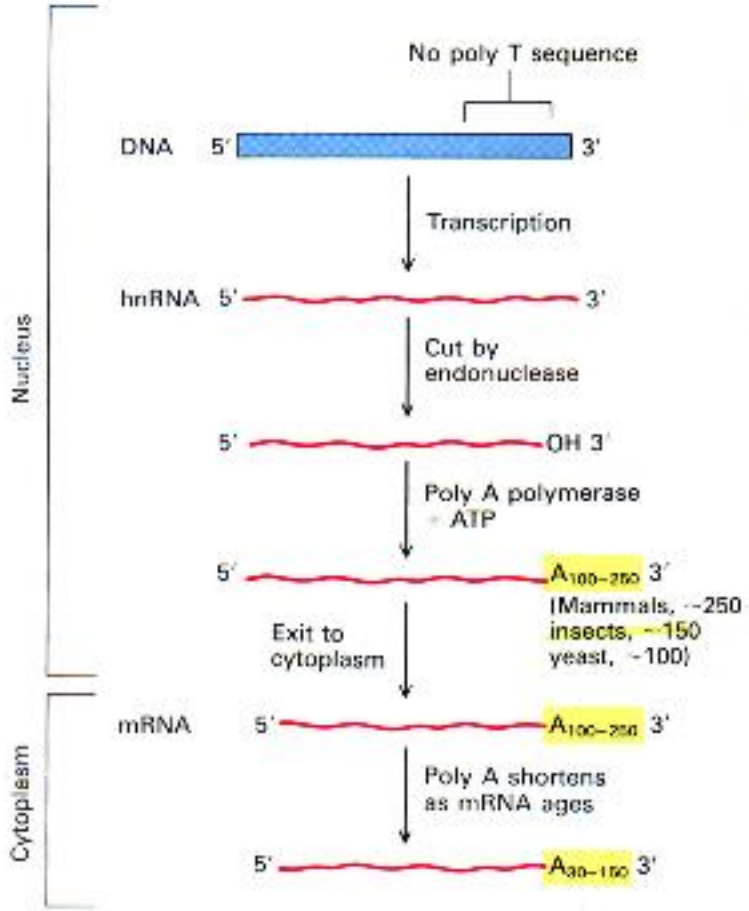
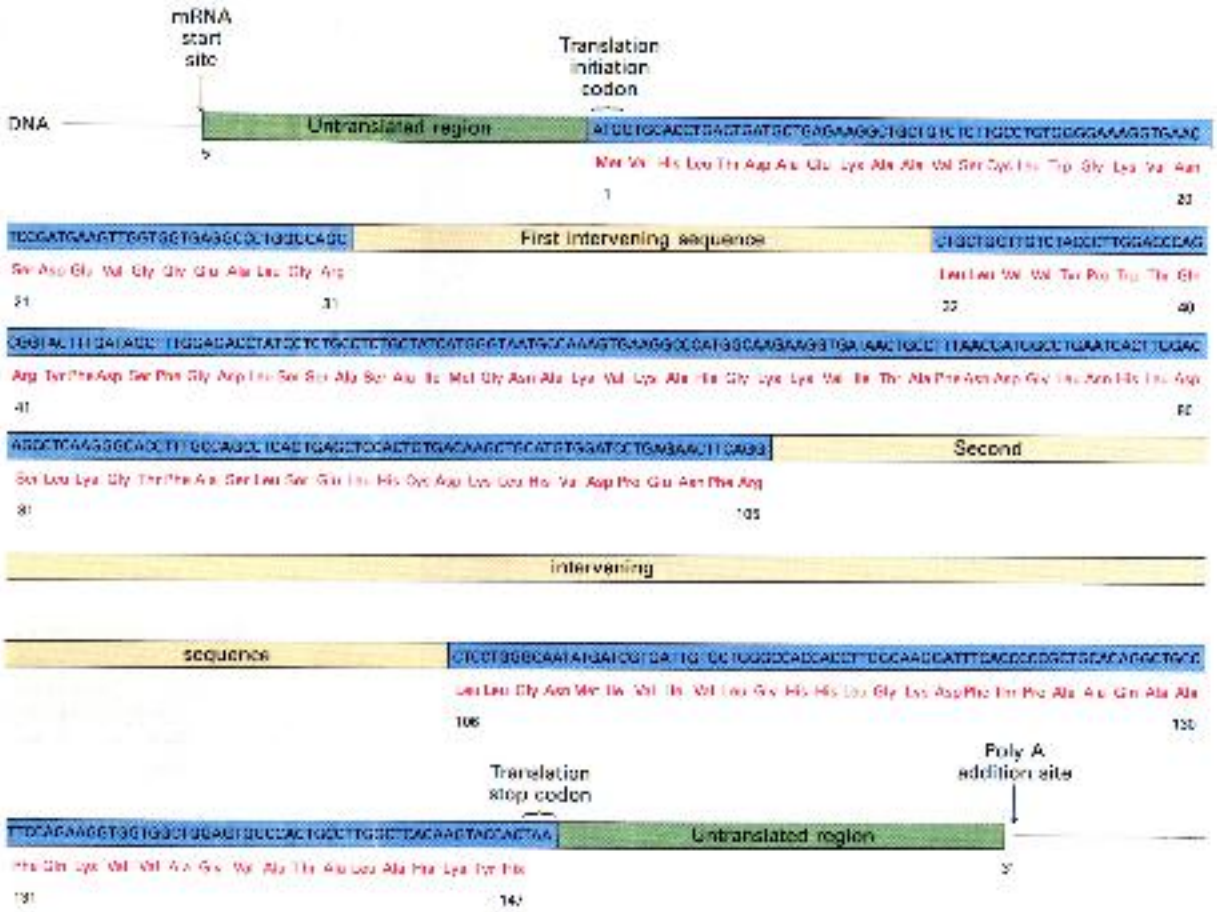
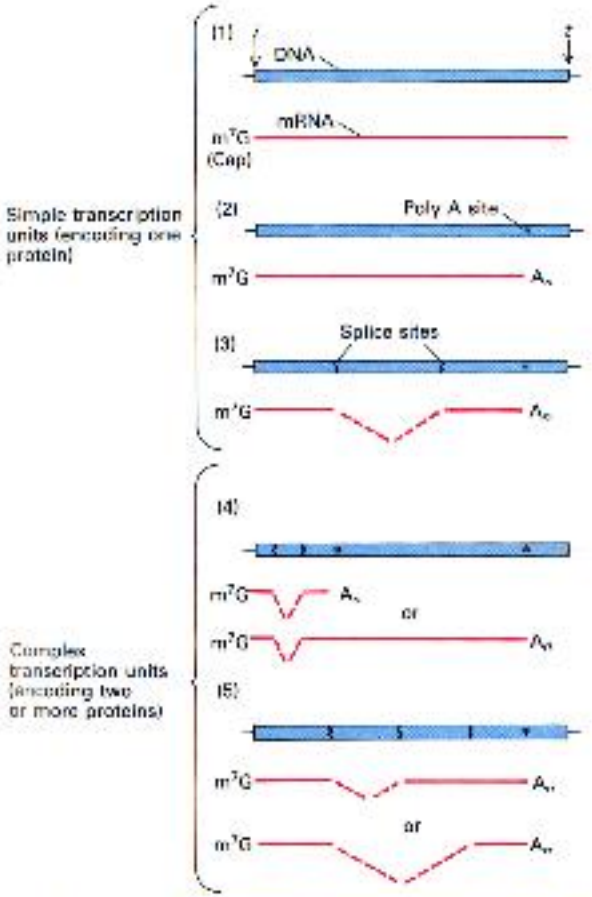


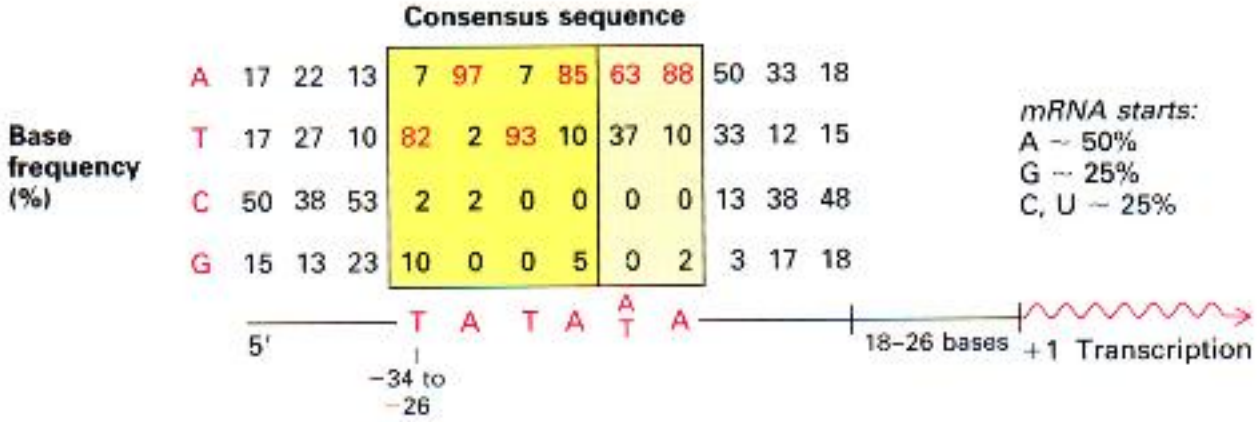
Figure 8-28 Steps in the addition of poly A during formation of eukaryotic mRNA. Note that the poly A sequence is not encoded in DNA. Cleavage at the 3' end of hnRNA followed by addition of A residues one at a time occurs in the nucleus. The only known eukaryotic mRNAs whose processing does not include poly A addition are those for histone proteins; however, because 3' A residues are lost as cytoplasmic mRNAs age, mRNAs with very short poly A are observed.



▲ **Figure 8-38** The β-globin gene has two intervening sequences. The 147 amino acids in the β-globin sequence were known; the DNA copied from the β-globin mRNA contained 5′ and 3′ untranslated regions and the exact coding information for the known amino acid sequence. The genomic DNA contained the coding sequence broken into three parts by two intervening sequences between amino acids 31 and 32 and 105 and 106. [See D. A. Konkel, S. Tilghman, and P. Leder, 1978, *Cell* 15:1125.]



▲ **Figure 8-39** Diagrams and examples of simple and complex transcription units encoding mRNAs. Each transcription unit is bounded by an initiation (*i*) and a termination (*t*) site; the mRNAs derived from the corresponding primary transcript are shown in red. Poly A sites are indicated by asterisks, and splice sites by zigzags.



▲ **Figure 8-41** Comparison of nucleotides upstream of start site in 60 different eukaryotic protein-coding genomes. Each sequence was aligned to yield maximum homology in the region from −35 to −20. The tabulated numbers give the percentage frequency of each base at each position. The maximum homology occurs over a six-base consensus sequence in which the first four bases are TATA.

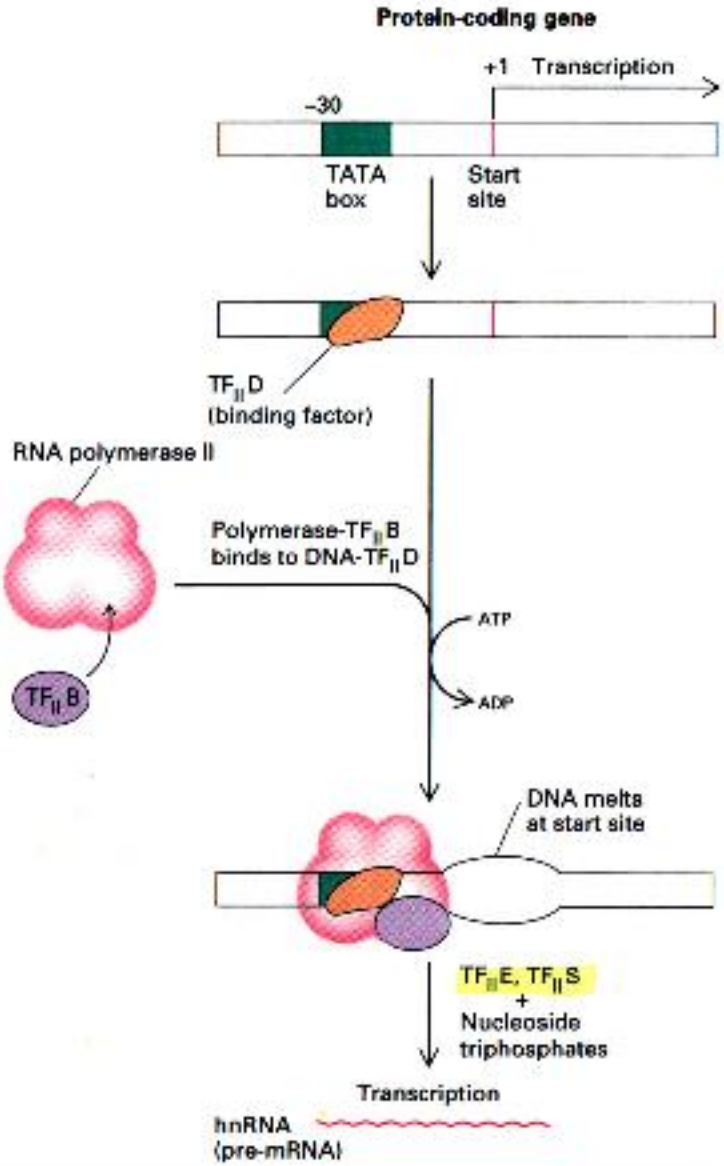


Figure 8-43 Model of formation of preinitiation complex for protein-coding genes containing a TATA box. The TATA-binding factor $TF_{II}D$ (orange) can bind on its own sufficiently tightly to form a stable complex. In other genes, it binds cooperatively with RNA polymerase II. Neither $TF_{II}B$ (purple) nor $TF_{II}E$ binds directly to DNA. $TF_{II}B$ can bind to RNA polymerase II in the absence of DNA. Once the polymerase-$TF_{II}B$ complex binds to the DNA-$TF_{II}D$ complex, it acts as an ATPase, perhaps causing the DNA to melt near the start site.

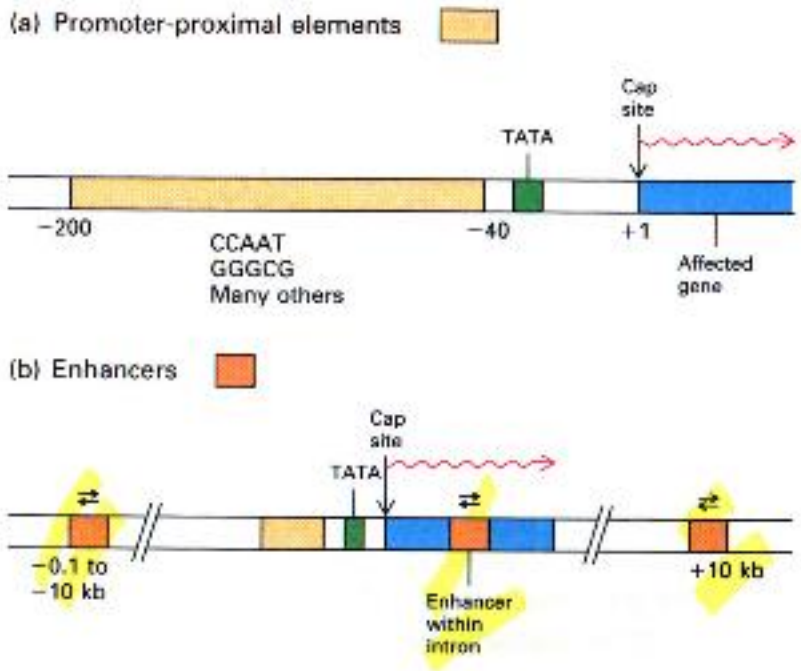
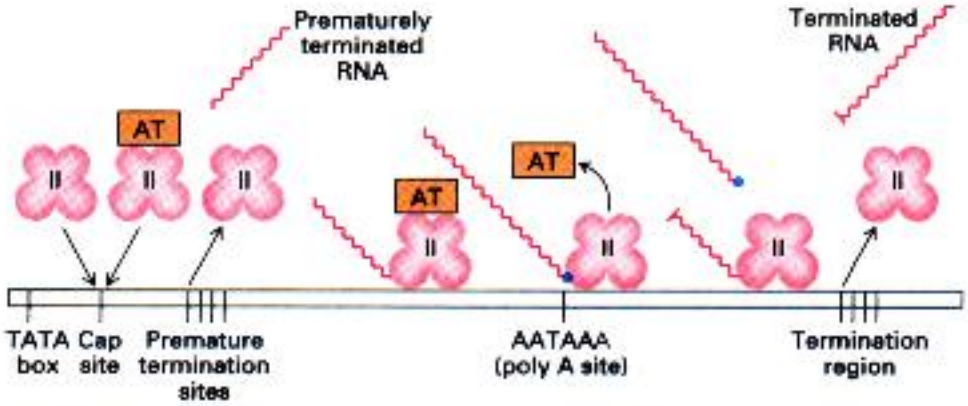


Figure 8-45 The classes of eukaryotic transcriptional activating elements. Many experiments with recombinant genes (of the type outlined in Figure 8-44) have shown that maximal transcription requires (a) upstream DNA regions close to the RNA start site, termed promoter-proximal elements, and (b) enhancers, which can be close to or far from RNA start sites (either upstream or downstream) and can exist in either orientation in the DNA. Almost all enhancer regions represent clusters of binding sites, which can bind several proteins. The most common promoter-proximal elements in the region −40 to −200 are CCAAT and GGGCG, but about 70 percent of genes so far examined have neither of these elements. Although most of the experiments that defined the two general activating elements were done in mammalian cells with genes from mammals or viruses infecting mammals, the same elements also exist in invertebrate genes. In the yeast genome, which is much smaller, the activating elements tend to be just upstream from the affected gene but often share with enhancers the property of functioning in either orientation. The yeast sequences are most often termed upstream activating sequences, or UASs. [See P. Bucher and E. N. Frifonov, 1986, *Nucleic Acids Res.* **14**:10009.]



▲ **Figure 8-46** Model for a mechanism coupling addition of poly A (A_n) and termination of transcription. The model proposes that two classes of RNA polymerase II transcription complexes (red) are directed to (cap) sites by upstream regulatory elements (including the TATA motif). The two classes differ by the presence or absence of an antiterminator (AT) factor (orange). Complexes lacking the AT factor cease transcription at premature termination sites. Complexes including the AT factor read through premature stop sites, but such complexes lose the factor at a poly A addition site (AATAAA) and then cease transcription at the next termination region.