GENE EXPRESSION

Ho Huynh Thuy Duong

University of Science

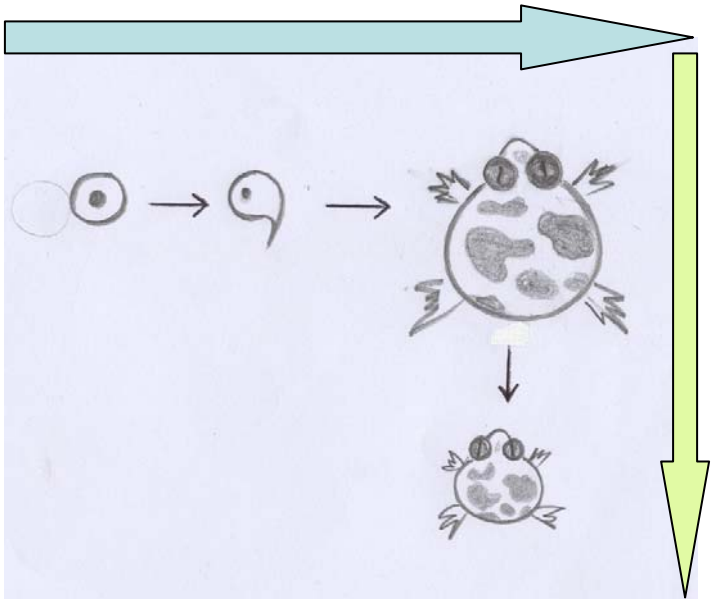
GENE EXPRESSION

TRANSCRIPTION

TRANSLATION

According to the Central Dogma there are two information flows : "… ***The transfer of information from nucleic acid to nucleic acid, or from nucleic acid to protein,*** **…" (Francis Crick, 1958).**

DECODING GENETIC INFORMATION GIVING RISE TO AN ORGANISM
(Transcription – Translation)



PERPETUATION OF GENETIC INFORMATION FROM GENERATION TO GENERATION
(DNA Replication)

HOW IS GENETIC INFORMATION CODED IN THE FORM OF NUCLEOTIDE STRETCHES CAN BE EXPRESSED INTO POLYPEPTIDE CHAINS, THAT IS, INTO PHENOTYPIC CHARACTERISTICS OF A LIVING ORGANISM ?

TRANSCRIPTION

- **GENERAL FEATURES OF TRANSCRIPTION PROCESS – TRANSCRIPTION IN PROKARYOTES**
  - **PROMOTER**
  - **RNA POLYMERASE**

- **TRANSCRIPTION IN EUKARYOTES**
  - **PROMOTERS**
  - **RNA POLYMERASES I, II, III**
  - **TRANSCRIPTIONAL FACTORS**

SOME CHARACTERISTICS OF THE TRANSCRIPTION PROCESS

Transcription is the synthesis of RNA. Chemically, transcription is similar to replication – that is the synthetic process generates a nucleic acid strand complementary to the template.

WHAT ARE THE DIFFERENCES BETWEEN THE TWO PROCESSES ?

ᔓ Transcription differs from replication in some features : (1) transcription does not need a primer to be initiated, (2) The RNA product does not base pairs with the DNA template, (3) RNA synthesis requires ribonucleotides rather than deoxyribonucleotides, with U replacing T.

ᔓ The purposes are different between the two processes – replication aims at perpetuating genetic information from generation to generation whereas transcription allows expression of these informations during the life time of an organism

→ Replication is regulated so that it copies the whole genome once and only once at each cell cycle. In transcription, only a portion of the genome is synthesized at one moment and the amount of RNA copies synthesized from this part varies depending on internal and external regulatory factors.

→ Transcription is less acurate than replication

ᔓ Unlike replication which requires many factors, transcription is performed by one enzyme called RNA polymerase, which assumes many functions

TRANSCRIPTION IN PROKARYOTES – *E. COLI* RNA POLYMERASE

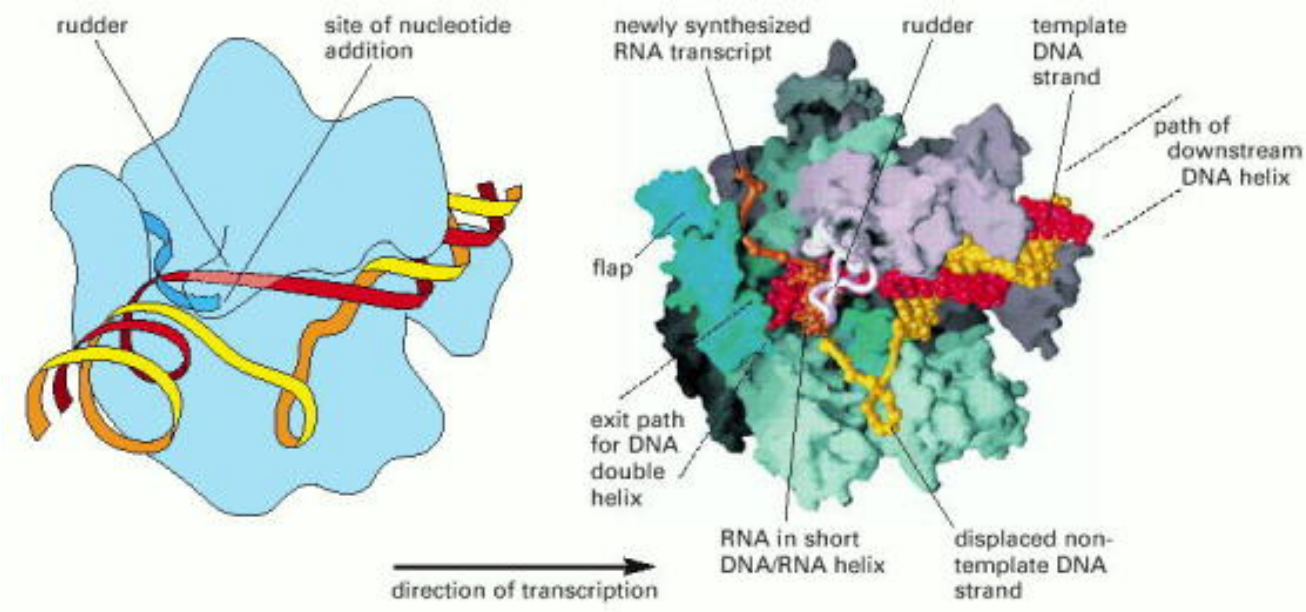
RNA polymerases found in different organisms have many common characteristics in shape and structure as well as in functions

ᔓ In *E.coli*, one RNA polymerase transcribes all bacterial genes.

ᔓ Bacterial RNA polymerase includes two components (1) & (2) ; both constitute the holoenzyme :

(1) The core enzyme composed of two α subunits, one β, one β' and one ω subunits

(2) σ factor



RNA pol subunits :

- α subunits (in green) : association of the multiple subunits, promoter recognition, binding some regulatory proteins

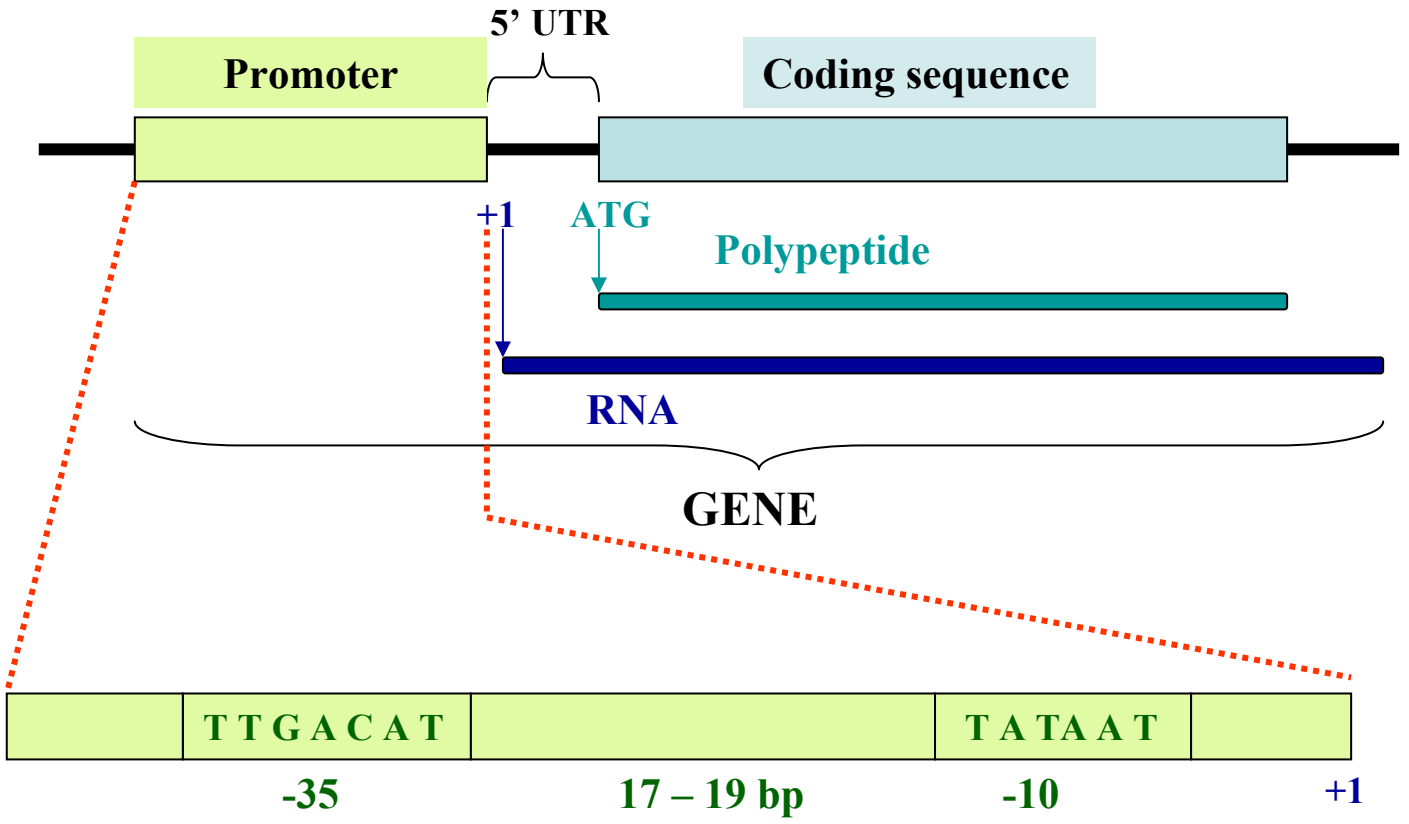
- β, β' subunits (in purple) : Constitution of active site where ribonucleotides are incorporated in the RNA growing chain and base pairing with the DNA template (in red)

April 2009

HOW DOES TRANSCRIPTION OF A GENE BE ACCURATE – BEGINNING AND ENDING AT DEFINED SITES ?

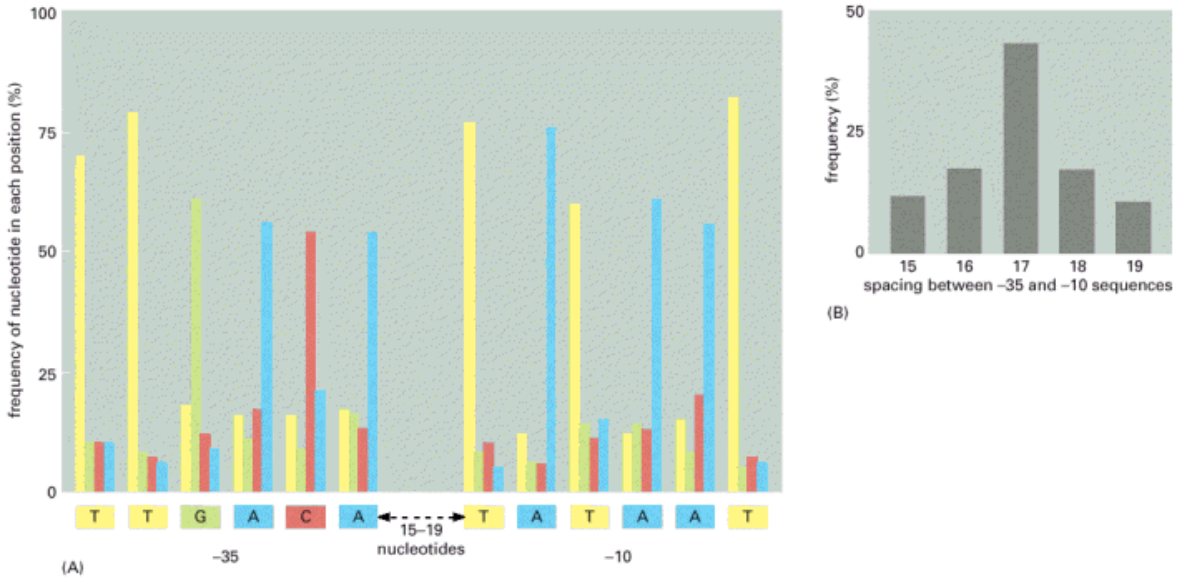
THE TRANSCRIPTION OF A GENE BEGINS AT A SEQUENCE NAMED PROMOTER

STRUCTURE OF PROKARYOTIC PROMOTER



Consensus sequence for most *E. coli* promoters

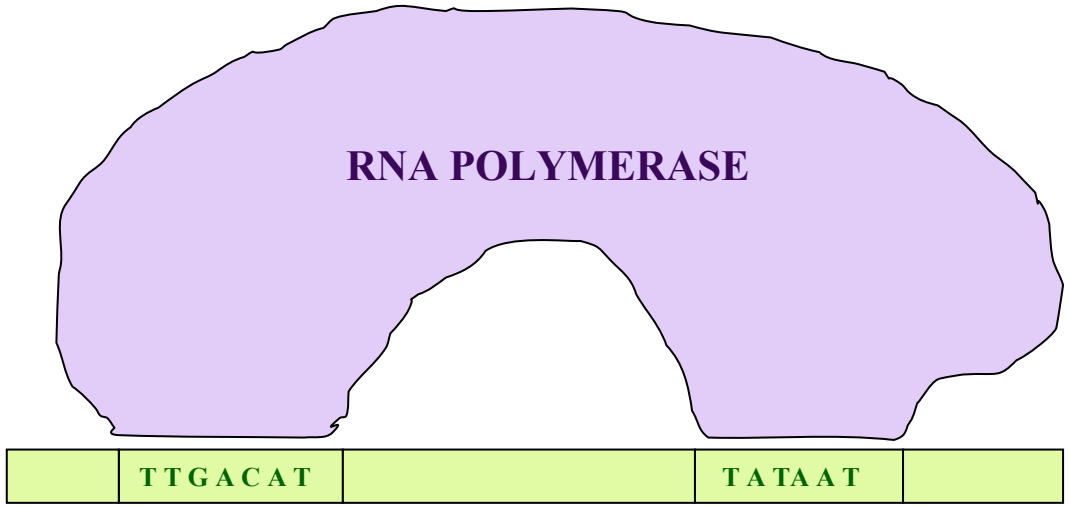
Almost all promoters have : (1) two conserved nucleotide sequences "-35" and "-10" named according to their position towards the transcription beginning site (+1), and (2) nearly conserved distance "17-19 bp" between "-35" and "-10"



"Copyright 2002 from Molecular Biology of the Cell by Alberts et al. Reproduced by permission of Garland Science/Taylor & Francis LLC."

The consensus sequence is defined by comparison of many *E. coli* promoters. Each nucleotide of the conserved sequences is the nucleotide which has the most important frequency among *E. coli* promoters

WHAT ARE THE ROLES OF THESE CONSERVED SEQUENCES AND DISTANCE ?



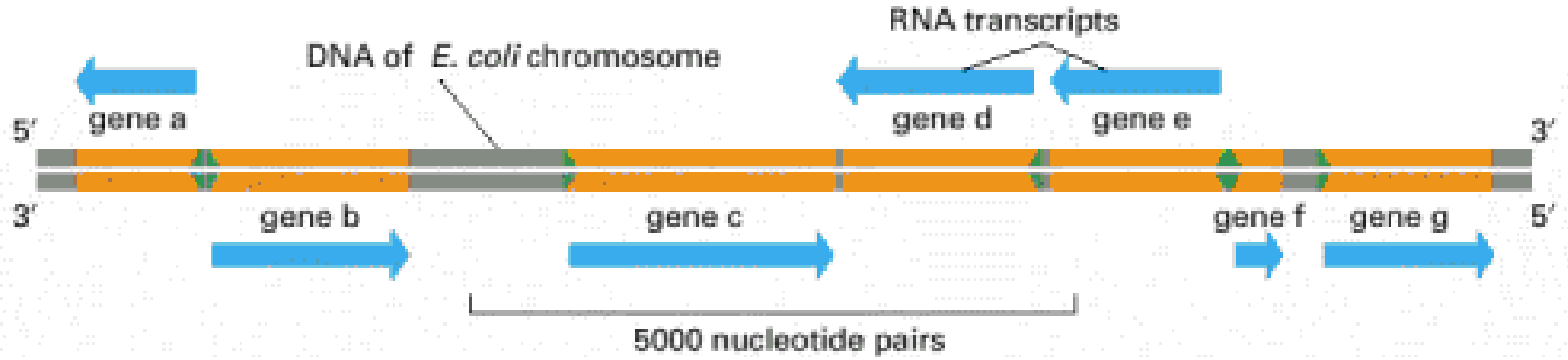
Conserved sequences are contact sites between promoter sequence and RNA polymerase.

The conserved distance between "-35" and "-10" regions fits the enzyme shape.

The picture in the left represents an *E. coli* RNA polymerase binding to the promoter

THE TRANSCRIPTION USES ONE DNA STRAND AS TEMPLATE

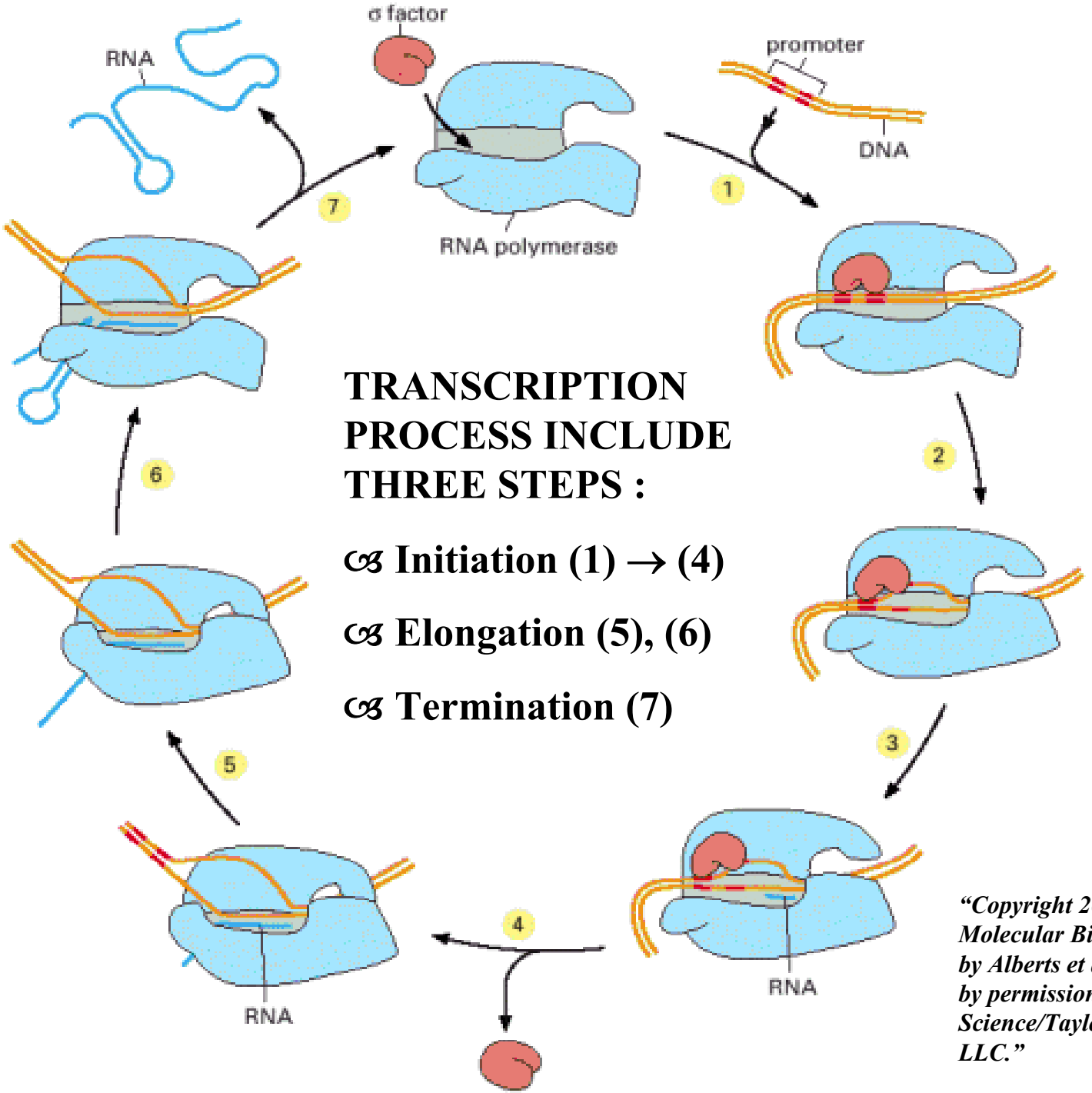
WHICH STRAND IS CHOSEN TO SERVE AS TEMPLATE FOR THE TRANSCRIPTION ?



- **Both strands of the DNA molecule serve as template for whole genome transcription. But for each gene, only one and always the same is used as template.**

- **The template DNA strand is determined by the position and directional sequence of the promoter (small green arrow)**

In the figure above, the genes a, d and e are transcribed from the upper DNA strand whereas the genes b, c, f and g are transcribed from the under strand



TRANSCRIPTION PROCESS INCLUDE THREE STEPS :

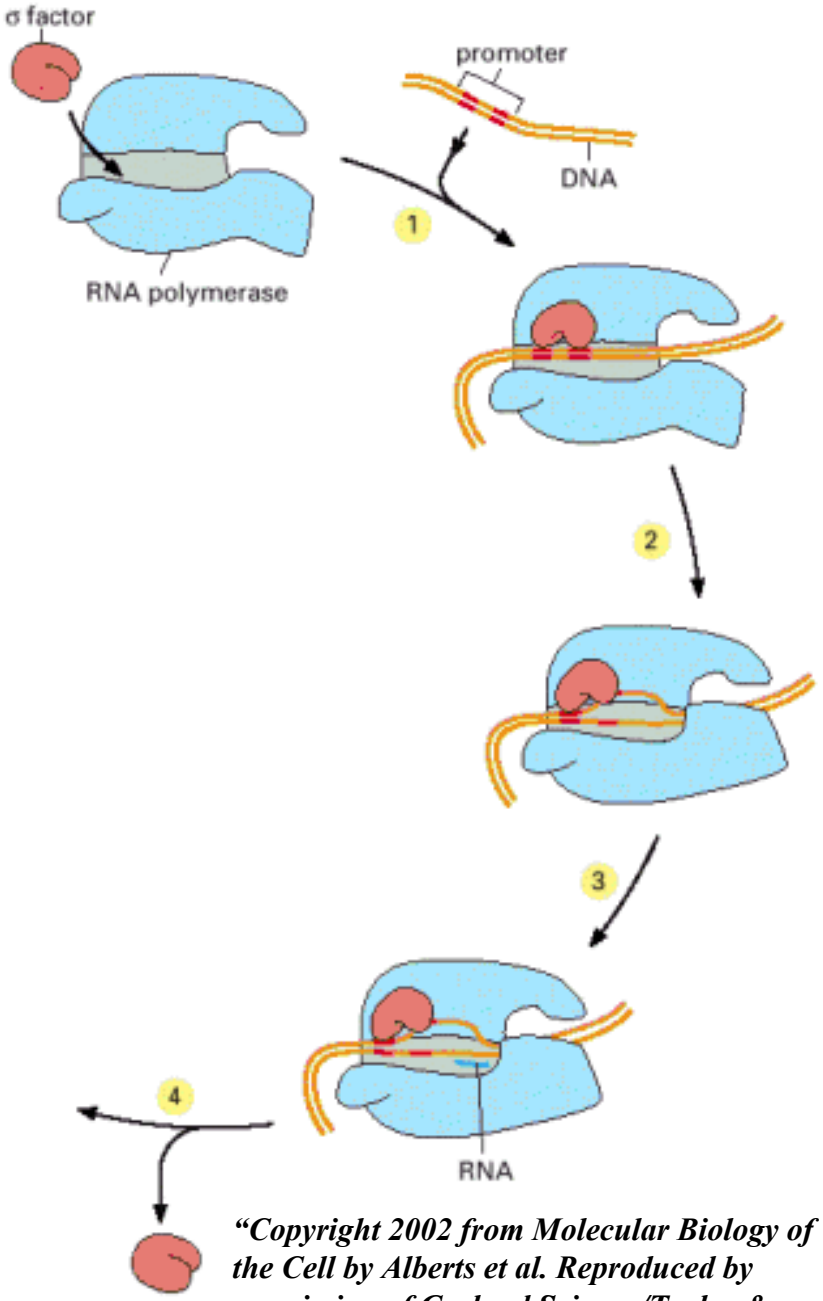
- **Initiation (1) → (4)**
- **Elongation (5), (6)**
- **Termination (7)**

TRANSCRIPTION INITIATION

Transcription initiation includes many "substeps":

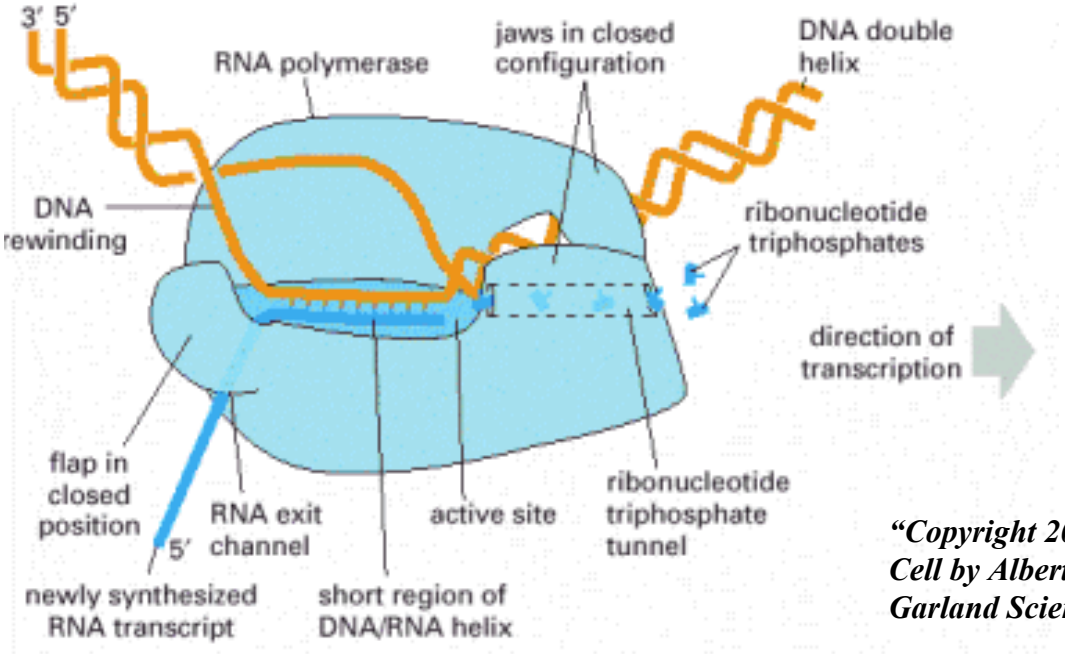
- **RNA polymerase binds to the promoter to form a closed complex ; DNA remains double-stranded (1)**
- **The closed complex becomes an open complex when the promoter region around the starting point is melted to form the transcription bubble (2)**
- **RNA polymerase begins to synthesize short ribonucleotide chains (9 – 10 nucleotides) many times. This period is called abortive initiation (3)**
- **When RNA polymerase succeeds in making a longer chain, it can be freed from the promoter. A stable ternary complex including the DNA template strand, RNA polymerase and the growing RNA chain is formed. The σ factor is released from the holoenzyme (4)**

Transcription initiation ends. Elongation begins



"Copyright 2002 from Molecular Biology of the Cell by Alberts et al. Reproduced by permission of Garland Science/Taylor & Francis LLC."

TRANSCRIPTION ELONGATION



- **The RNA polymerase moves along the DNA template, unwinding the double helix ahead and rewinding it behind the transcription bubble and synthesizing the RNA chain in the 5' – 3' direction.**
- **Only a fragment of 8 – 9 ribonucleotides of the newly synthesized RNA molecule base-pair with the DNA template at the transcription bubble.**
- **RNA polymerase has proofreading activities : (1) pyrophosphorolytic editing : removal of the newly incorporated wrong nucleotide, (2) hydrolytic editing : removal of the wrong nucleotide or a short sequence containing the error.**

TRANSCRIPTION TERMINATION

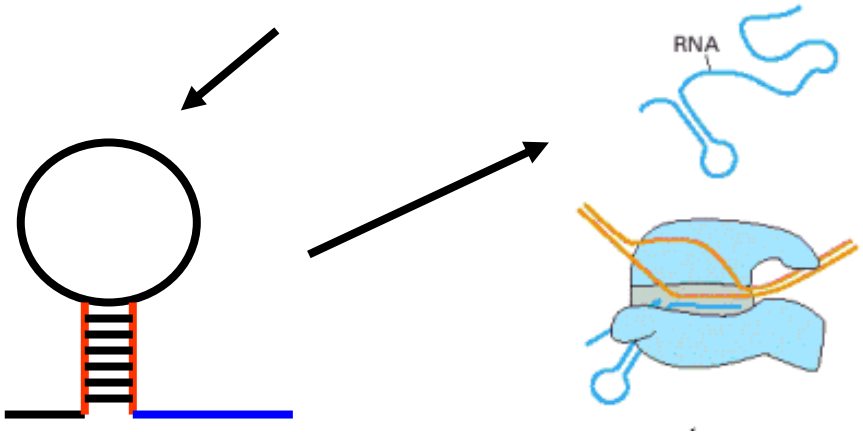
Transcription is terminated by two sequences of the RNA chain : Rho-independent terminator and rho-dependent terminator

DNA

CCCAAGCGCCGCTAATGACGGCGCTTTTTTTTTGAACAAAA
GGGTTCGCGGCGATTACTGCCGCGAAAAAAAAAACTTGTTTT

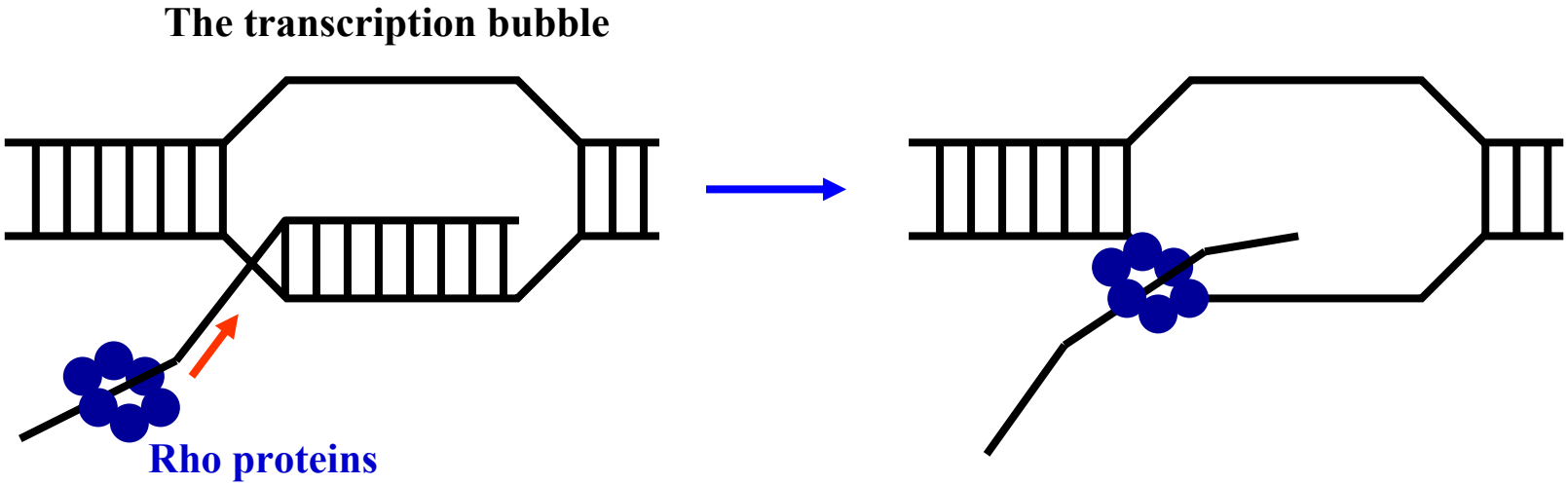
RNA

CCCAAGCGCCGCUAAUGACGGCGCUUUUUUUUUGAACAAAA



- **Rho-independent terminator is a "hairpin" structure including a stem constituted by the base-pairing of two inverted repeats followed by a sequence of 8-9 U**
- **When the termination hairpin structure is formed, it ends the transcription, maybe by weakening and then disrupting the interactions between the newly synthesized RNA and the DNA template**

TRANSCRIPTION TERMINATION (continued)



- **Rho–dependent terminator involves the formation of a ring constituted of six rho (ρ) protein molecules.**
- **Rho proteins have ATPase activities. When they reach the transcription machinery at the transcription bubble, they disrupt hydrogen bonds between the DNA template and RNA chain, so release the transcript.and terminate the transcription.**

TRANSCRIPTION IN EUKARYOTES

- Gene expression in prokaryotes is naturally "on"whereas in eukaryotes, when needed gene expression has to be "switch on" from the normal "off" status

- Transcription initiation in eukaryotes requires the initial participation of many general transcription factors which then recruit the RNA polymerase

- There are three RNA polymerases (RNA pol) responsible for the synthesis of different RNAs and proteins

ENZYMES	LOCATION	SYNTHESIZED PRODUCTS
RNA pol I	Nucleoli	Precursors of most rRNAs
RNA pol II	Nucleoplasm	mRNA precursors, some small nuclear RNAs (snRNA)
RNA pol III	Nucleoplasm	5S rRNA, tRNAs precursors, U6 snRNA and other small nuclear RNAs,

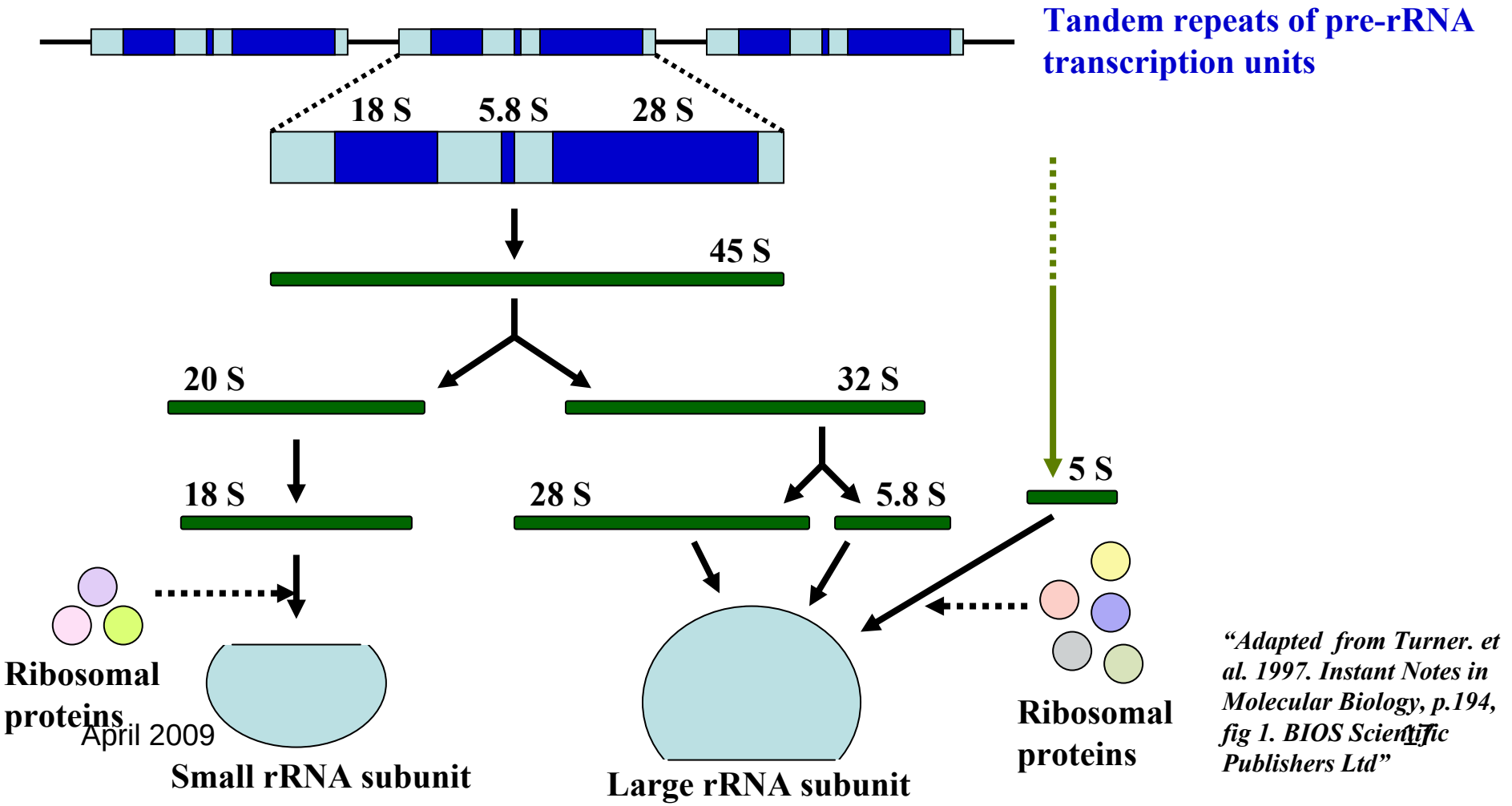
The three RNA polymerases are composed of many subunits, some of them are common for the three enzymes, e.g the TBP (TATA-Binding Protein)

Some of these subunits are homologous to *E. coli* RNA polymerase subunits, especially the two largest subunits of the three eukaryotic RNA polymerases are similar to the β and β' subunits in *E. coli*.

RNA POLYMERASE I - *TRANSCRIPTION INITIATION*

Products synthesized by RNA pol I : pre-rRNA transcripts containing 18S, 5.8S and 28S rRNA coding sequences.

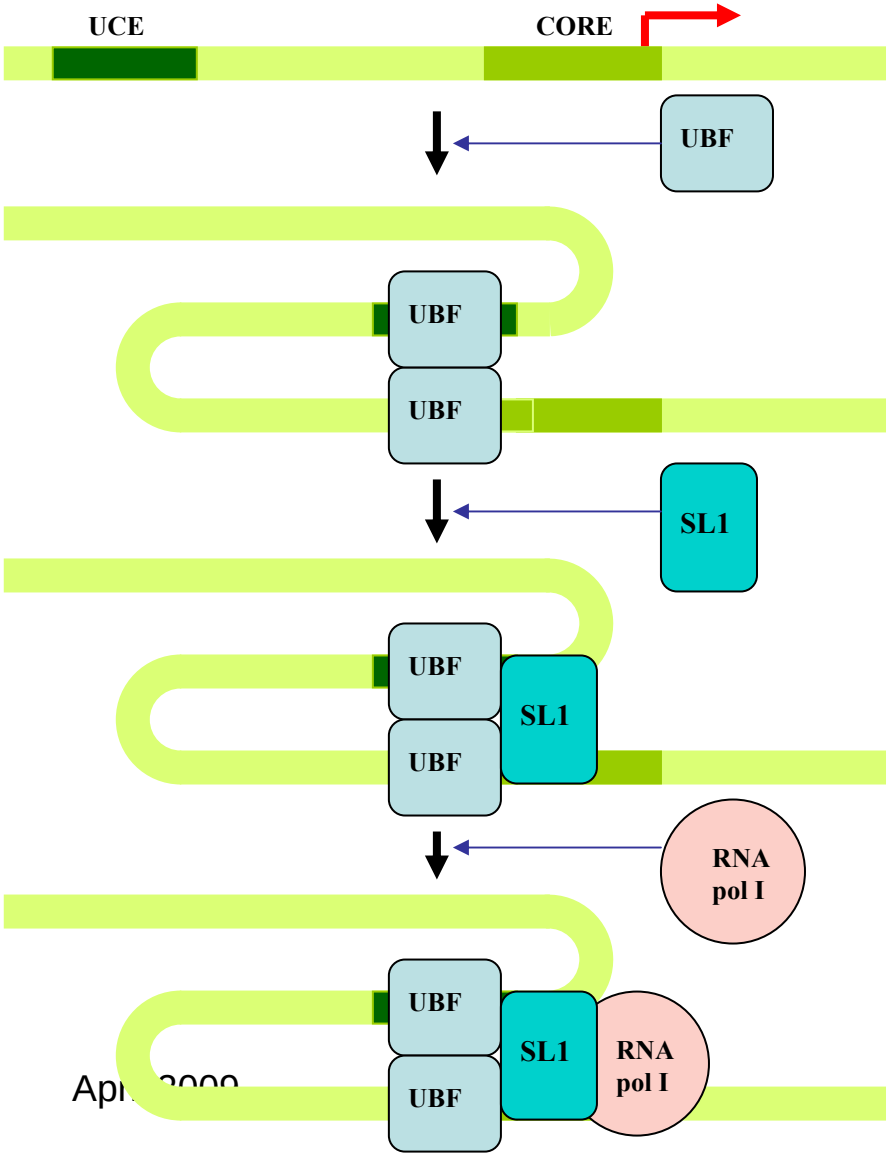
Pre-rRNA transcription units are regrouped in tandem repeats in some regions of the genome and form the nucleolus



"Adapted from Turner. et al. 1997. Instant Notes in Molecular Biology, p.194, fig 1. BIOS Scientific Publishers Ltd"

RNA POLYMERASE I (continued)

Promoters recognized by RNA pol I are composed of two sequences : UCE (Upstream Control Element) and the Core promoter



The recognition of the promoter include many steps :

➔ The proteins UBF recognize and bind to UCE sequence and the upstream part of the Core promoter. Binding of UBF proteins creates a loop on the promoter region and enhances transcription rate

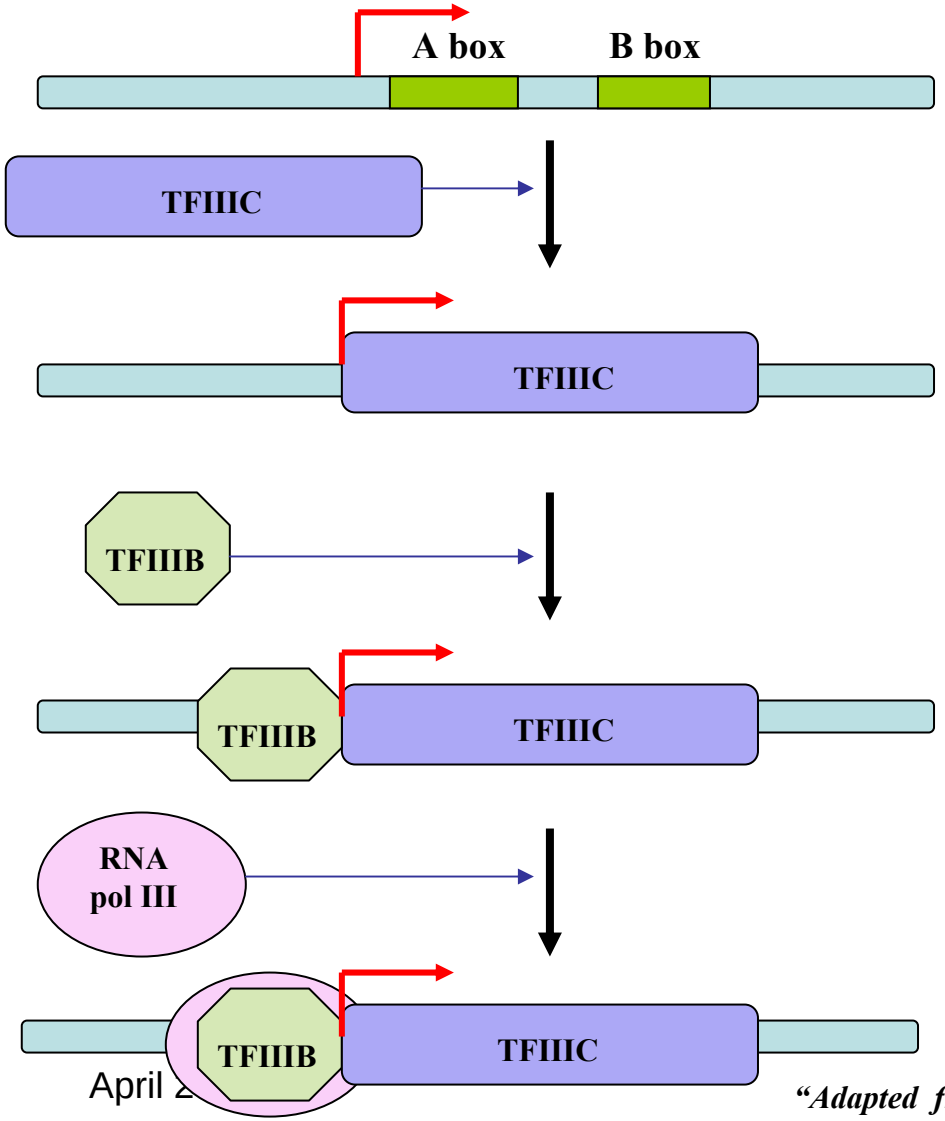
➔ SL1 complex, including TBP (TATA-Binding Protein), binds to the downstream part of the Core promoter, stabilizes UBF-UCE interaction and recruits RNA polymerase I

➔ RNA polymerase I binds and initiates the transcription

"Adapted from Turner. et al. 1997. Instant Notes in Molecular Biology, p.195, fig 3. BIOS Scientific Publishers Ltd"

RNA POLYMERASE III – *TRANSCRIPTION INITIATION*

- Products synthesized by **RNA polymerase III** : precursors of **5S rRNA, tRNAs**, and **small nuclear RNAs (snRNAs)**

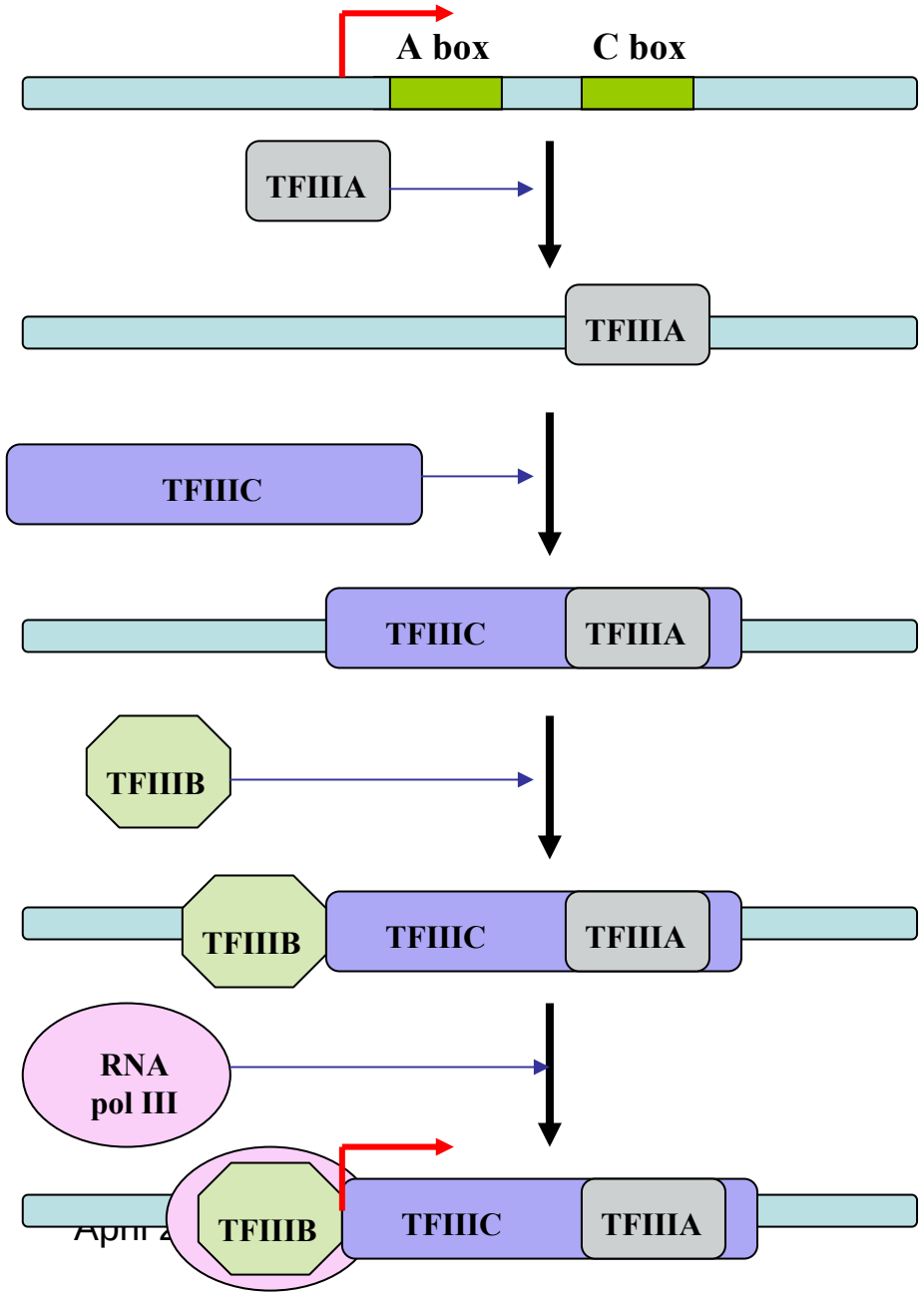


- Transcription control region of **tRNA genes** is located inside the coding region and is composed of two conserved sequences A box (5'TGGCNNAGTGG3') and B box (5'GGTTCGANNCC3').
- A box and B box encode the conserved D-loop and the TψC loop of tRNAs.
- Transcription initiation begins when TFIIIC binds to A box and B box ; this binding determines the binding site of TFIIIB which contains TBP (TATA-Binding Protein).

TFIIIB then recruits RNA polymerase III which initiates transcription

"Adapted from Turner. et al. 1997. Instant Notes in Molecular Biology, p.198, fig 1. BIOS Scientific Publishers Ltd"

RNA POLYMERASE III (continued)



- 5S rRNA genes exist in tandem repeats in the genome.
- The promoters of 5S rRNA genes contain two conserved sequences, te A box and the C box
- Transcription initiation occurs when a protein, TFIIIA, binds to the C box.

The order of intervention of the remaining factors is almost the same as for tRNA genes

"Adapted from Turner. et al. 1997. Instant Notes in Molecular Biology, p.199, fig 2. BIOS Scientific Publishers Ltd"

RNA POLYMERASE II : *TRANSCRIPTION INITIATION*

Products synthesized by RNA polymerase II are mRNAs

DNA sequences controlling transcription initiation by RNA pol II include :

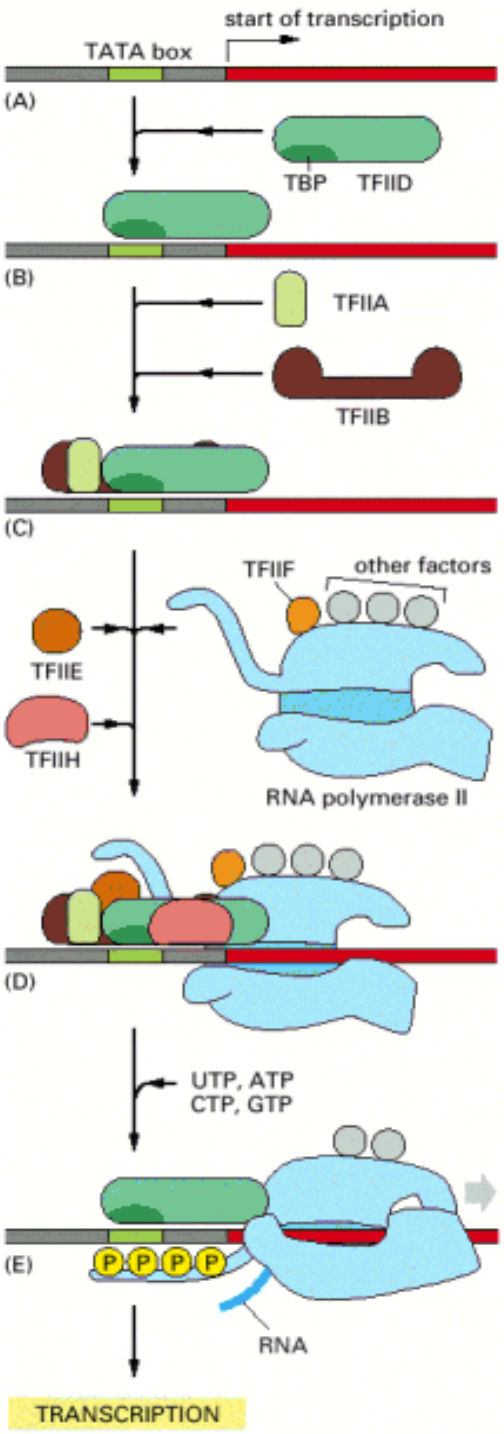
- **The Core promoter composed of some of the following sequences : (1) BRE (TFIIB Recognition Element), (2) TATA box, (3) Initiator, (4) DPE (Downstream Promoter Element).**

TATA box, common to nearly all promoters, is situated around 25-35 bp upstream of the start site. It is the equivalent of -10 sequence in prokaryotic promoter and has the consensus sequence 5' TATA(A/T)A(A/T) 3'.

The basal transcription, especially in *in vivo* conditions, is largely enhanced or repressed. These enhanced and repressed effects involve the interaction between regulatory proteins and :

- **Other regulatory sequences (mediator, enhancer, silencer, …)**

RNA POLYMERASE II (continued)



The formation of pre-initiation complex on the promoter include many steps :

TFIID - more specifically its TBP (TATA Binding Protein) subunit – binds to the TATA box, followed by the binding of : TFIIA, TFIIB, TFIIF-RNA polymerase complex, TFIIE/H.

TFIIH induces promoter melting

→ Abortive initiation occurs

→ When RNA polymerase successfully synthesizes a RNA chain longer than 10 nucleotides → Initiation ends, elongation begins.

The elongation step is controlled by elongation factors such as hSPT5, TFIIS. This step involves the phosphorylation of the CTD (C-terminal Domain) constituted of many repeats "Tyr-Ser-Pro-Thre-Ser-Pro-Ser" by TFIIH and other kinases involving hSPT5. TFIIS has a role in proofreading during elongation. It stimulates an RNase activity center of RNA polymerase which removes misincorporated nucleotides.

RNA POLYMERASE II – *BASAL TRANSCRIPTION FACTORS AND OTHER REGULATORY PROTEINS*

WHAT ARE THE ROLES OF OTHER REGULATORY PROTEINS IN TRANSCRIPTION INITIATION ?

Besides the general transcription factors, other regulatory proteins also participate in transcription initiation. The reason is :

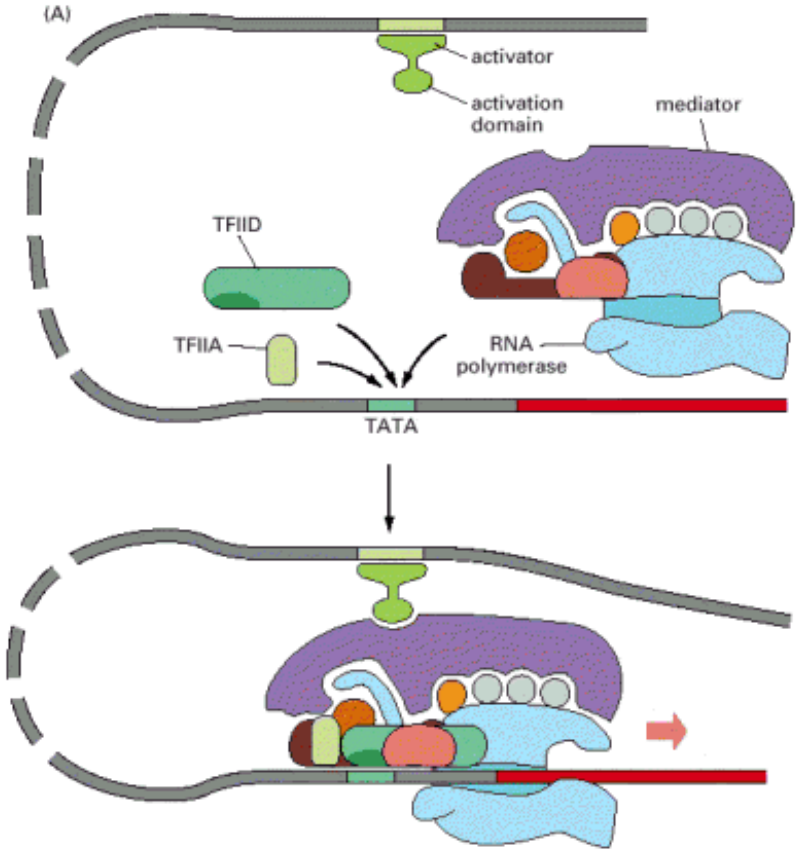
(1) In eukaryotes, genomic DNA is packaged into nucleosomes and more complex structure and needs additional factors to make promoter region more accessible to initiators

(2) Multicellular eukaryotes require precise regulation of gene expression which is time and tissue specific.

These other regulatory proteins include : Mediator complex, special transcriptional regulatory proteins, chromatin modifiers, …

The special transcriptional regulatory proteins and the chromatin modifiers are detailed in "*Regulation of gene expression in eukaryotes*"

THE ROLE OF MEDIATOR IN TRANSCRIPTION



"Copyright 2002 from Molecular Biology of the Cell by Alberts et al. Reproduced by permission of Garland Science/Taylor & Francis LLC."

In the cell, transcription initiation needs additional factors. Mediator complex is one of them. The Mediator complex :

- **Is a protein complex composed of many subunits, some of these subunits are conserved between different species.**
- **Has some characteristics : its action is general for all RNA pol II promoters, its action is positive, its action is required at the same time of the general transcription factors.**
- **Associates with RNA polymerase II through one surface and interacts with other regulatory proteins at the other surface**
- **May have some functions : stabilizes and speeds the initiation complex formation, enhances the rate of transcription initiation through activation of the CTD kinase of TFIIH**

RNA POLYMERASE II – *ELONGATION FACTORS*

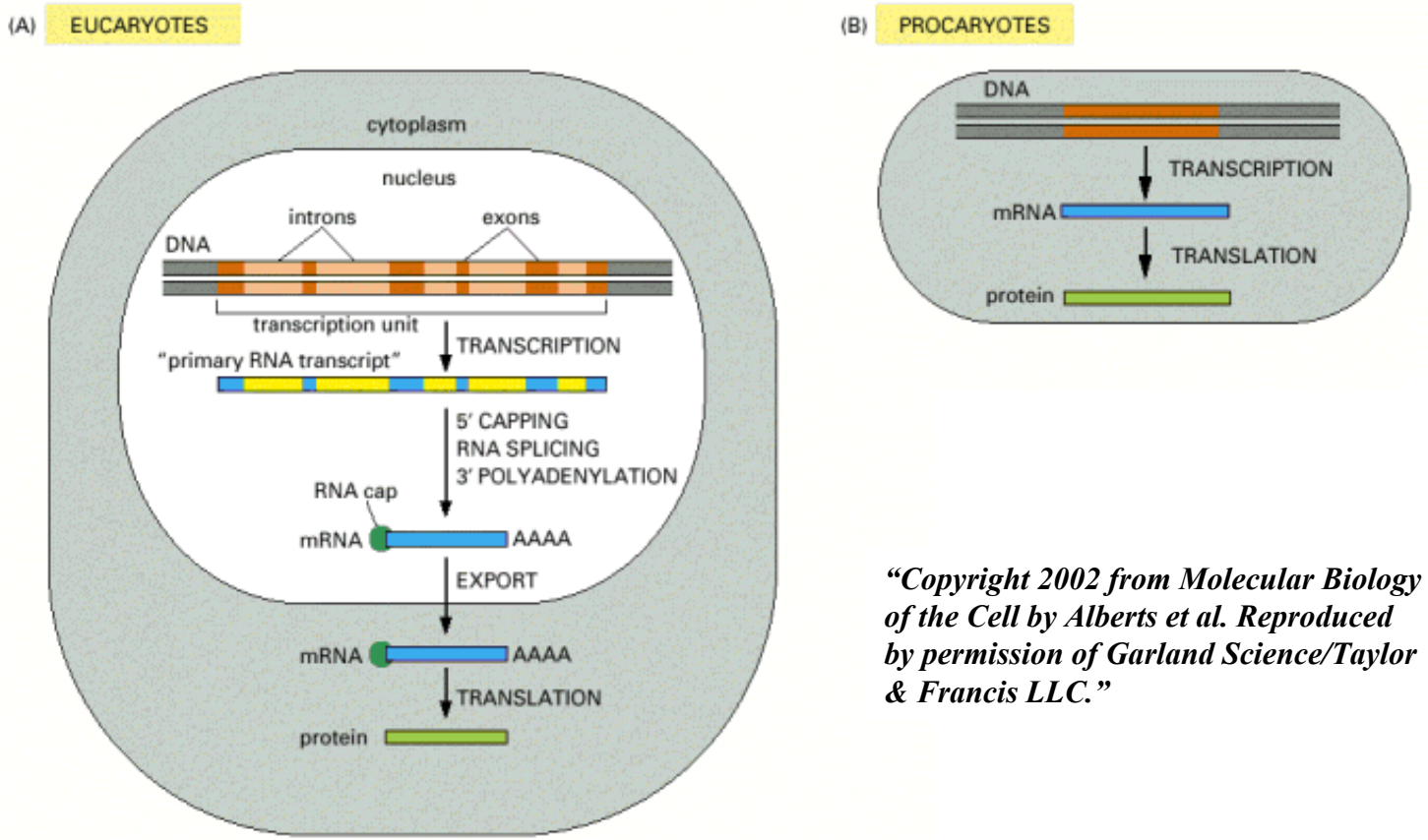
After initiating transcription, RNA polymerase enters the elongation stage. Elongation stage involves the replacement of initiation factors by elongation factors. This factor exchange necessary for elongation is induced by phosphorylation of the RNA polymerase CTD.

Elongation by RNA polymerase II is stimulated by some factors : kinase P-TEFb and TFIIS

- **Kinase P-TEFb acts by three ways :**
  1. **It phosphorylates, that is activates, the serine residue at position 2 of the CTD**
  2. **It phosphorylates the protein hSPT5 which is an elongation factor**
  3. **It recruits another elongation factor, TAT-SF1**
- **TFIIS has two functions : (1) it increases elongation rate, (2) it contributes to proofreading process by activating a RNase activity of the RNA polymerase which can remove misincorporated ribonucleotides.**

- **Besides the binding of elongation factors, the elongating polymerase also binds other protein factors involved in mRNA processing.**

mRNA PROCESSING IN EUKARYOTES

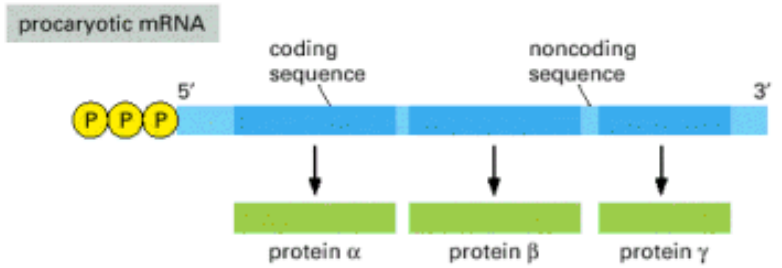


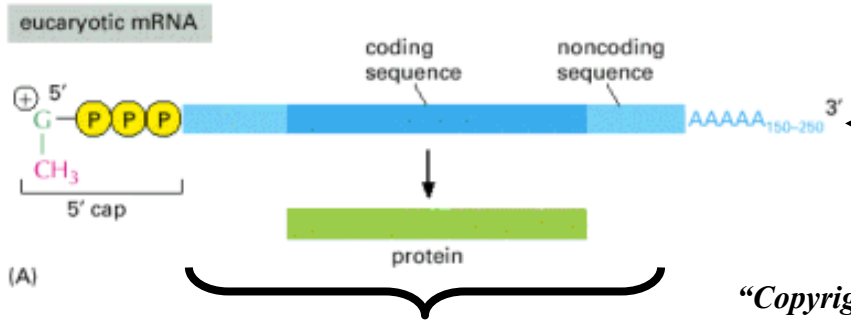
Since transcription and translation process are closely linked in prokaryotes, mRNAs are almost translated without any processing steps.

In eukaryotes, the "primary RNA transcripts" (pre-mRNAs) are processed before exporting from the nucleus to the cytoplasm to be translated.

mRNA PROCESSING

mRNA processing includes : (1) 5' capping, (2) RNA splicing, (3) 3' polyadenylation





5' capping is the first event

3' polyadenylation is the last event

RNA splicing is the second event

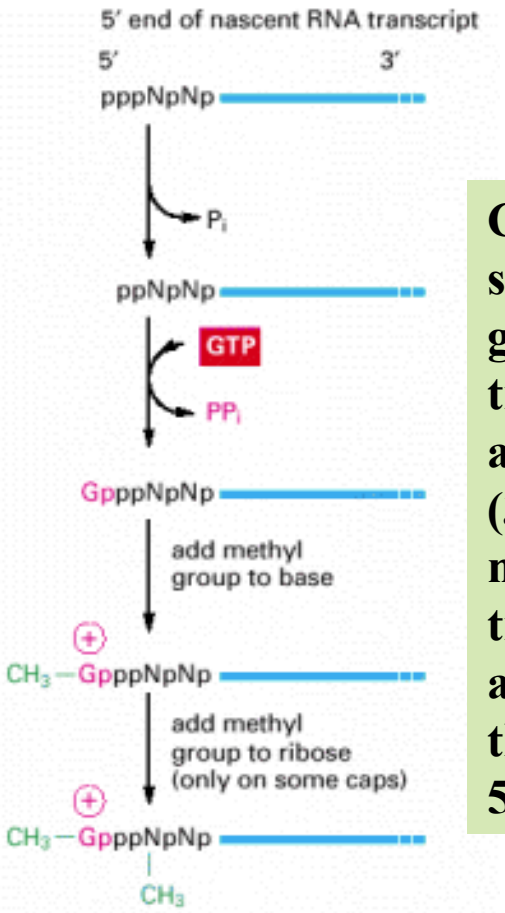
"Copyright 2002 from Molecular Biology of the Cell by Alberts et al. Reproduced by permission of Garland Science/Taylor & Francis LLC."

The enzymes responsible for RNA processing are recruited to the polymerase CTD during elongation stage.

Depending on the phosphorylation sites of the CTD, different enzymes are recruited, e.g phosphorylation of the serine at position 5 stimulates the recruitment of capping factors whereas phosphorylated serine at position 2 is associated with splicing factors recruitment. These recruited factors will induce structural changes of the mRNA

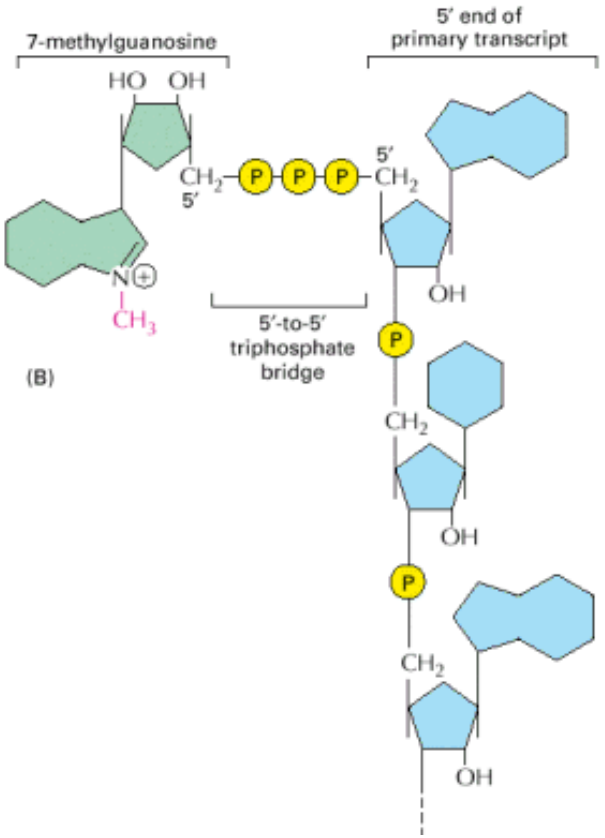
5’ CAPPING

Capping is the adding of a methylated guanine to the 5’end of an mRNA. Capping begins when RNA is only 20-40 nucleotides long



“Copyright 2002 from Molecular Biology of the Cell by Alberts et al. Reproduced by permission of Garland Science/Taylor & Francis LLC.”

Capping reaction involves three steps : (1) The 5’phosphate group is removed by RNA triphosphatase, (2) a GTP is added by guanylyl transferase, (3) The added G is then methylated by methyl transferase ; methylation can also occurs at the 2’OH group of the three riboses located at the 5’end of the mRNA



“Copyright 2002 from Molecular Biology of the Cell by Alberts et al. Reproduced by permission of Garland Science/Taylor & Francis LLC.”

Functions of 5’capping : (1) Regulation of nuclear transport, (2) Protection of the mRNA 5’end from exonuclease digestion, (3) Regulation of translation, (4) Regulation of 5’ intron excision.

RNA SPLICING

BIOLOGICAL FUNCTIONS OF RNA SPLICING ?

Eukaryotic genes are interrupted genes made of exons (coding region) alternating with introns (non-coding region) ; coding sequences make up less than 10% of the total genome length.

The RNA processing event consisting of eliminating introns and joining exons together is called RNA splicing

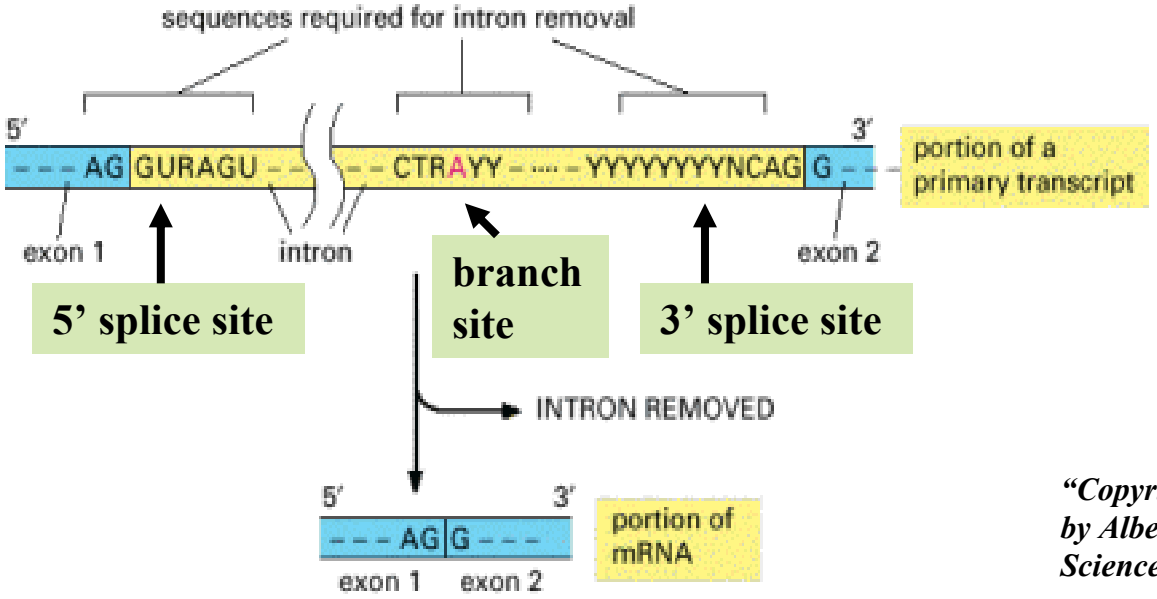
RNA splicing has two functions :

- **The translational machinery is not capable of distinguishing exon from intron, thus the introns have to be removed before the mRNA is translated**
- **Alternative splicing can generates many mRNAs encoding different proteins from one gene → increasing diversification**

The "machinery" responsible of RNA splicing is called spliceosome, a large complex composed of about 150 proteins and 5 RNAs called U1, U2, U4, U5 and U6. These RNAs combine with proteins to form the snRNPs (small nuclear ribonuclear proteins)

It seems that biological activities of the spliceosome are carried out by the RNA components rather than by proteins

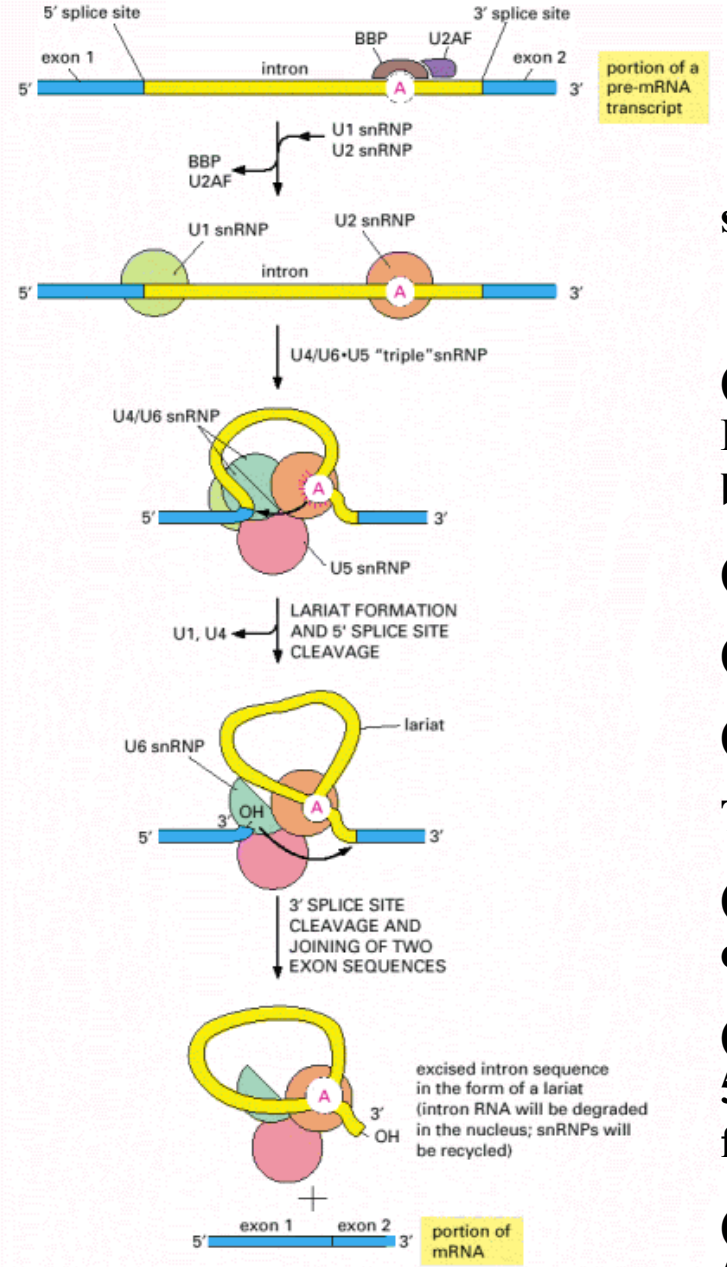
SPLICING PROCESS



- **Sequences on the pre-mRNA which determines the splicing sites include : a 5’ splice site and a 3’ splice site which determine exon-intron boundary and a branch point site.**

The most conserved sequences are the GU at the 5’splice site, A at the branch point site and AG at the 3’ splice site.

- **Splicing includes two successive transesterification reactions : (1) firstly, the 2’OH of A at the branch site, as a nucleophile, attacks the phosphoryl group of the G at the 5’splice site, this results in the break of phosphodiester bond at the junction of exon 1 and the intron, the intron forms a loop at its 5’end, (2) secondly, the 3’OH of exon 1 acts as a nucleophile to attack the phosphoryl group at the junction of the intron and exon 2, this will link together exon 1 and exon 2 and liberates the intron in the form of a lariat**



SPLICING PROCESS (continued)

The splicing process includes two stages : assembly of the spliceosome and reaction catalysis

The assembly stage :

(1) A helper protein, U2AF helps BBP (Branch Point Binding Protein) to bind to the branch site ; U1 snRNP binds to the 5'splice site

(2) U2 snRNP binds the branch site and displaces BBP

(3) U4/U6 and U5 snRNPS join the complex

(4) U1 is replaced by U6.

The catalysis of esterification reactions :

(1) U4 is ejected from the complex, allowing the interaction of U6 and U2 which forms the active site

(2) The first esterification consists of interaction between the 5'end of the intron at the 5'splice site and the branch site to form a "lariat"

(3) In the second reaction, the 3'end of the exon 1 at the 5'splice site and the 5'end of the exon 2 at the 3'splice site are joined together.

HOW CAN SPLICING BE AN EXTREMELY PRECISE MECHANISM ?

Eukaryotic genes are made of alternating exons and introns. The most challenging problem for the cell is to accurately recognize the exon-intron boundaries and to perform correct splicing process.

➔The very complicated splicing pathway contributes to prevent inappropriate splicing. For example, the active site is only formed after many steps of the spliceosome assembly when all the splice sites are duly recognized.

The two kind of splicing error and prevention mechanisms :

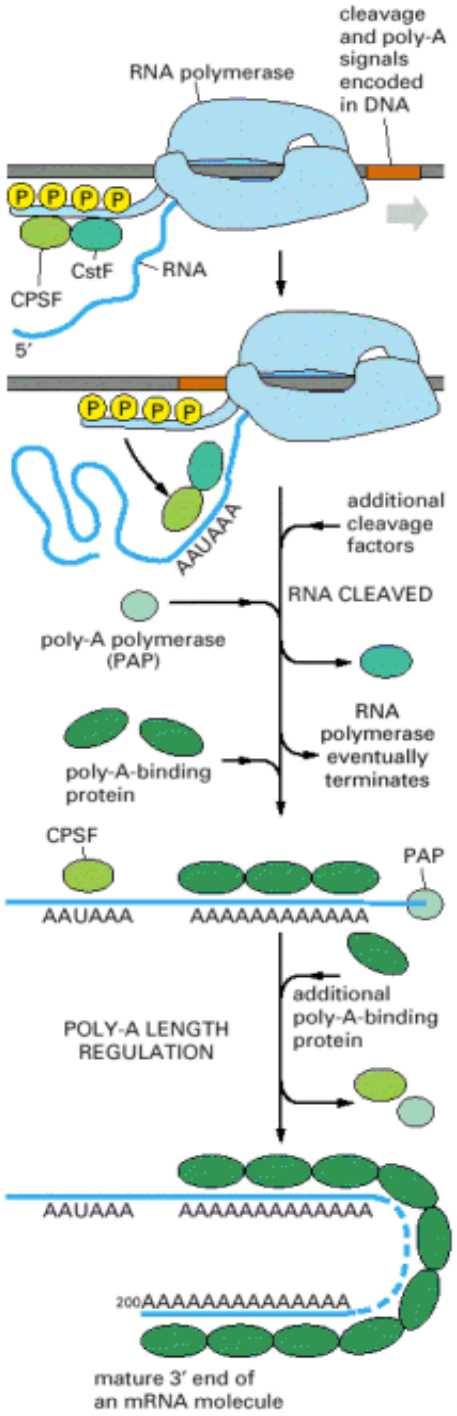
(1) skipping of splice sites, e.g splicing occurs between the 5'splice site of exon 1-intron 1 and the 3'splice site of intron 2-exon 3 instead of intron 1-exon 2.

➔ During transcription, the CTD tail of RNA polymerase is associated with the splicing proteins. Thus, the 5'splice site sequence newly synthesized interacts with the successive 3'splice site synthesized just after. This combined transcription-RNA processing prevents splice site skipping.

(2) misrecognition of other sites resembling splice sites.

➔ A protein called SR (Serine Arginine rich) binds to a sequence called ESE (Exonic Splicing Enhancer) located within the exons. Bound SR enhances the recruitment of the splicing machinery to the correct splice sites. This prevents the incorrect sites not close to the exons to be used in splicing reaction. Remind that introns are much more larger than exons.

POLYADENYLATION



Polyadenylation is the last RNA processing event, and is needed for the transcription termination. This is the formation of a poly-A tail at the 3'end of the pre-mRNA.

Polyadenylation concerns the cleavage of the pre-mRNA followed by the addition of many A residues to its 3'end. Reactions occur as follows :

- **CPSF (Cleavage and Polyadenylation Specificity Factor) binds to the AAUAAA and CstF (Cleavage Stimulation Factor) binds to a GU-rich region of the poly-A signal sequence in the pre-mRNA. These factors cleave the pre-mRNA.**
- **The binding of CPSF and CstF also recruit other proteins, such as the PAP (poly-A polymerase) which then adds many A residues (about 200 A) to the cleaved site.**

The mature RNA is formed and transcription ends.

SUMMARY

Transcription, the synthesis of RNA from DNA template, involves two elements : the promoter, situated upstream of the coding sequence, decides for the template strand selection and the beginning of RNA synthesis, and transcription proteins involving RNA polymerases and transcription factors.

Transcription in prokaryotes and eukaryotes have some differences.

- **One RNA polymerase transcribes all the prokaryotic genes. In eukaryotes, there are three RNA polymerases – RNA pol I is responsible for pre-rRNA transcription, RNA pol II produces mRNA and RNA pol III transcribes the pre-5S rRNA, tRNA, and snRNA genes.**

- **Prokaryotic promoter is essentially the binding site for RNA polymerase whereas promoters in eukaryotes are composed of binding sites for many transcription factors as well as RNA polymerases.**

The transcription process includes three steps : initiation, elongation and termination

- **The initiation of transcription in prokaryotes begins when RNA polymerase binds to the promoter. In eukaryotes, the initiation requires binding of te general (basal) transcription factors prior to the recruitment of RNA polymerases.**

- **The mechanism of elongation is nearly the same for all RNA polymerases. RNA polymerases move along the DNA template in the 5'-3' direction, unwinding the double helix ahead and rewinding it behind the transcription bubble.**

SUMMARY (continued)

ℭ Transcription in prokaryotes is terminated when RNA polymerase encounters one of the two termination signals : ρ-dependent and ρ-independent termination structures. In eukaryotes, after RNA polymerases terminate transcription, the transcripts are processed before being used in translation. The most important RNA processing concerns pre-mRNA. The pre-mRNAs undergo three processing pathways : 5' capping, RNA splicing and polyadenylation. The processed products, mature mRNAs, are transported to the cytoplasm to the translational machinery.

Another type of RNA modification is RNA editing which is discussed in "Regulation of gene expression in eukaryotes"