Transcription

Praveen Deepak
Assistant Professor of Zoology
Swami Sahjanand College
Jehanabad

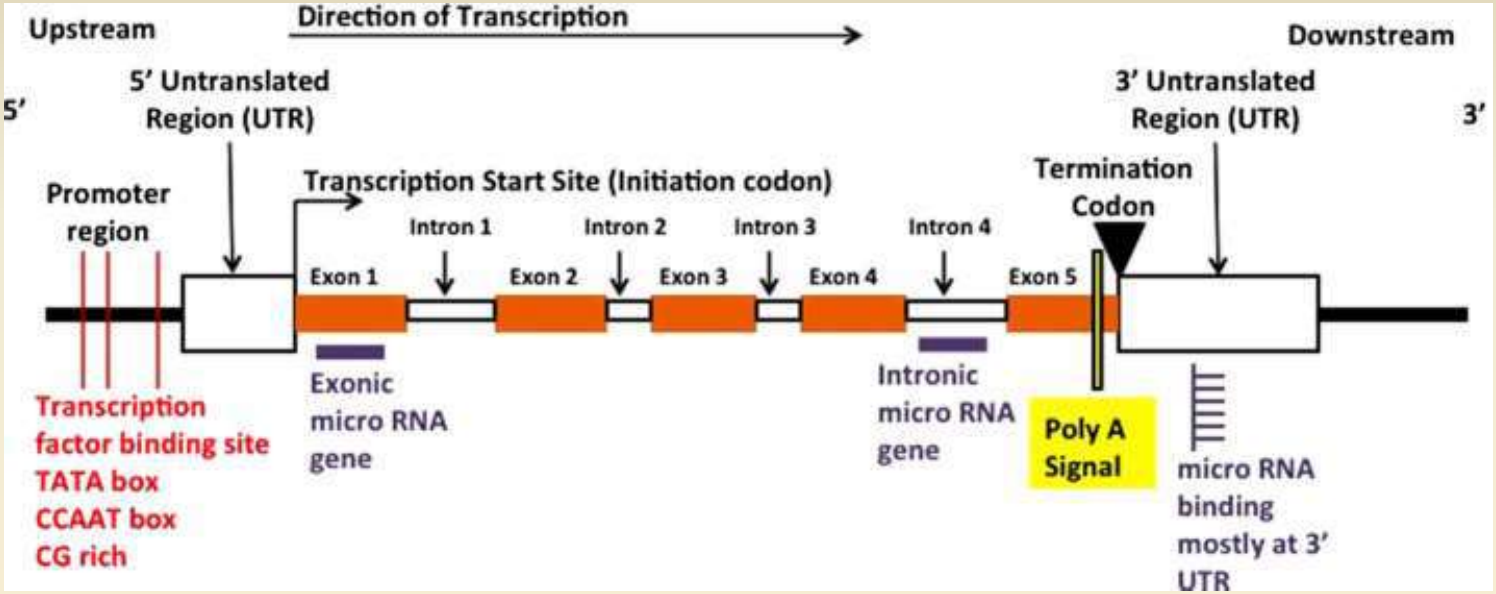
Transcription

Contents

Introduction

- It is the process of synthesis of messenger RNA transcripts by copying the template strand of DNA in the prokaryotes. Messenger RNA is then translated to produce proteins (some are not translated e.g. ncRNA or non-coding RNA).
- It occurs in the protoplasm/cytoplasm alongside translation (In eukaryotes, translation can not occur simultaneously with the transcription).
- During replication entire genome is copied but in transcription only the selected portion of genome , i.e. called as gene, is copied.
- The enzyme involved in transcription is RNA polymerase. Unlike DNA polymerase it can initiate transcription by itself, it does not require primase (More exactly it is a DNA dependent RNA polymerase).
- Like DNA replication, it is also a multistep process involving several enzymes and a number of regulatory molecules, called transcription factor (some factors are homodimers containing helix-turn-helix DNA-binding motifs).
- There are three steps in transcription:
    - Initiation – Formation of Transcription Initiation Complex (TIC)
    - Elongation – Formation of mRNA complementary to the DNA template
    - Termination – Termination of transcription

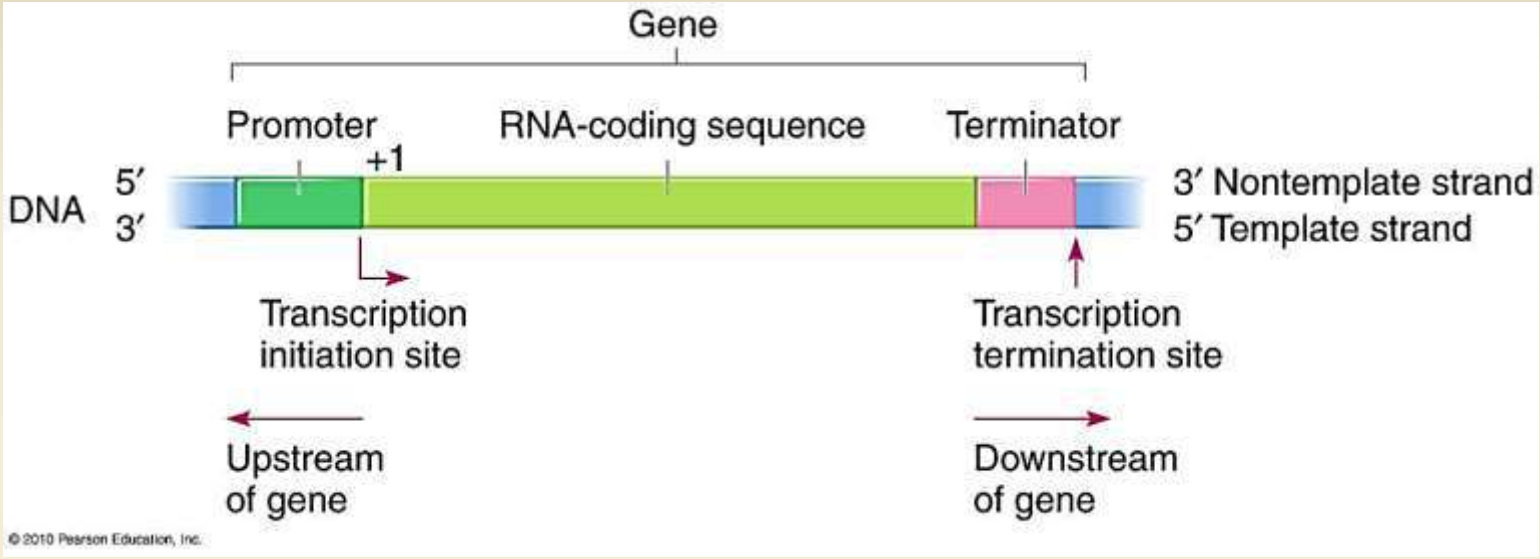
Structure of Gene



- Genes tend to be arranged in cluster along the chromosome (a typical chromosome may be upwards 1.5 x 10^8 bp in length and may contain between 2 – 3 x 10^3 genes)
- A gene cluster may contain upwards of 1.5 x 10^6 bp.
- Each gene is a unique DNA sequence that consists of a 5' regulatory region, e.g. promoter region, *untranslated region (UTR)*, *termination codon region*, exons (red) and introns (white boxes). An average gene may be comprised of 2 – 5 x 10^4 bp.
- The coding region of eukaryotic genes typically consists of a series of ***ex***pressed **exons** with ***int***ervening **introns**.

https://doi.org/10.1016/B978-0-12-812537-3.00004-4

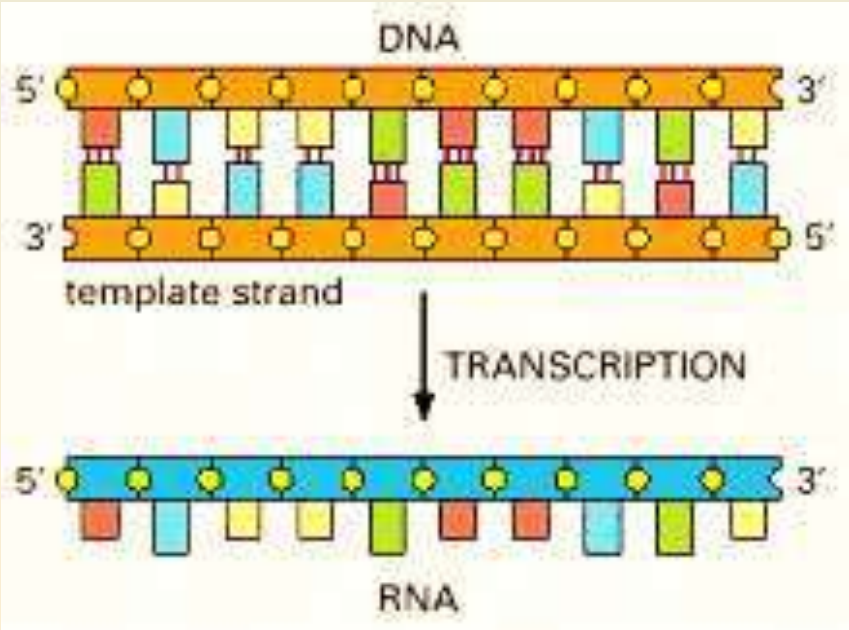
Structure of gene



- A "***gene***" occurs over a particular physical region called locus of dsDNA molecule.
- It includes an RNA-coding region, where transcription occurs from the DNA template strand .
- It runs between the initiation and termination sites in **5‘** to **3'** direction.
- Control regions for transcription, i.e., promoters and terminators occur immediately outside the coding sequence.
- By convention, the gene is written "*left to right*" in the **5'** to **3'** direction of the sense strand.
- The region to the '*left*' of the gene is called '***upstream***' and the region to the '*right*' is '***downstream***'.

DNA to RNA – Transcription

DNA transcription produces a single single-stranded RNA molecule that is complementary to one strand of DNA

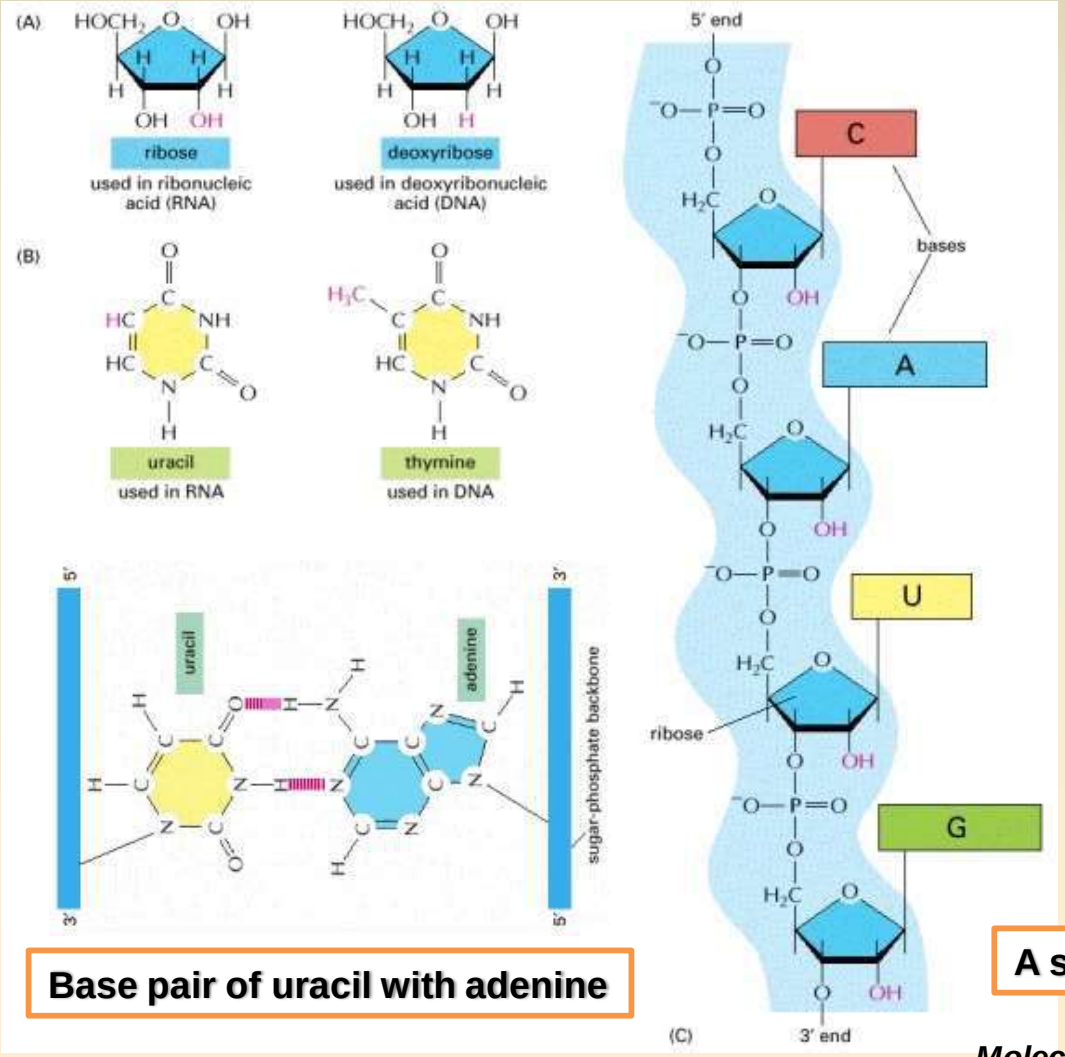


Poly-deoxyribonucleotide

Poly-ribonucleotide

Molecular Biology of the Cell. 4th edition.

Structure of RNA

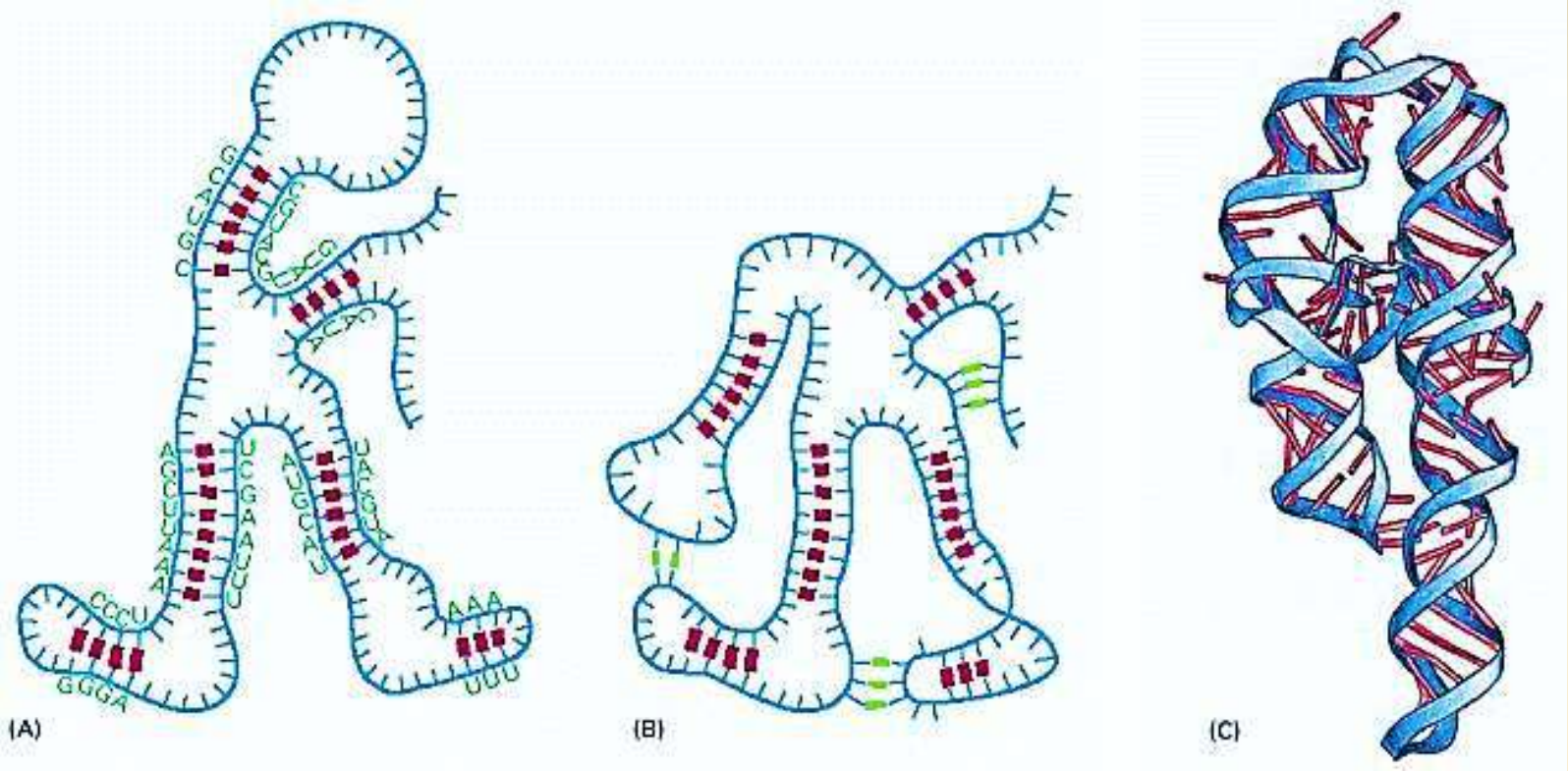


- ***RNA contains the sugar ribose, which differs from deoxyribose, (sugar used in DNA), by the presence of an additional -OH group.***
- ***RNA contains the base uracil, which differs from thymine, the equivalent base in DNA, by the absence of a - CH_3 group.***
- ***The <u>phosphodiester chemical linkage</u> between nucleotides in RNA is the same as that in DNA.***

A short length of RNA

Molecular Biology of the Cell. 4th edition.

Structure of RNA

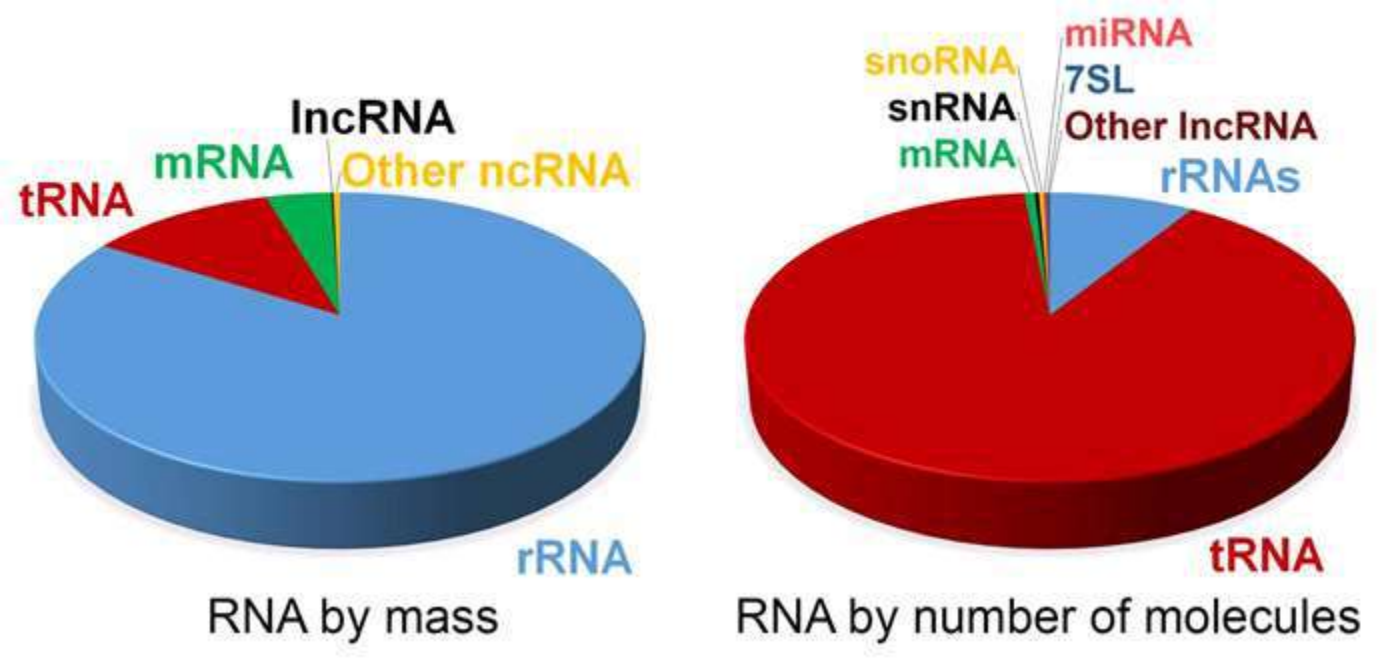


RNA can fold into specific structures

RNA is largely single-stranded, but it often contains short stretches of nucleotides that can form conventional base-pairs with complementary sequences and "nonconventional" base-pairs with modified bases on the same molecule that allow an RNA molecule to fold into a three-dimensional structure.

Molecular Biology of the Cell. 4th edition.

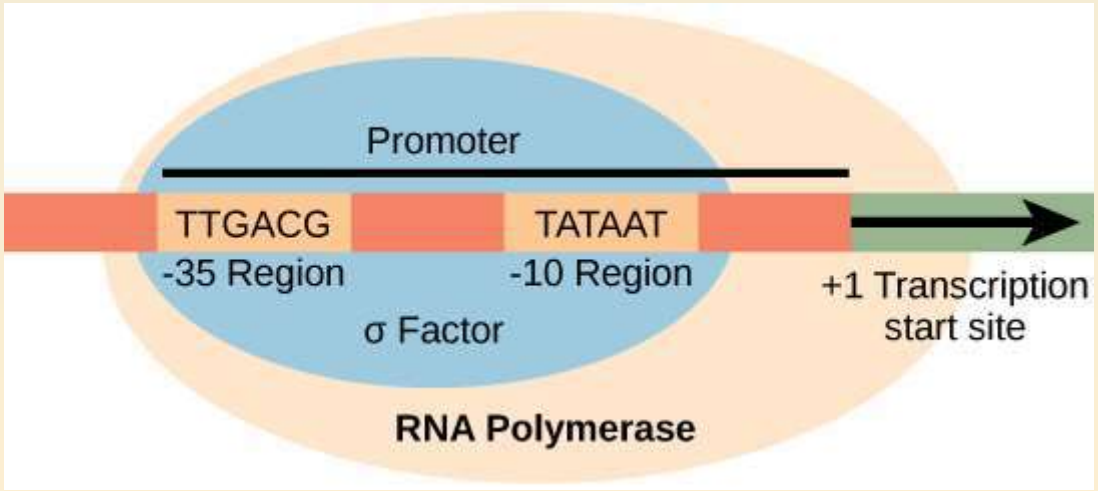
Types of RNAs



- Three main types of RNAs, viz. mRNA, rRNA and tRNA are found both in prokaryotes and eukaryotes.
- Transfer-messenger RNAs (tmRNA) are found only in bacteria.
- Small nuclear RNAs (snRNA) are found only in nucleated organism, such as eukaryotes.
- Some RNAs are involved in protein synthesis, some are in post-transcriptional modification, some are in replication and some are involved in regulation of gene expression and others.

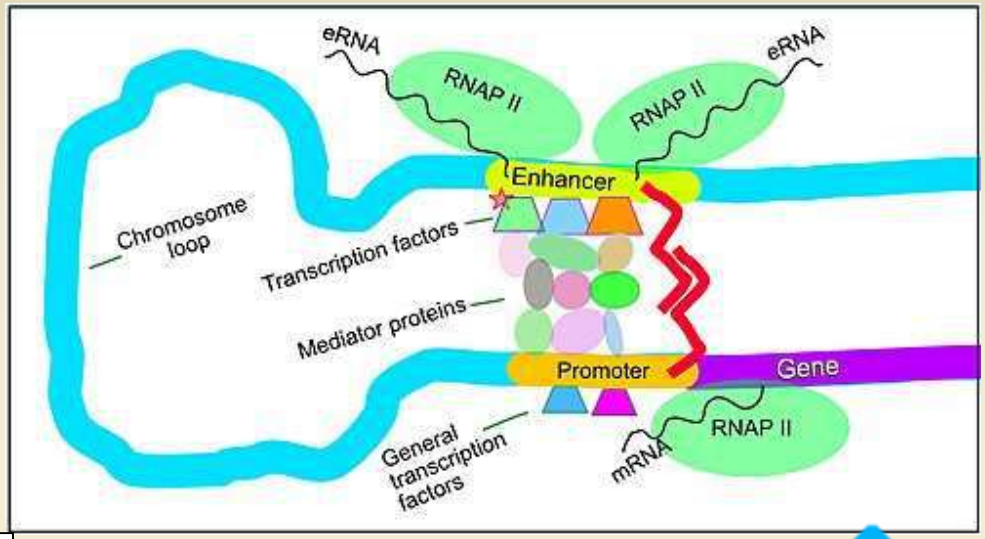
Initiation of Transcription

- Transcription in requires the DNA double helix to partially unwind in the region of RNA synthesis.
- It starts with the binding of proteins and enzymes at the ***promoter*** to form a transcription bubble, which is relatively simple in the prokaryotes, e.g. Pribnow box in prokaryotes and TATA box in eukaryotes which is present 10 bases before the transcription start site (-10). Most of the genes contain TATA-box sequence.
- TATA-box sequence is recognized and bound by TATA-box binding protein (TBP) subunit of transcription factor which helps position the RNA polymerase machinery and initiates transcription.

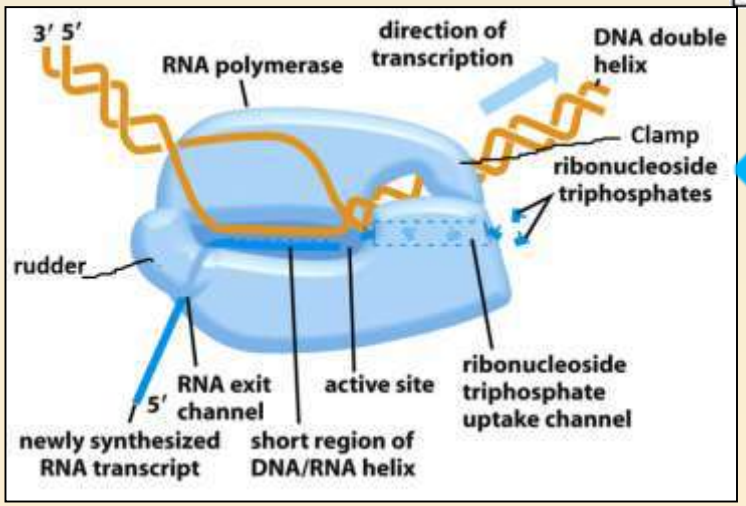


Initiation of Transcription

- The RNA polymerase recognize and binds to Pribnow box in prokaryotes, whereas a TATA box binding protein (TBP) recognize and bind to TATA box in eukaryotes).
- In prokaryotes, there are no transcription factor



Closing complex in eukaryotes



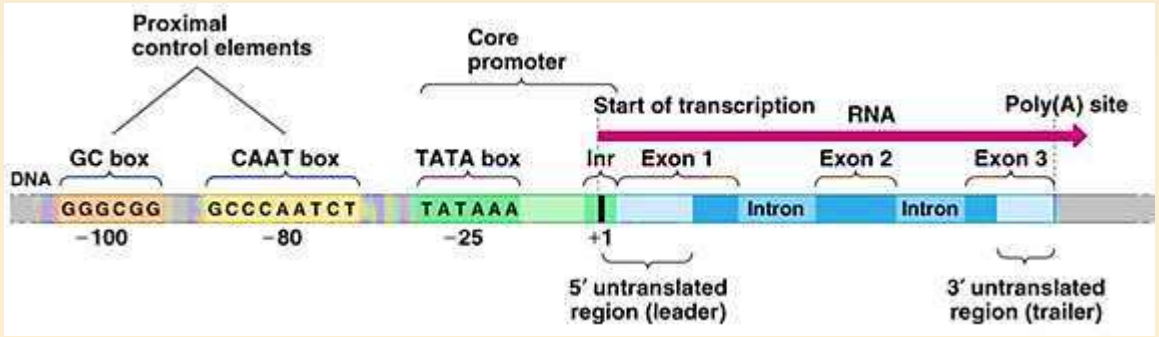
Closing complex in prokaryotes

- In eukaryotes, there are six general transcription factors namely TFIIA, TFIIB (ortholog of Archeal TFB), **TFIID** (a subunit ortholog of Archeal TPB), TFIIE (ortholog of archeal TFE), TFIIF and **TFIIH**.

Initiation of Transcription

Promoter region

- Region that contains specific DNA sequences recognized by a transcription factors.
- Transcription factor binds to promoter sequence and recruits RNA polymerase.
- It is of two types:
  - ***Core promoter:*** a minimal portion of the promoter required to properly initiate transcription. It has following characteristics:
    - It has Transcription Start Site (TSS) upstream to the gene
    - It has stretch of approximately -34 where RNA polymerase binds
    - It is a binding site for general transcription factor
  - ***Proximal promoter:*** is the proximal sequence upstream of the gene that tends to contain primary regulatory elements.
    - It may have up to approximately -250 bp of stretch
    - It is a binding site for specific transcription factor



Basal Promoter Element

Initiation of Transcription

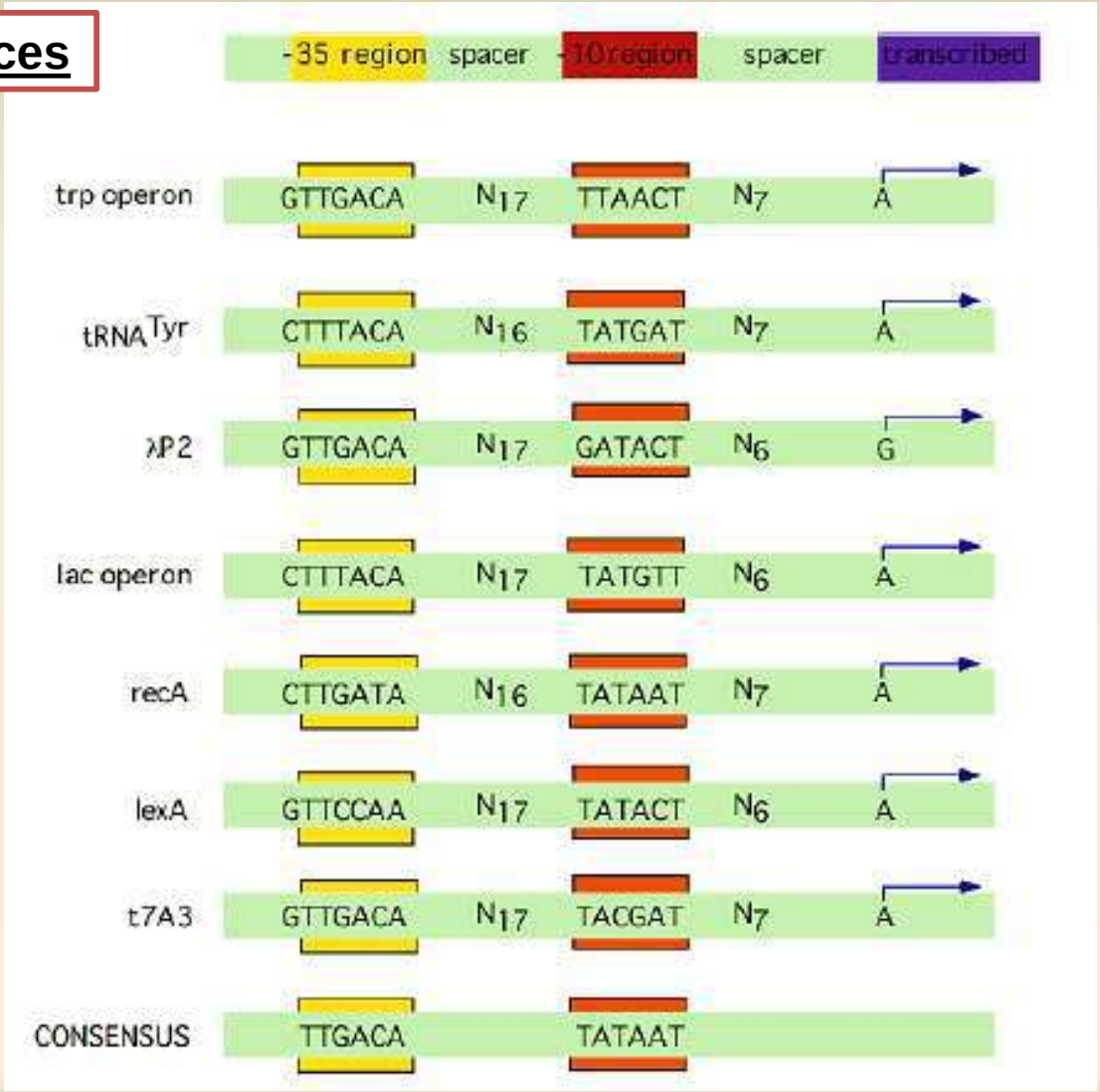
Promoter and Promoter Complexes

1. Promoter – DNA sequence that binds RNA polymerase to initiate transcription
2. Transcription initiation – Synthesis of first phosphodiester bond in nascent RNA
3. Position +1 – Position of nucleotide in DNA template that encodes the first nucleotide of mRNA.
4. Typical prokaryotic promoters recognized by E. coli σ70 – RNA polymerase share important polymerase recognition sequences.
    a. -10 region (Pribnow box): TATAAT consensus sequence
    b. -35 region: TTGACA consensus sequence
    c. Different promoters have similar, but not identical -10 and -35 region sequences
    d. Mutations within these regions alter promoter strength and function
    e. Distance between -10 and -35 regions important
    f. Strength of promoter mostly determined by affinity of RNA polymerase for promoter DNA sequences
    g. Region unwound by polymerase appears to be between -9 and +3 (includes right end of -10 sequence and extending to just downstream of transcription initiation site).
5. Synthesis of RNA in 5′ → 3′ direction; nucleotides added to 3′ end from ribonucleotide triphosphate precursors

Initiation of Transcription

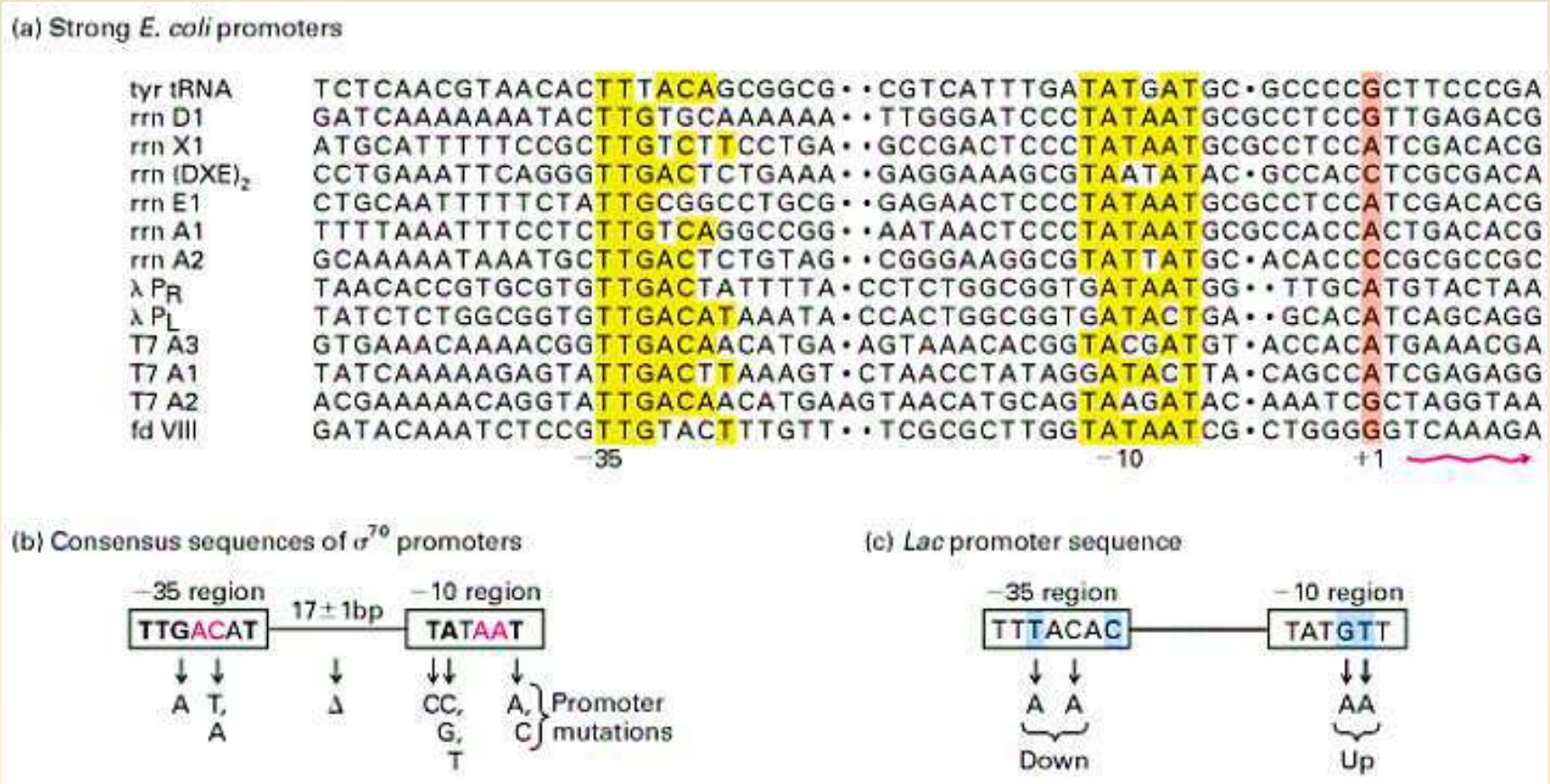
Some Promoter Sequences

These are prokaryotic promoters



Initiation of Transcription

Some Promoter Sequences



Initiation of Transcription

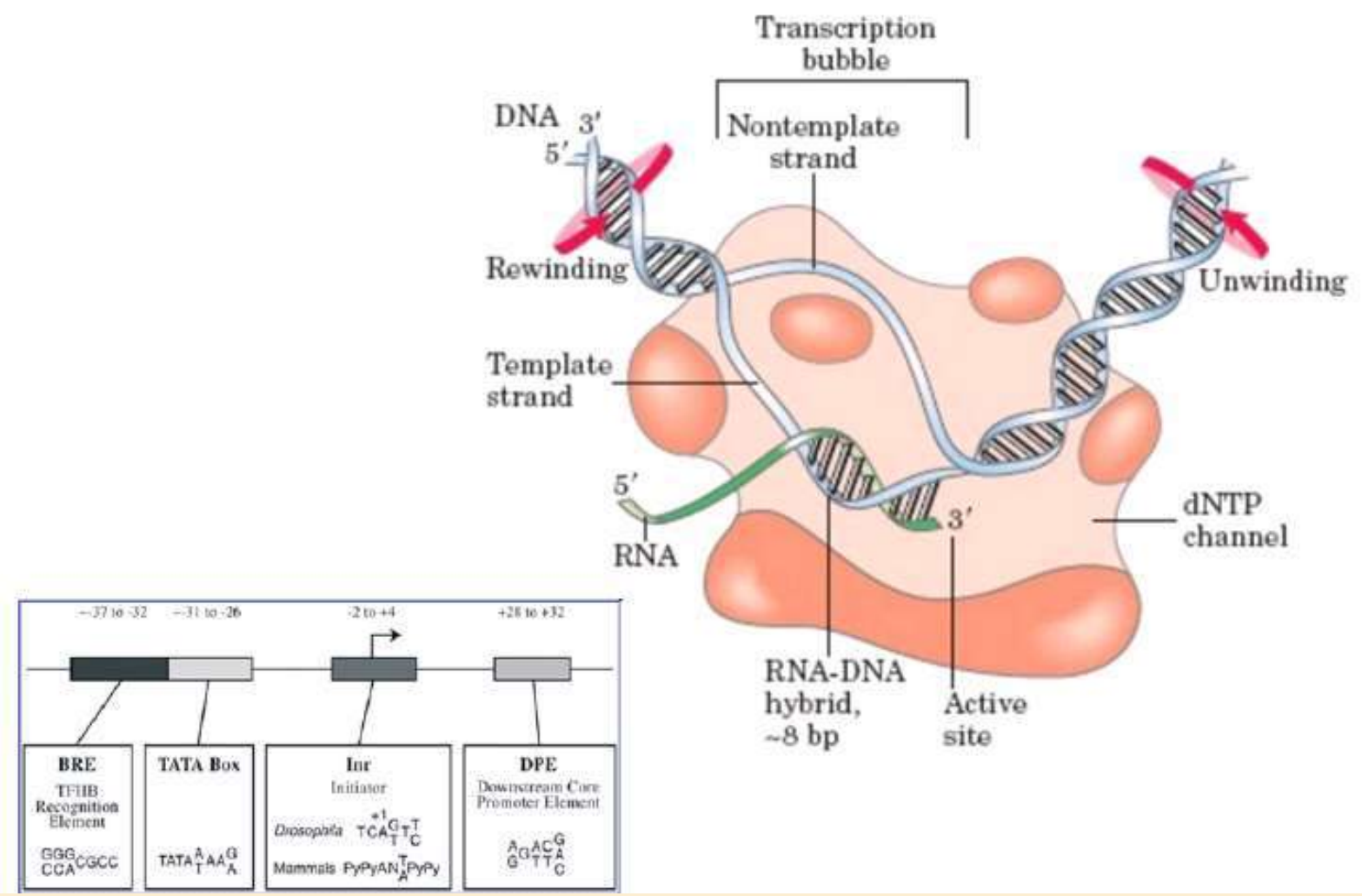
Difference between prokaryotic and eukaryotic promoters

- **Prokaryotic promoters:** In prokaryotic, the promoter consists of two short sequences at -10 and -35 positions from the transcription start sites (TSS).
  - → The sequence at -10 is called the Pribnow box, or the -10 element, which usually consists of six nucleotides TATAAT. It is essential to start transcription in prokaryotes.
  - → The sequence at -35 is called -35 element which usually consists of the six nucleotides TTGACA. Its presence allows a very high transcription rate.

- **Eukaryotic promoters:** In eukaryotic, the promoters are extremely diverse and are difficult to characterize. They lie upstream to gene and may have regulatory elements several kilobases away from the transcription start site. transcriptional complex can cause the DNA to bend back on itself, which allows for placement of regulatory sequences far from the actual site of transcription.
  - → It contains a TATA box (sequence TATAAA), which binds to TATA binding protein that assists in the formation of the RNA polymerase transcriptional complex. TATA box typically lies very close to the transcription start site (often within 50 bases).

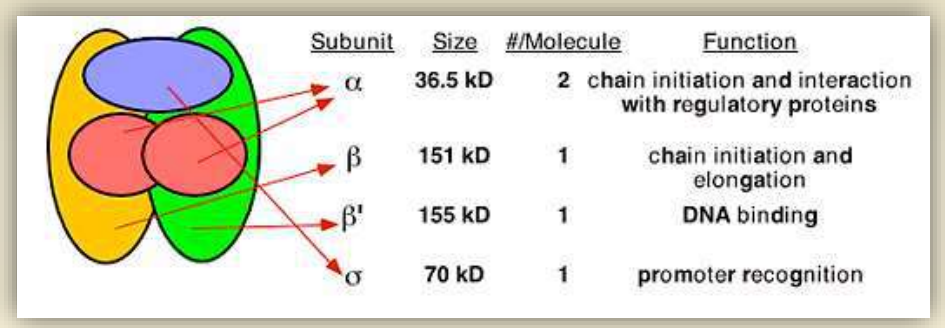
http://www.scfbio-iitd.res.in/tutorial/promoter.html

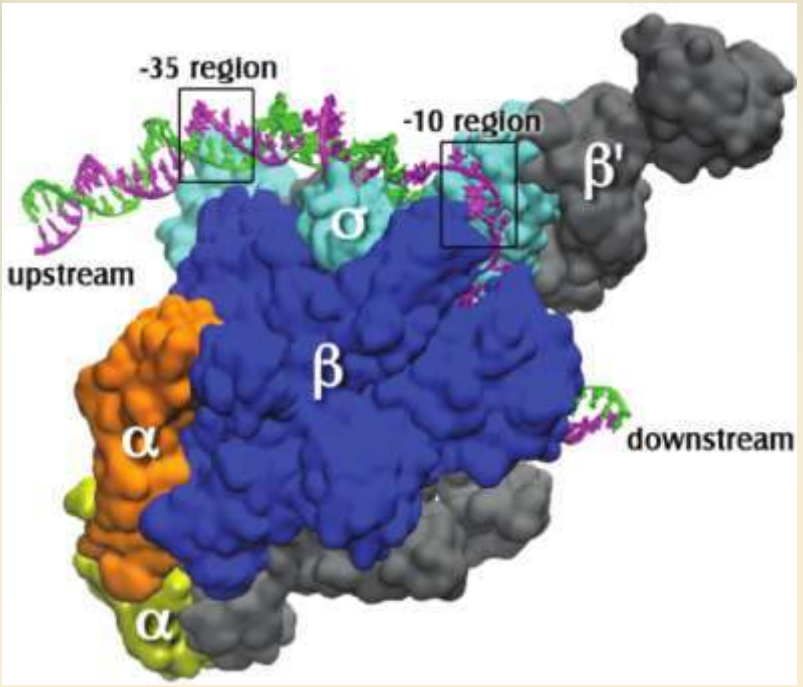
Initiation of Transcription



Initiation of Transcription

Prokaryotic RNA polymerase - Holoenzyme





Karpen M.E. Biomolecules 2015; 5(2):668-678

- RNA polymerase binds to one of the several specificity factor, σ, to form a holoenzyme.
- Holoenzyme then recognize and bind to specific promoter regions in the DNA to form a complex referred to as closed complex.
- The dissociation of *σ* allows the core enzyme to proceed along the DNA template, synthesizing mRNA by adding RNA nucleotides according to the base pairing rules.
- The transcribed strand of DNA is called the template strand to which mRNA product is complementary and is almost identical to the other DNA strand, called the non-template strand.

Initiation of Transcription

Types of RNA polymerase

Unlike prokaryotes where all RNA is synthesized by a single RNA polymerase, the eukaryotes have 3 RNA polymerases responsible for transcribing different types of RNA, which are:

I. **RNA polymerase I (RNA Pol I)** - located in the nucleolus and transcribes the 28S, 18S, and 5.8S rRNA genes.
II. **RNA polymerase II (RNA Pol II)** - located in the nucleoplasm and transcribes protein-coding genes, to yield pre-mRNA, and also the genes encoding small nucleolar RNAs (snoRNAs) involved in rRNA processing and small nuclear RNAs (snRNAs) involved in mRNA processing, except for U6 snRNA.
III. **RNA polymerase III (RNA Pol III)** - also located in the nucleoplasm. It transcribes the genes for tRNA, 5S rRNA, U6 snRNA, and the 7S RNA associated with the signal recognition particle (SRP) involved in the translocation of proteins across the endoplasmic reticulum membrane.

Each of the three eukaryotic RNA polymerases contains 12 or more subunits and so these are large complex enzymes. They show DNA sequence similarities with the genes of subunits of E. coli core enzymes (4 to 7 other subunits show no similarity either with bacterial RNA polymerase subunits or with the subunits of other eukaryotic RNA polymerases).

Initiation of Transcription

RNA polymerase subunuts and promoter recognition

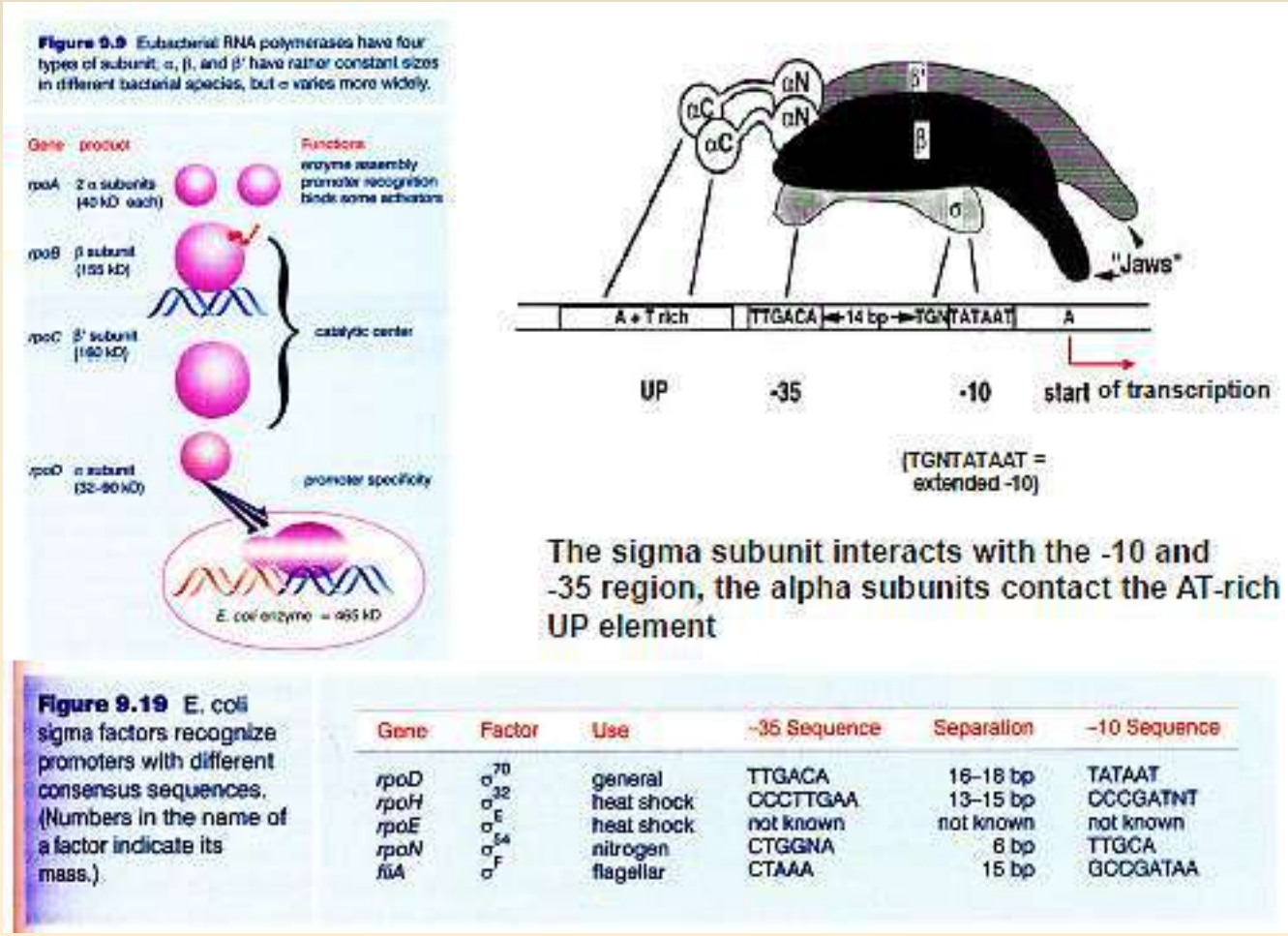
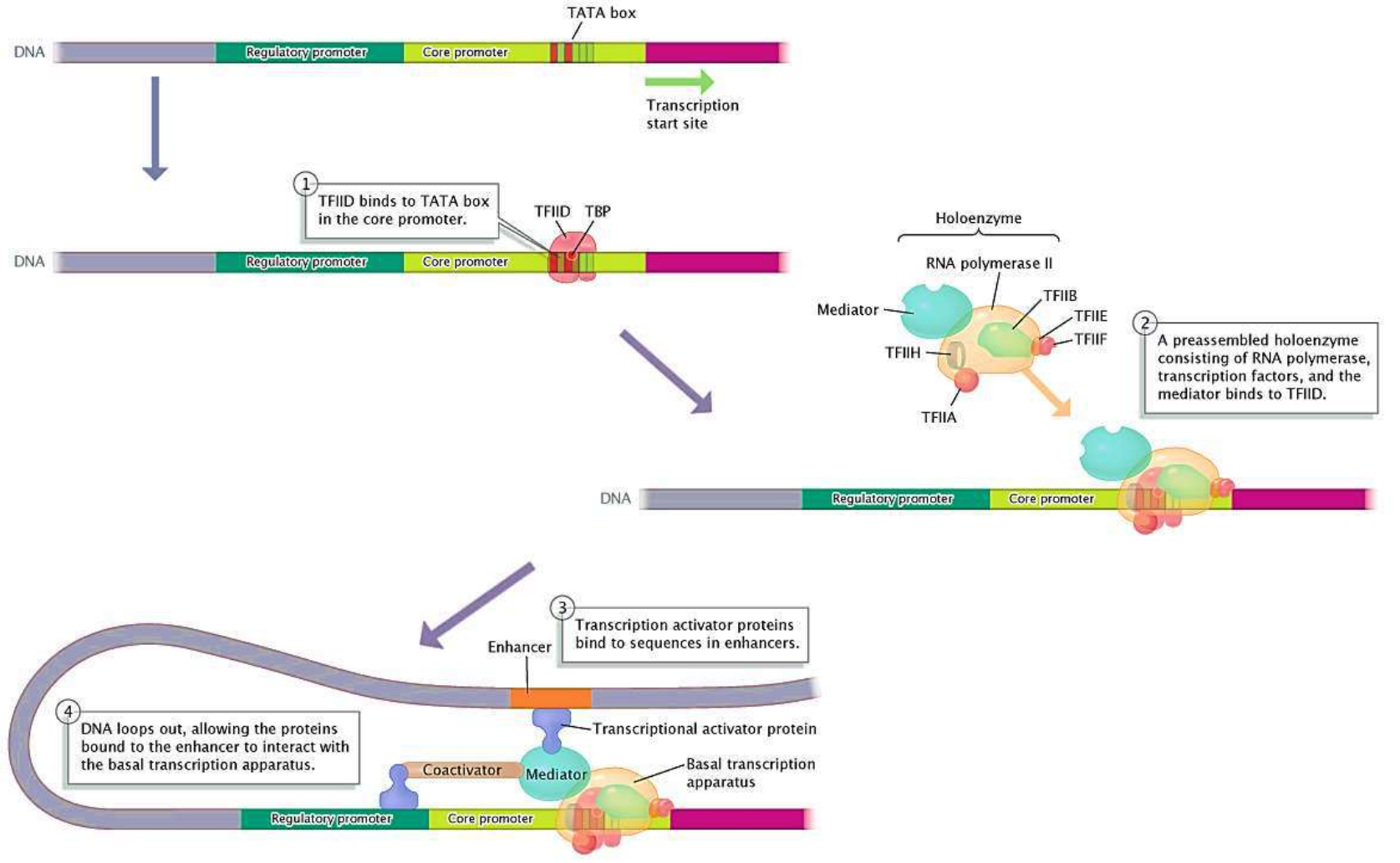


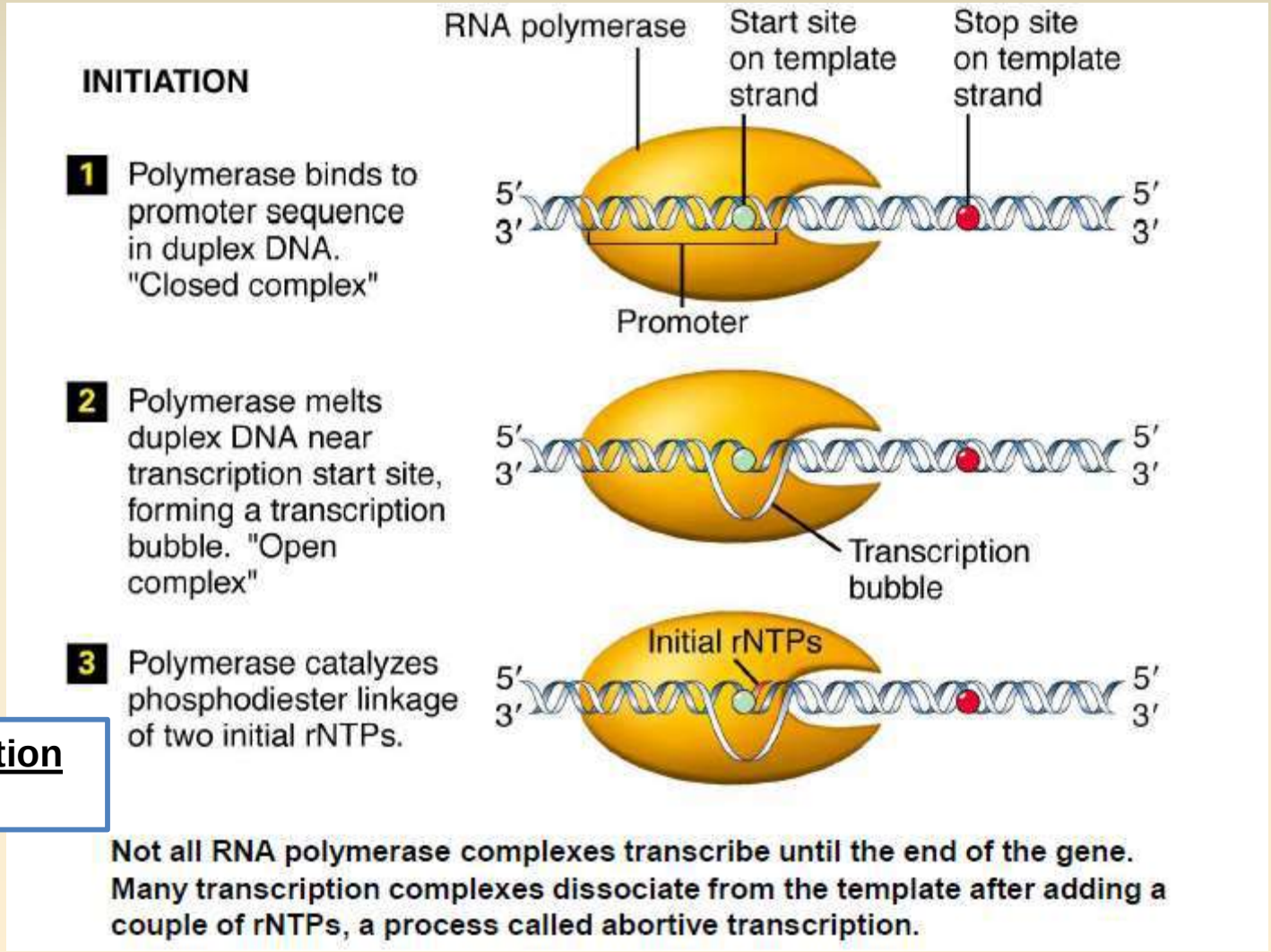
Figure 9.19 E. coli sigma factors recognize promoters with different consensus sequences. (Numbers in the name of a factor indicate its mass.)

Gene	Factor	Use	−35 Sequence	Separation	−10 Sequence
rpoD	σ^{70}	general	TTGACA	16–18 bp	TATAAT
rpoH	σ^{32}	heat shock	CCCTTGAA	13–15 bp	CCCGATNT
rpoE	σ^{E}	heat shock	not known	not known	not known
rpoN	σ^{54}	nitrogen	CTGGNA	6 bp	TTGCA
fliA	σ^{F}	flagellar	CTAAA	15 bp	GCCGATAA

Initiation of Transcription



Initiation of Transcription



Transcription Cycle

Initiation of Transcription

Transcription factors

- Transcription factors are proteins that influence the ability of RNA polymerase to transcribe a given gene.
- There are two main types of transcription factors:

  - ***General transcription factors:***
    - Required for the binding of the RNA pol to the core promoter and its progression to the elongation stage.
    - They are necessary for basal transcription.

  - ***Regulatory transcription factors:***
    - Serve to regulate the rate of transcription of nearby genes
    - They influence the ability of RNA pol to begin transcription of a particular gene.
    - They recognize *cis* regulatory elements near the core promoter, such as response elements, control elements or regulatory elements – binding affects the transcription of associated gene.

Initiation of Transcription

Transcription factors

- Regulatory transcription factors are of three types:
  - ***Activator:*** A regulatory proteins that increases the rate of transcription is termed as activator.
    - The sequence it binds is called ***enhancer***.
  - ***Repressor:*** A regulatory protein that decreases the rate of transcription is termed as repressor.
    - *The sequence it binds is called a* ***silencer****.*

Initiation of Transcription

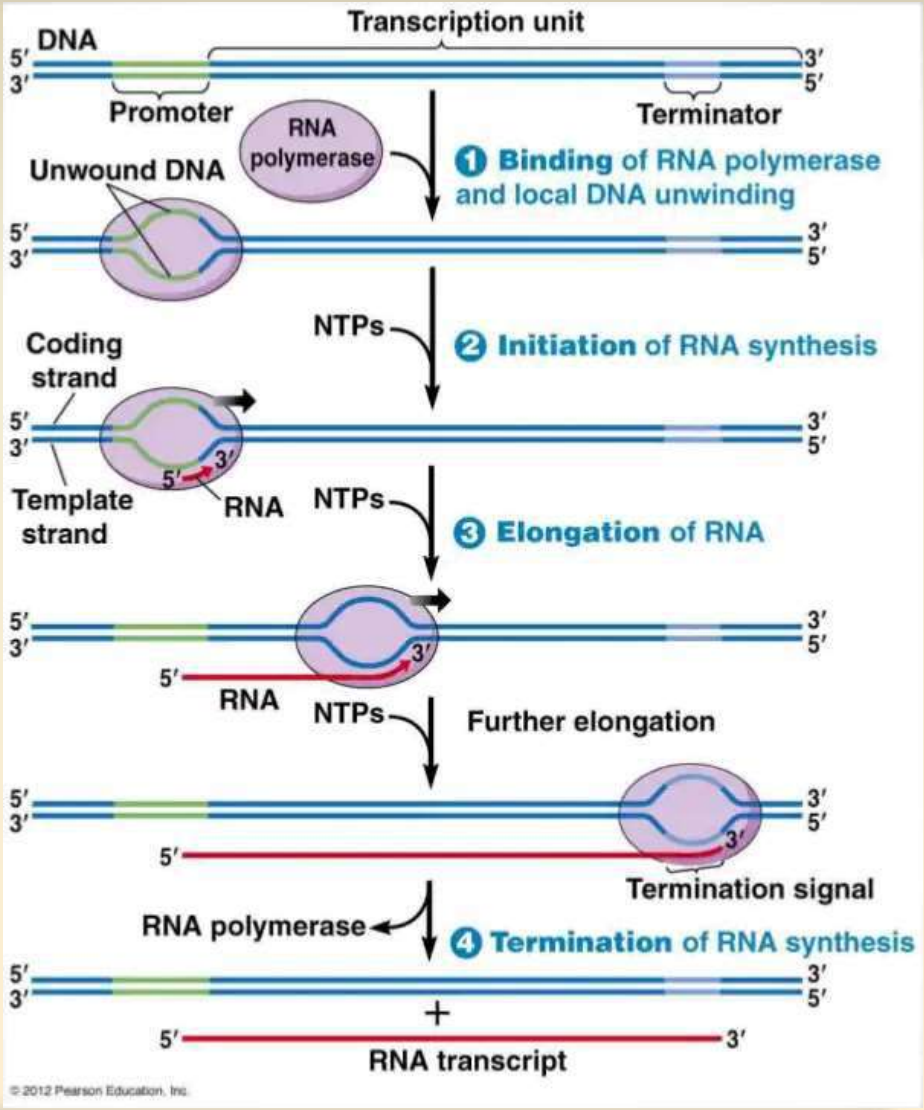
Functional domainds of eukaryotic transcription factors

- ***DNA binding domain (DBD)***
  - Binds specific sequence of base pairs
- ***Transcriptional activating domain (TAD)***
  - Interacts with basal transcription factors or directly with RNA ploymerase II
- ***Protein-protein interaction domain (PPID)***
  - Home- and Hetero-dimerization domains
  - Interaction with other transcription factors (besides the basal apparatus)

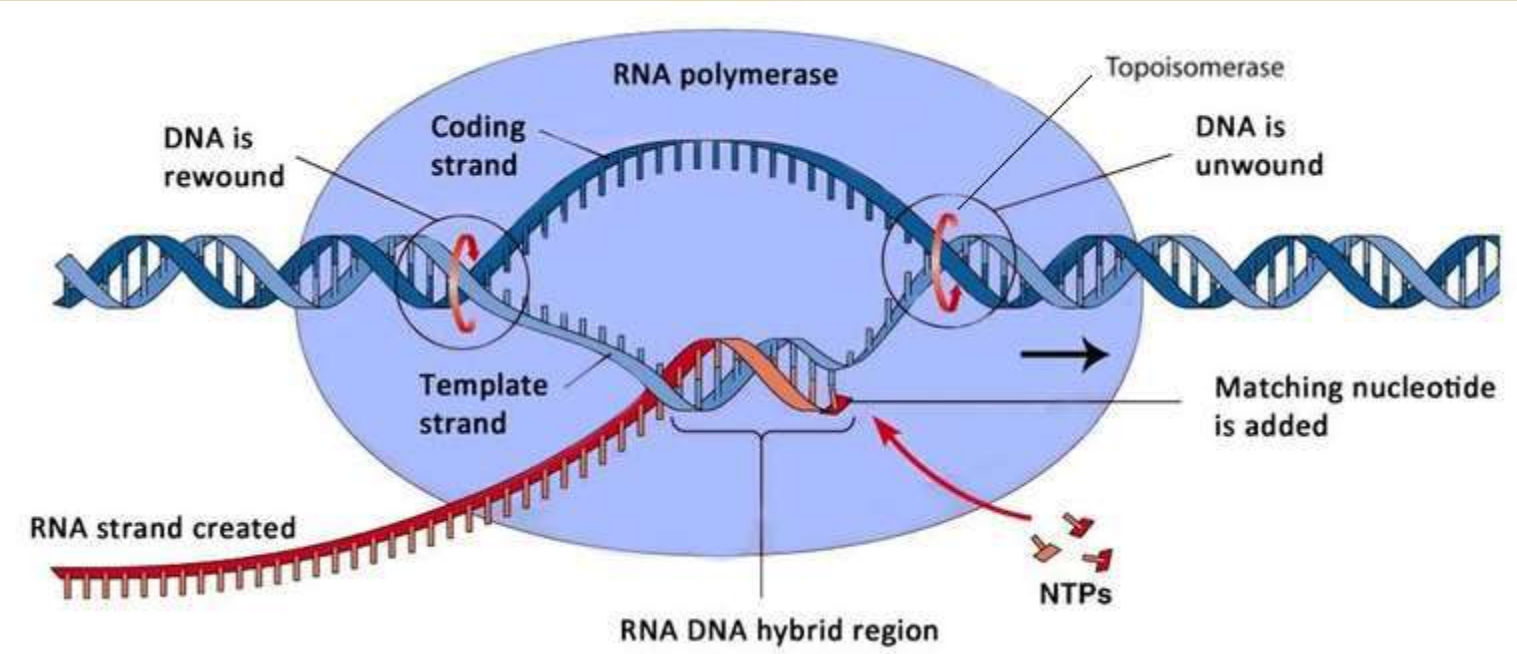
- A domain is a region on a transcription factor that have specific functions.
  - One domain could be for DNA binding
  - Another could provide a binding site for effector molecules.
- A motif is a domain or portion of it that has a very similar structure in many different proteins.

Elongation of Transcripts

- The RNA polymerase transcribes the DNA (β subunit initiates the synthesis), but produces about 10 abortive (short, non-productive) transcripts which are unable to leave the RNA polymerase because the exit channel is blocked by σ-factor.
- The σ-factor eventually dissociates from the core enzyme and elongation proceeds.
- This enzyme has no exo/endonuclease activity and cannot repair the mistakes as DNA polymerase in replication.

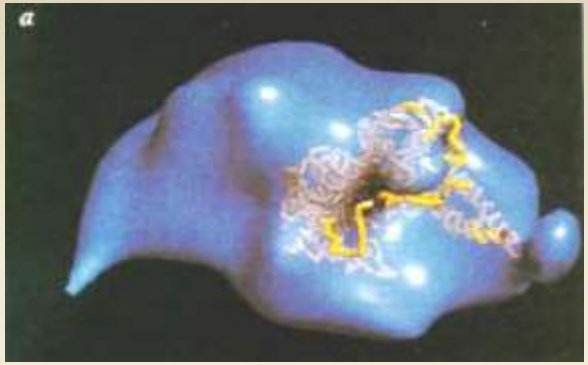


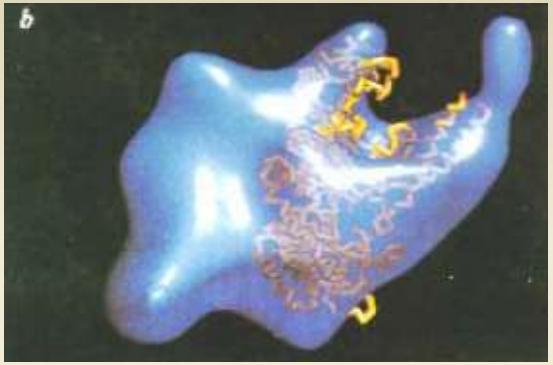
Elongation of Transcripts



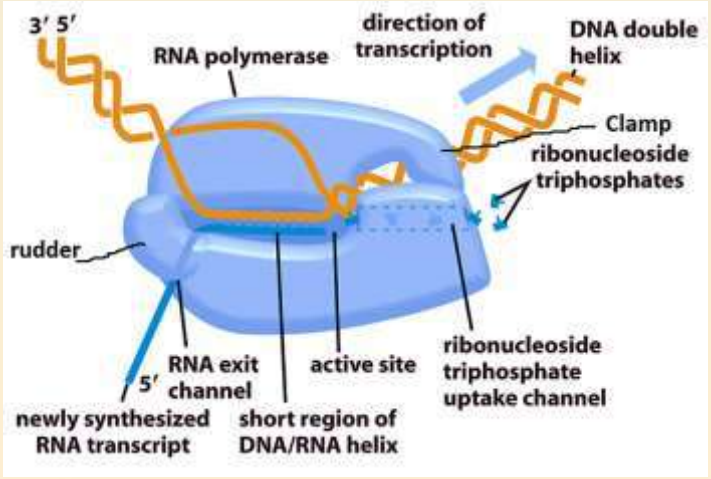
- RNA polymerase add complementary bases to the template strand of DNA. It adds Thiamine for Adenine (T = A), Guanine for Cytosine (G ≡ C), Cytosine for Guanine (C ≡ G) and Adenine for Uracil (A = U).
- Most transcripts originate using adenosine-5'-triphosphate (ATP) and to a lesser extent, guonosine-5'-triphosphate (GTP) at the +1 site. Uridine-5'-triphosphate (UTP) and cytidine-5'-triphosphate are disfavored at the initiation site.

Elongation of Transcripts



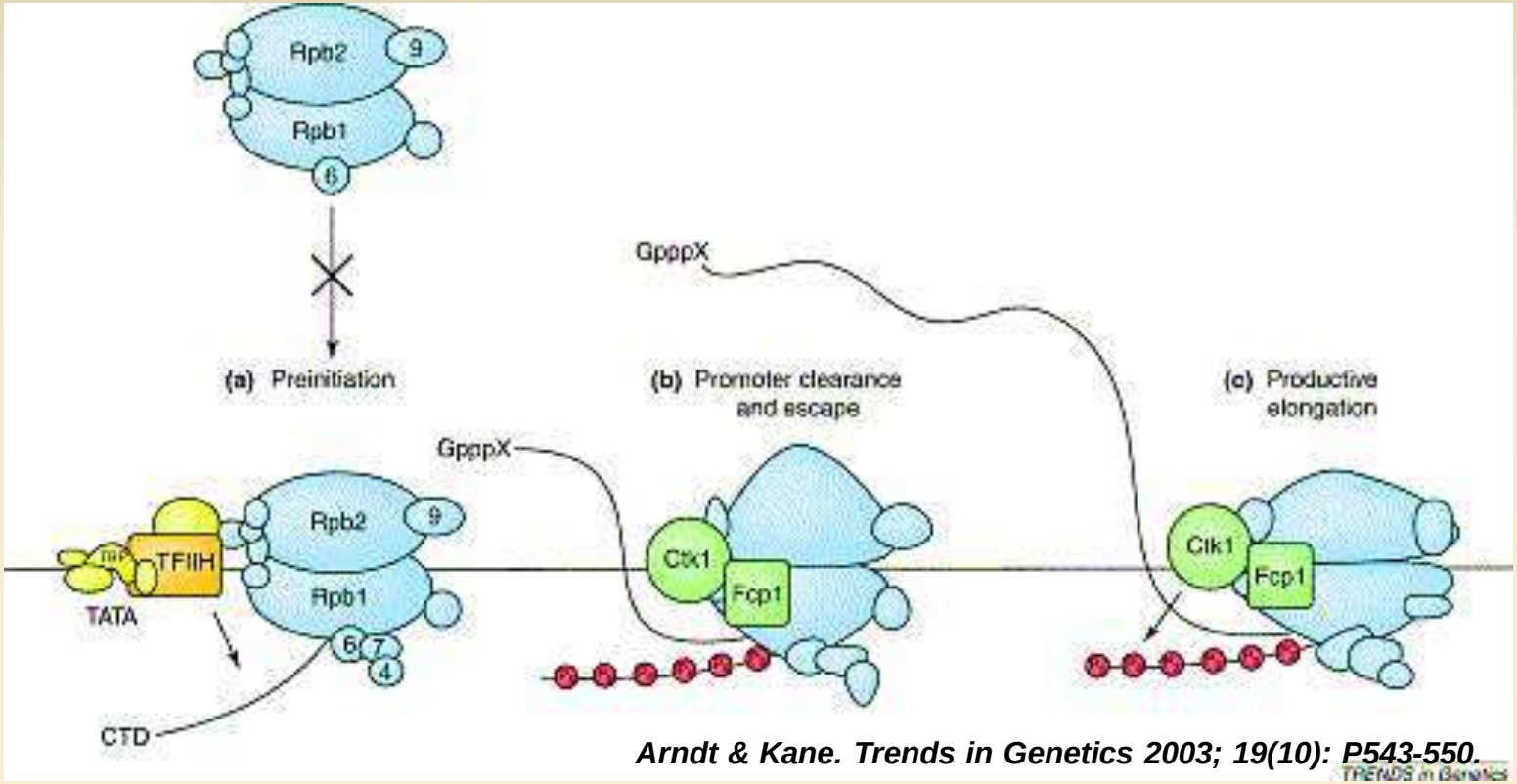


- **RNA polymerase exhibits groove which have length of 55 Å and diameter of 25 Å.**
- **This groove fits well the 20 Å double strand of DNA.**
- **The 55 Å length can accept 16 nucleotides.**



- Formation of <u>closed complex</u> by binding of the promoter sequence.
- Thereafter, the RNA polymerase holoenzyme separates 10-14 bases extending from -11 to +3 to form a open complex called **<u>melting</u>** . The changing from closed to open complex is called **<u>isomerization</u>**.
- RNA polymerase starts synthesizing RNA.

Elongation of Transcripts

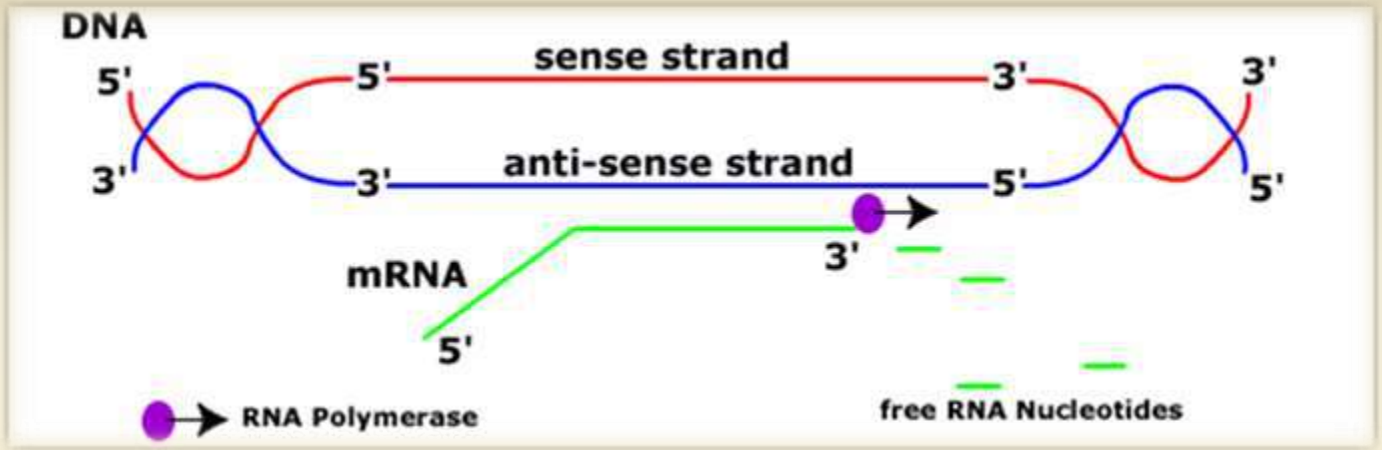


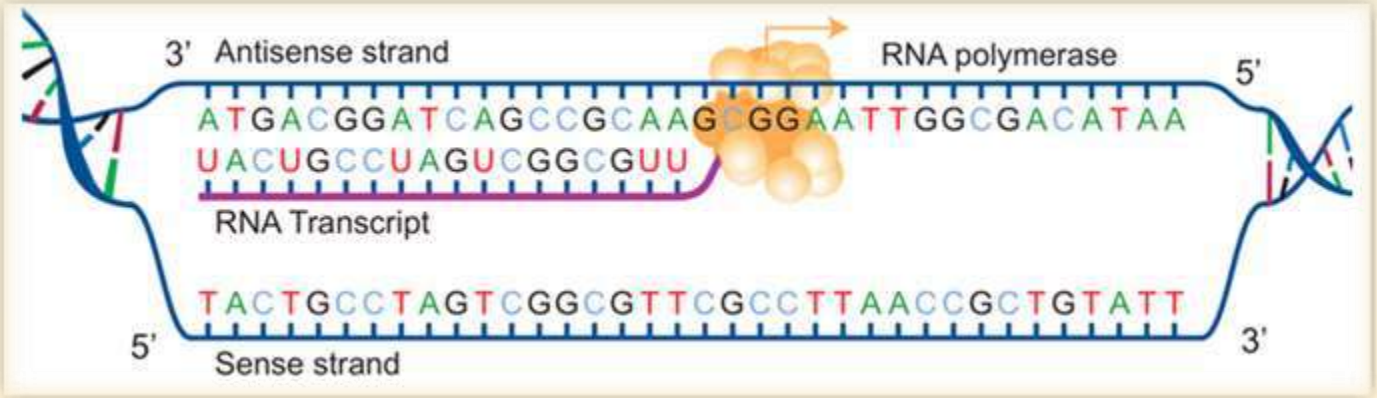
Arndt & Kane. Trends in Genetics 2003; 19(10): P543-550.

Eukaryotic transcript elongation

- Rate of elongation with E. coli RNA pol = ~40 nucleotides/sec; T3 RNA pol = ~200 nucleotides/sec
- E. Coli RNA pol covers (footprints) ~28-35 bp of DNA during elongation

Elongation of Transcripts

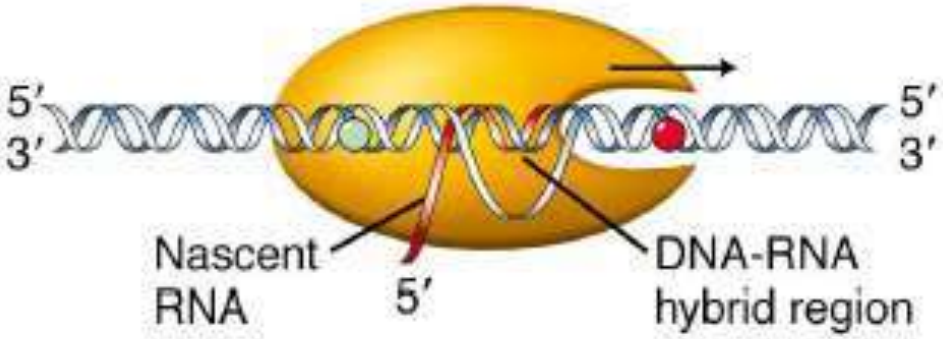




Antisense strand act as template for transcription, and therefore transcripts (RNAs) are complimentary to the sense strand of the DNA.

Elongation of Transcripts

4 Polymerase advances 3′ → 5′ down template strand, melting duplex DNA and adding rNTPs to growing RNA.

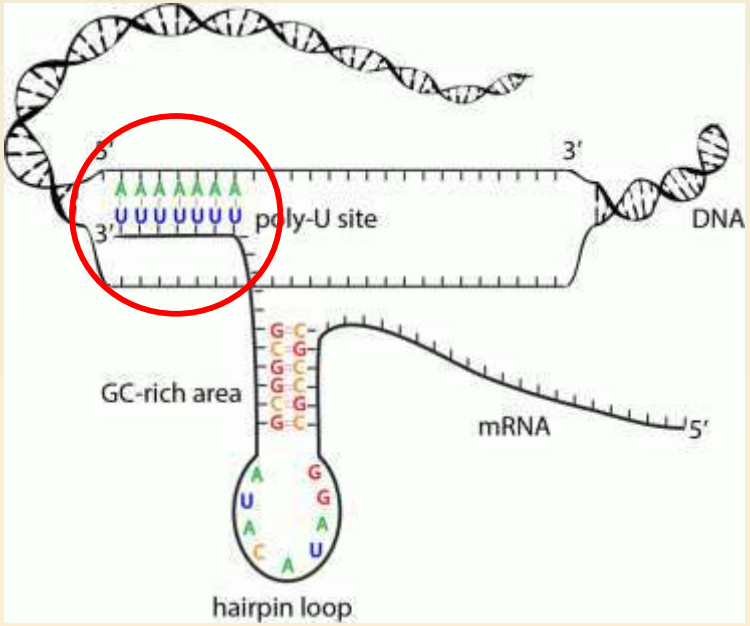


Pyrophosphorolytic Editing: The polymerase backtracks and removes an incorrectly inserted ribonucleotide by reincorporation of PPi.

Hydrolytic editing: The polymerase backtracks and cleaves the RNA, removing error-containing sequence. The process is stimulated by Gre factors, which also function as elongation stimulators.

Termination of Transcription

- At transcription stop site, polymerase releases completed RNA and dissociates from DNA. In prokaryotes, termination of transcription is stopped by two way; one is rho-independent and other is rho-dependent way.
- **Rho-independent termination:** It involves terminator sequences consisting of a short inverted repeat (about 20 nucleotides) followed by a stretch of 8 A:T base pairs. The resulting RNA forms a stem-loop structure, which disrupts the elongation complex. A stretch of A:U base pairs in the DNA/RNA hybrid are weaker than other base pairs and are more easily disrupted as a consequence of stem loop nformation.

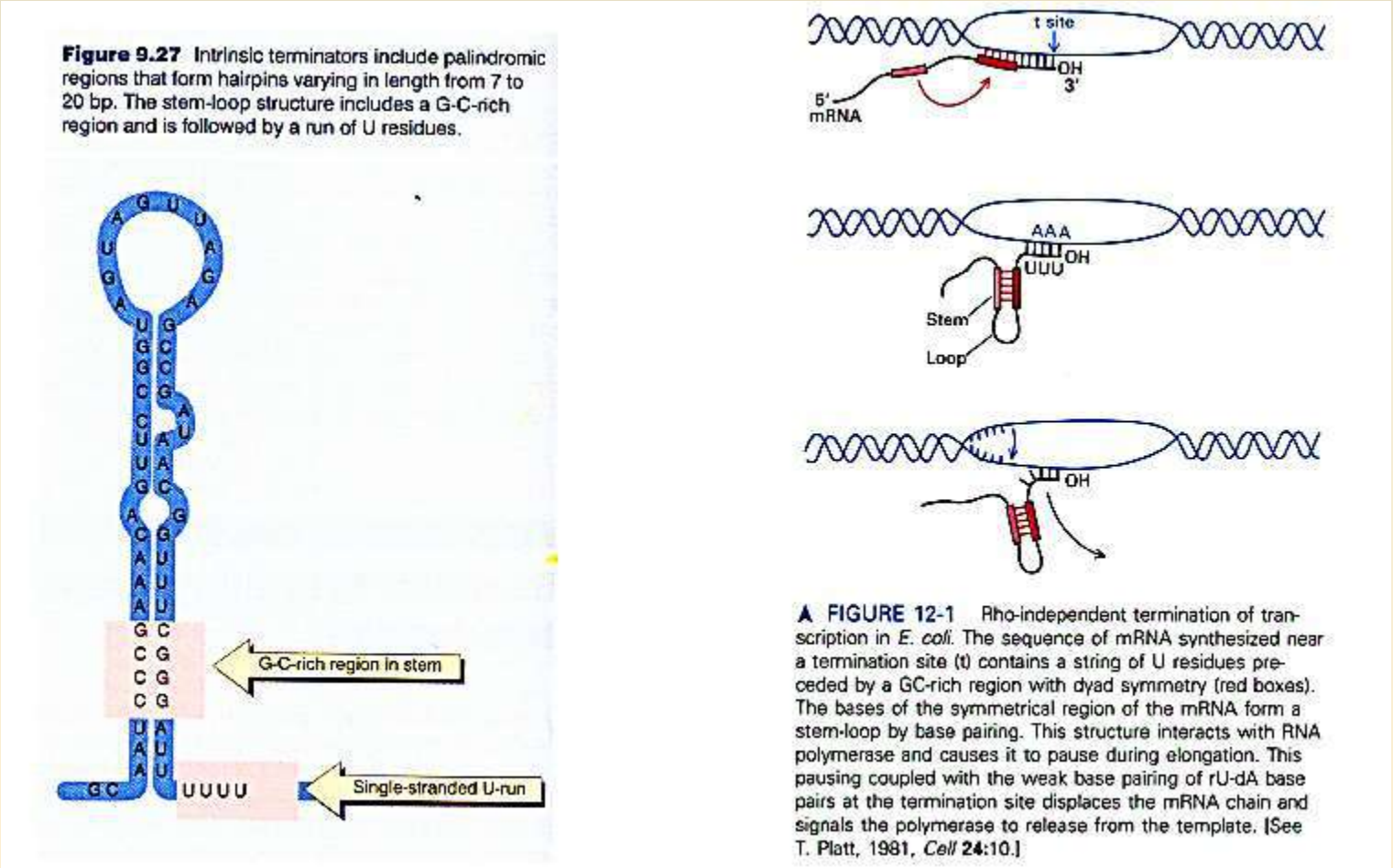


It has following characteristics:

- A strong G-C rich stem and loop.
- A sequence of 4-6 U-residues in the RNA, which are transcribed from a corresponding stretch of a DNA as in the template.

Termination of Transcription

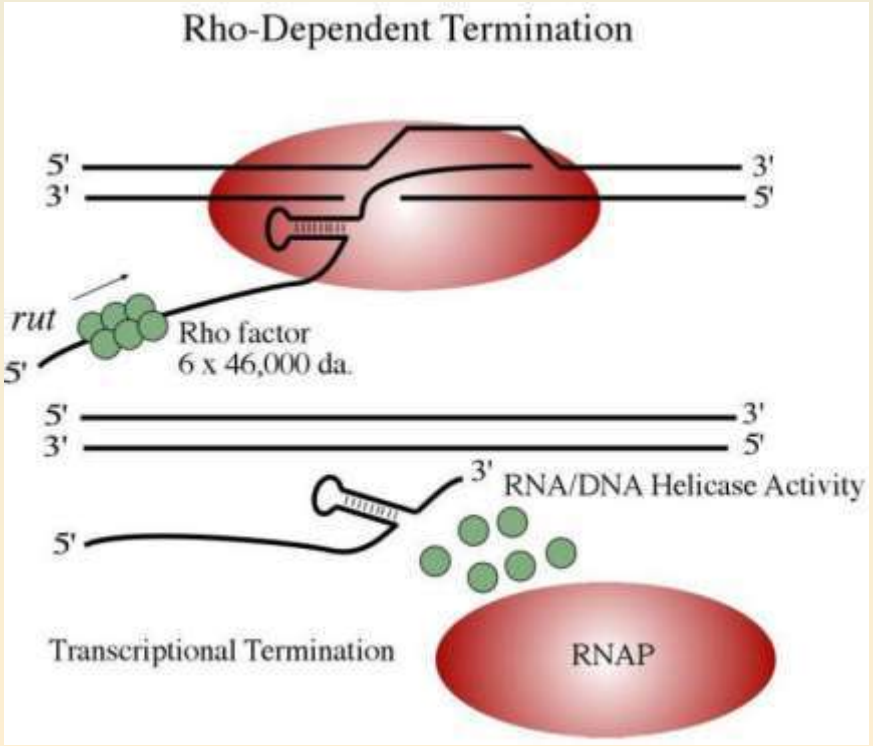
Rho Independent Termination



▲ FIGURE 12-1 Rho-independent termination of transcription in *E. coli*. The sequence of mRNA synthesized near a termination site (t) contains a string of U residues preceded by a GC-rich region with dyad symmetry (red boxes). The bases of the symmetrical region of the mRNA form a stem-loop by base pairing. This structure interacts with RNA polymerase and causes it to pause during elongation. This pausing coupled with the weak base pairing of rU-dA base pairs at the termination site displaces the mRNA chain and signals the polymerase to release from the template. [See T. Platt, 1981, *Cell* **24**:10.]

Termination of Transcription

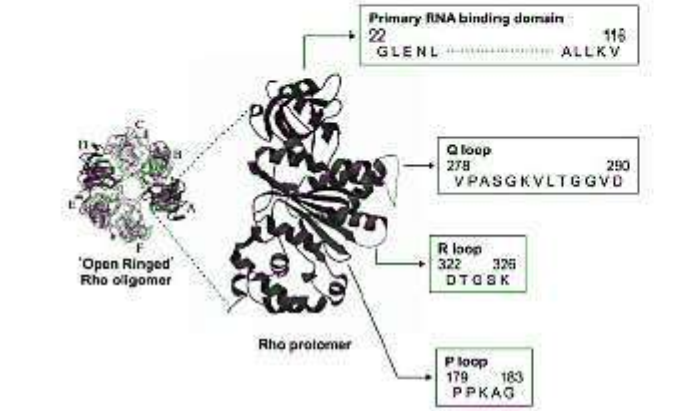
- **Rho-dependent termination:** Terminators are not characterized by specific RNA elements that fold in secondary structure**.** It involves termination factor called ρ factor (rho factor) which is a protein to stop RNA synthesis at specific sites, known as rut sites (Rho utilization) on the single stranded RNA.
- Rho factor binds to rut site on the nascent RNA strand and runs along the mRNA towards the RNA polymerase.
- A stem loop structure upstream of the terminator region pauses the RNA polymerase, when ρ factor reaches the RNA polymerase, it causes RNAP to dissociate from the DNA.

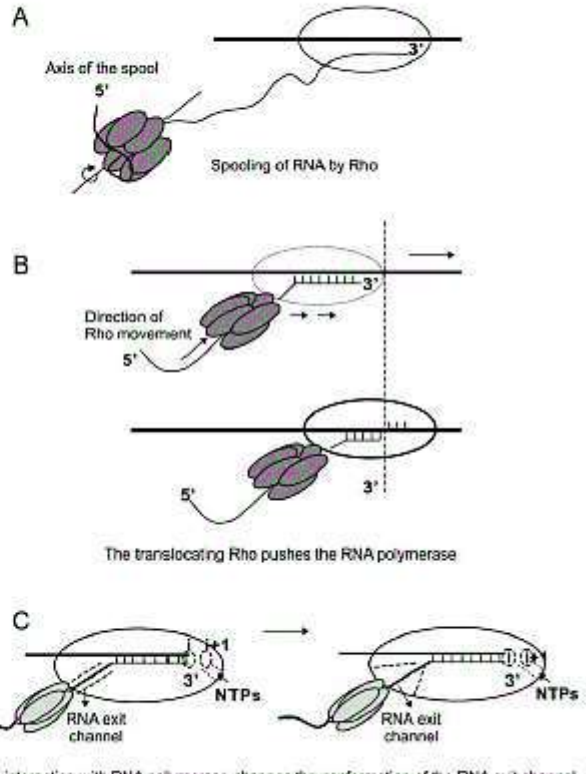


Termination of Transcription

Rho dependent Termination

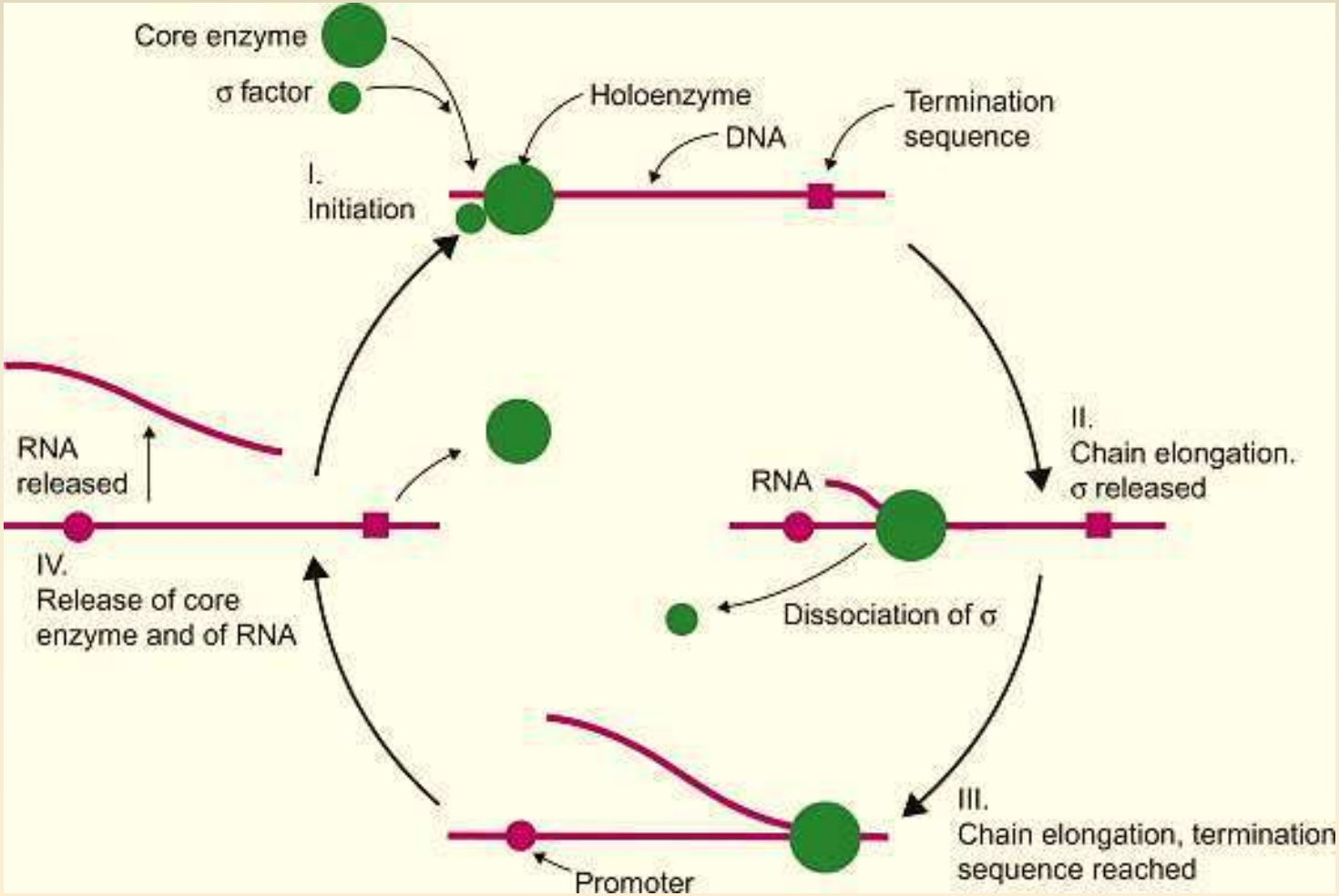
Rho binds as a hexameric protein complex to specific sequences called RUT (rho utilization) sites. The complex also binds ATP and moves along the RNA ultimately disrupting the interactions between the RNA polymerase and the RNA.





Banerjee et al., Journal of Microbiology, 2007

An Overview of Transcription



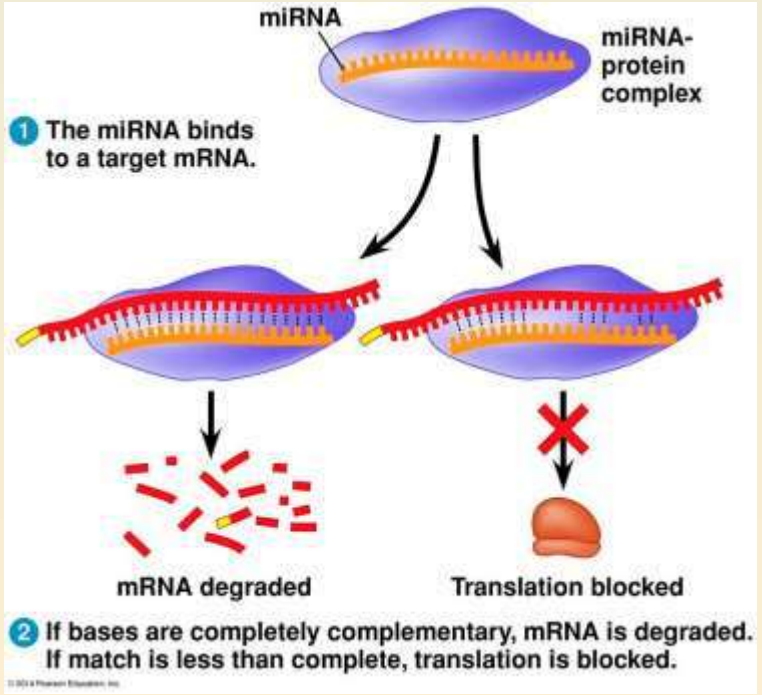
Regulation of Transcription

→ **Enhancer** – control element far from a gene or intron

→ **Activator** – bind to enhancers to **turn on** transcription of a gene
- Transcription factors + enhancer + activator + RNA Polymerase II = **transcription initiation complex**
- Needed for transcription to begin

→ **Repressors** – inhibit gene expression
- Turn off transcription
- Block activators from binding to enhancers

→ **Micro RNA (miRNA)** – It can degrade or block translation.
- It causes mRNA to fold on itself and base pair to create dsRNA which is then digested with an enzyme.



→ **Short interfering RNA (siRNA)** – It also degrades mRNA or block translation (blocking by siRNA is called RNAi, or RNA interferance).

Further reading

- Willey J., Sherwood L., Woolverton C.J. 2017. Prescott's Microbiology 10th Edition, McGraw Hill Publication, New York, USA
- Krebs J.E., Goldstein E.S., Kilpatrick S.T. 2017. Lewin's Genes XII. Jones and Bartlett Publishers, Inc., Burlington, MA, USA
- Snyder L.R., Peters J.E., Henkin T.M., Champness W. 2013. Molecular Genetics of Bacteria, 2nd ed., ASM Press, Washington DC, USA, 2003.
- Graumann P.L. Chromosome architecture and segregation in prokaryotic cells. Microbial Physiology 24(5-6).
- Griswold A. 2008. Genome Packaging in Prokaryotes: the Circular Chromosome of *E. coli*. Nature Education 1(1):57.
- Kuzminov A. 2014. The precarious prokaryotic chromosome. Journal of Bacteriology 196(10):1793-1806.