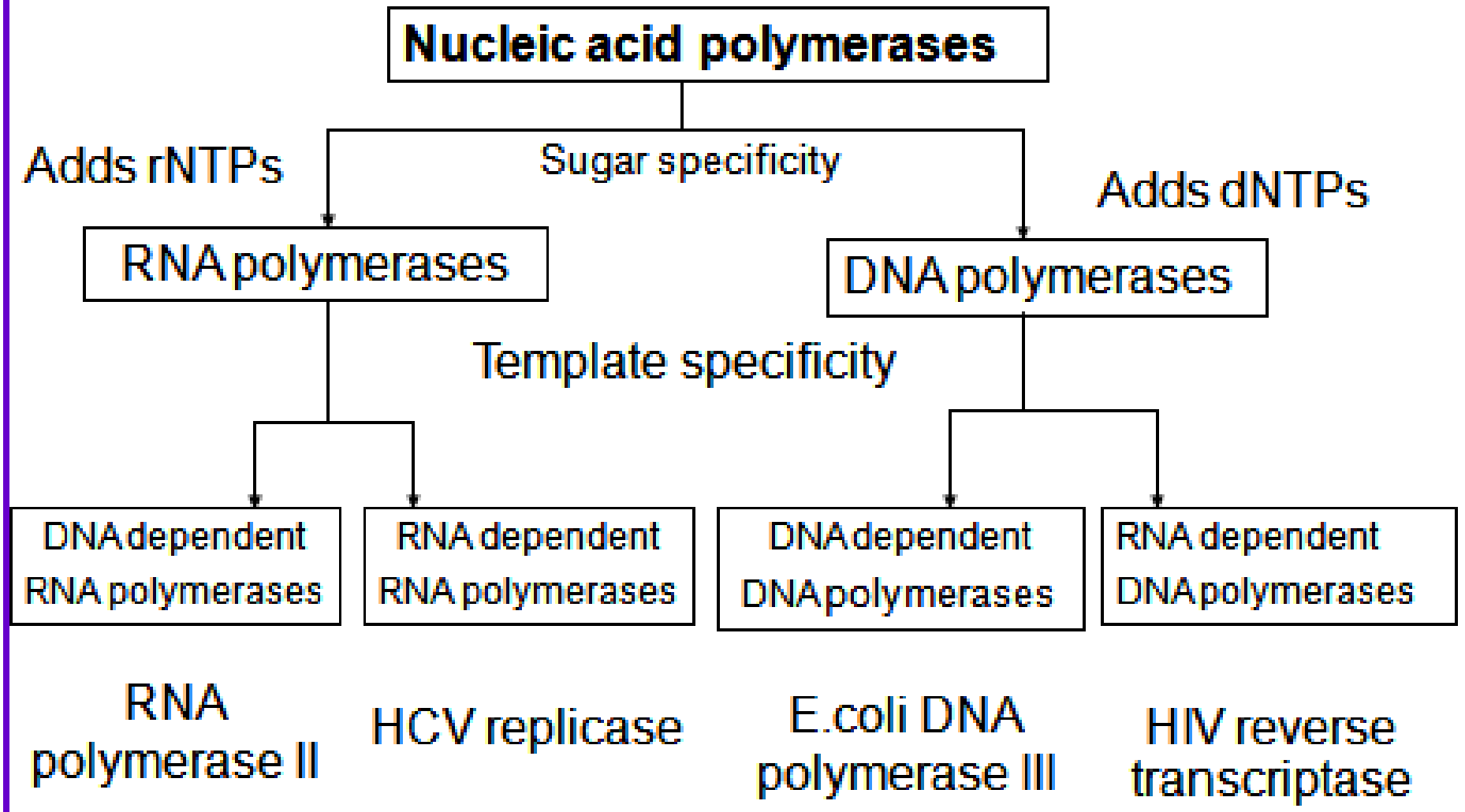
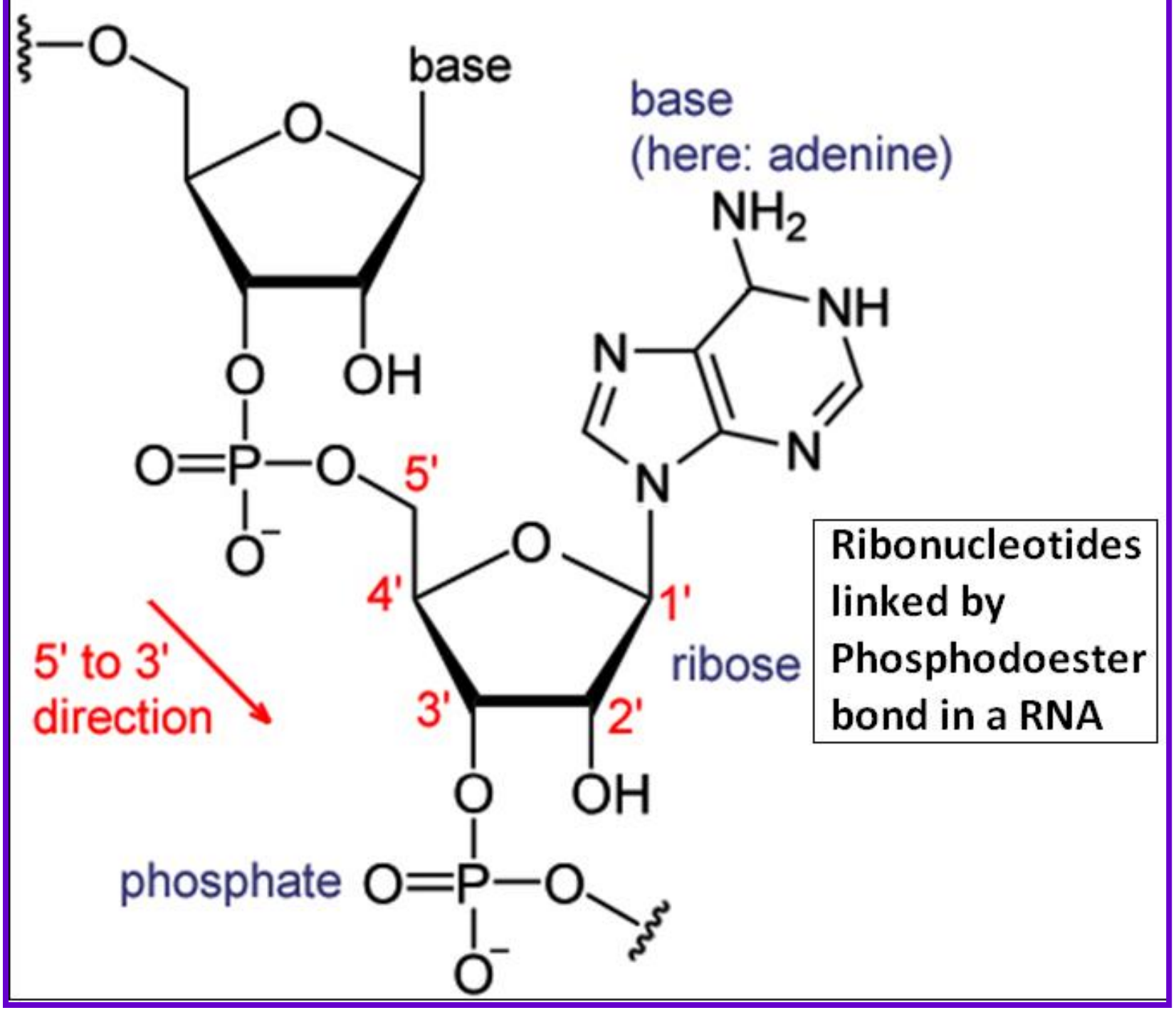
TRANSCRIPTION
IN
EUKARYOTES

Nucleic acid polymerases: classification




base
base
(here: adenine)
NH2
NH
N
N
N
O
OH
O
O=P—O
O⁻
5'
4'
1'
3'
2'
ribose
5' to 3'
direction
phosphate
Ribonucleotides
linked by
Phosphodoester
bond in a RNA

RNA Polymerases in Eukaryotes

- RNA polymerase I
- RNA polymerase II
- RNA polymerase III
- RNA polymerase IV and
- RNA polymerase V
- RNA polymerases in mitochondria and chloroplasts

RNA Polymerases in Eukaryotes

- All eukaryotic RNA polymerases seen in the nucleus have at least 12 subunits and are aggregates of >500 kD.
- Some subunits are common to all three RNA polymerases.

RNA Polymerase I

- RNA polymerase I has 14 subunits and its molecular weight is more than 500 kD.
- It is present in nucleus/ nucleolus
- It is resistant to alpha –amanitin a toxin found in the poisonous "**Deathcap mushroom**" - *Amanita phalloides*
- It synthesizes a pre-rRNA 45S, which matures into 28S, 18S and 5.8S rRNAs

RNA Polymerase II

- It is present in nucleus.
- It is very sensitive to alpha –amanitin.
- It synthesizes hnRNA - precursors of mRNAs, mi RNAs (micro RNA) and most snRNAs
- There are 12 subunits in RNA polymerase II.
- The subunits Rpb1, 2, and 3 are homologous to β, β’, and α of bacterial RNA polymerase.
- The subunits Rpb5, 6, 8, 10, and 12 are found in all three eukaryotic polymerases.

RNA Polymerase II

- Rpb1 -largest subunit. Its CTD (carboxyl terminal domain) has a heptamer sequence of YSPTSPS (Tyr-Ser-Pro-Thr-Ser-Pro-Ser) repeated several times. This sequence is repeated for 26 times in yeast and 50 times in humans.
- CTDs are phosphorylated at several Ser and Tyr residues after transcription initiation.
- Phosphorylation of CTDs is important for promoter escape, elongation and post-transcritption events
- CTDs act as a “landing pad” for different proteins, involved in elongation & post-transcription events
- CTDs’ Base is located near the exit groove of RNA.

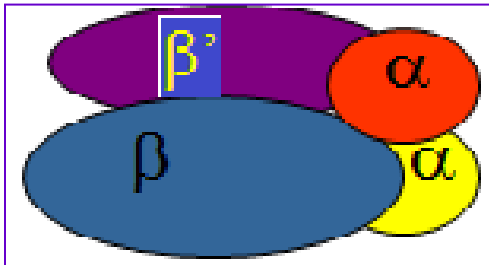
RNA Polymerase III

- RNA Pol III has 17 subunits.
- It is moderately sensitive to alpha amanitin.
- It is present in nucleus
- It synthesizes small RNA such as tRNA, 5S rRNA and snRNAs.

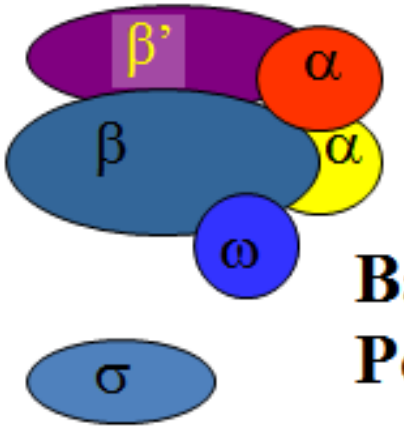
Other RNA Polymerases in Eukaryotes

- RNA polymerase IV is present in the nucleus. It synthesizes siRNA in plants.
- RNA polymerase V is present in the nucleus. It synthesizes RNAs involved in siRNA directed heterochromatin formation in plants.
- Chloroplast and Mitochondria have their own RNA Polymerases.

Chloroplast RNA polymerase



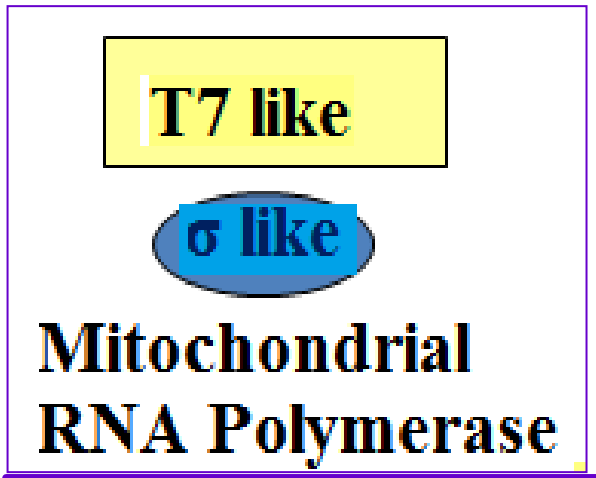
Chloroplast RNA Polymerase

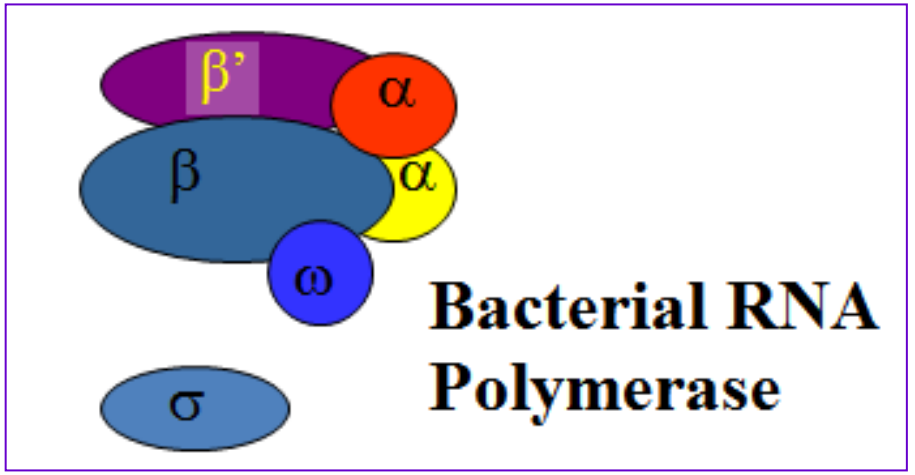


Bacterial RNA Polymerase

- Chloroplast RNA polymerases are encoded by chloroplast genome and have 4 subunits.
- These subunits have considerable homology with α, β, β' subunits of *E. coli*.
- There aren't any sigma like factors or general transcription factors in Chloroplast RNA polymerases.

Mitochondrial RNA polymerase





- Mitochondrial RNA polymerase is encoded in nuclear RNA & transported to mitochondrial matrix.
- It contains two subunits, one similar to T7 (bacteriophage) RNA polymerase and other similar to bacterial sigma factor
- Note: T7 RNA polymerase is a single polypeptide (98,800 Daltons) and is specific for its own promoters, a conserved 23-base-pair (bp) sequence.

Transcription by RNA Polymerase II

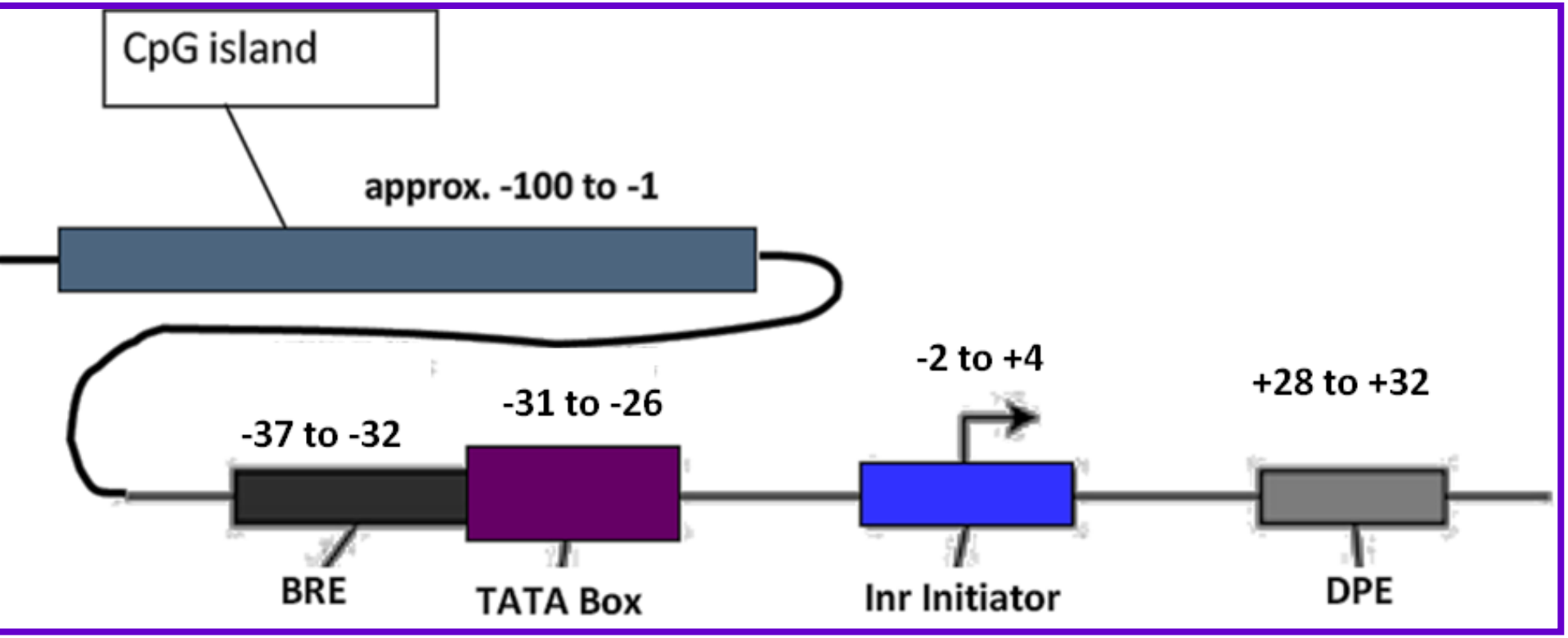
Promoters of RNA Polymerase II

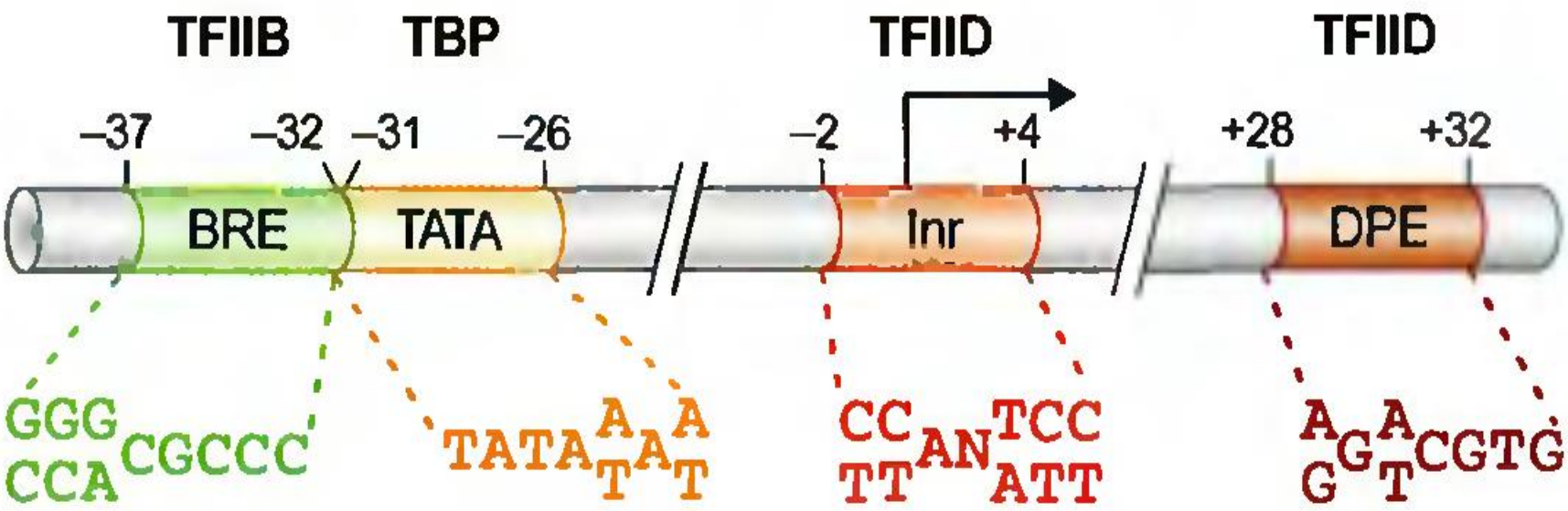
- Eukaryotic core promoter is the “minimal set of DNA sequences required for initiating transcription by RNA Polymerase holoenzyme”.
- A core promoter lies within 40 nucleotides either upstream or downstream of the transcription start site.
- The four common elements found in Pol II core promoters are
  (1) **TFIIB R**ecognition **E**lement (BRE)
  (2) **TATA** element (or box)
  (3) **initiato**r (inr) and
  (4) **D**ownstream **P**romoter **E**lement (DPE).

Promoter of RNA Polymerase II

- Typically, a promoter includes only **two** or **three** of these four elements.
- A core promoter for RNA polymerase II includes the Inr and, commonly, either a TATA box or a DPE.
- It may also contain other minor elements.

Summary of promoter elements





Pol II core promoter showing location of BRE, TATA, Inr and DPE relative to the transcription start site. The consensus sequence (below) and the recognizing general transcription factor (above) were shown.

TFIIB Recognition Element (BRE)

- **TFIIB Recognition Element (BRE)** is a DNA sequence found in the promoter region of most genes in eukaryotes and Archaea.
- The BRE is a cis-regulatory element that is found at -37 to -32 immediately upstream of the TATA box, and consists of 7 conserved nucleotides.
- It's consensus sequence is

(G/C)(G/C)(G/A)CGCCC

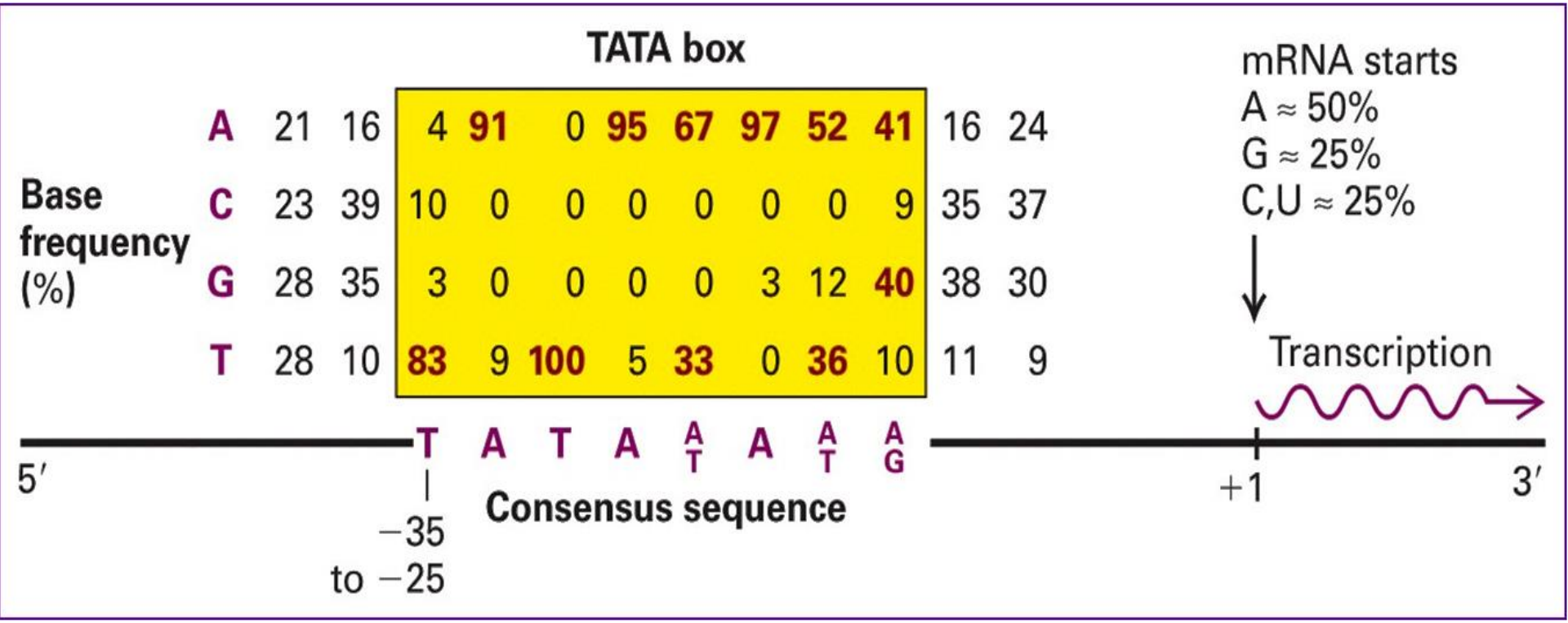
TFIIB Recognition Element (BRE)

- Transcription Factor IIB (TFIIB) recognizes this sequence in the DNA, and binds to it.
- The fourth and fifth alpha helices of TFIIB intercalate with the major groove of the DNA at the BRE.
- TFIIB is one part of the pre-initiation complex that helps RNA Polymerase II, bind to the DNA.
- BRE increases the affinity of TFIIB for the promoter.

TATA box (Goldberg-Hogness box)

- **TATA box** is a short nucleotide sequence seen at -25 to -35 bp upstream.
- It's consensus sequence is **TATAAAA**.
- Transcription factor TFIID binds to TATA box.
- This is the eukaryotic equivalent of Prokaryote Pribnow box (-10 region).
- This sequence is present in many eukaryotic promoters.
- The TATA-box-binding domain of yeast has 80% identity with human counterpart.

TATA box (Goldberg-Hogness box)



IUPAC nucleotide code	Base
A	Adenine
C	Cytosine
G	Guanine
T (or U)	Thymine (or Uracil)
R	A or G
Y	C or T
S	G or C
W	A or T

Contd.

IUPAC nucleotide code	Base
K	G or T
M	A or C
B	C or G or T
D	A or G or T
H	A or C or T
V	A or C or G
N	any base
. or -	gap

Initiator element (Inr)

- The **initiator element (Inr)** or **initiator motif**, is a core promoter with the consensus sequence YYANWYY.
- Note: In nucleic acid notation for DNA, Y (pYrimidine) - stands for C/T, R- puRine (A/G), W (Weak) stands for A/T which form only two hydrogen bonds and N for any nucleotide.
- The **Inr** element extends from -2 to +4.
- The **Inr** element is bound by transcription Factor II D (TBP).

Downstream Promoter Element (DPE)

- The **Downstream Promoter Element (DPE)** is a common component of RNA polymerase II promoters that do not contain a TATA box (**TATA-less promoters**).
- **DPE** has the consensus sequence (A/G)G(A/T)CGTG.
- **DPE** extends from +28 to +32
- **DPE** is bound by Transcription Factor II D (TBP)

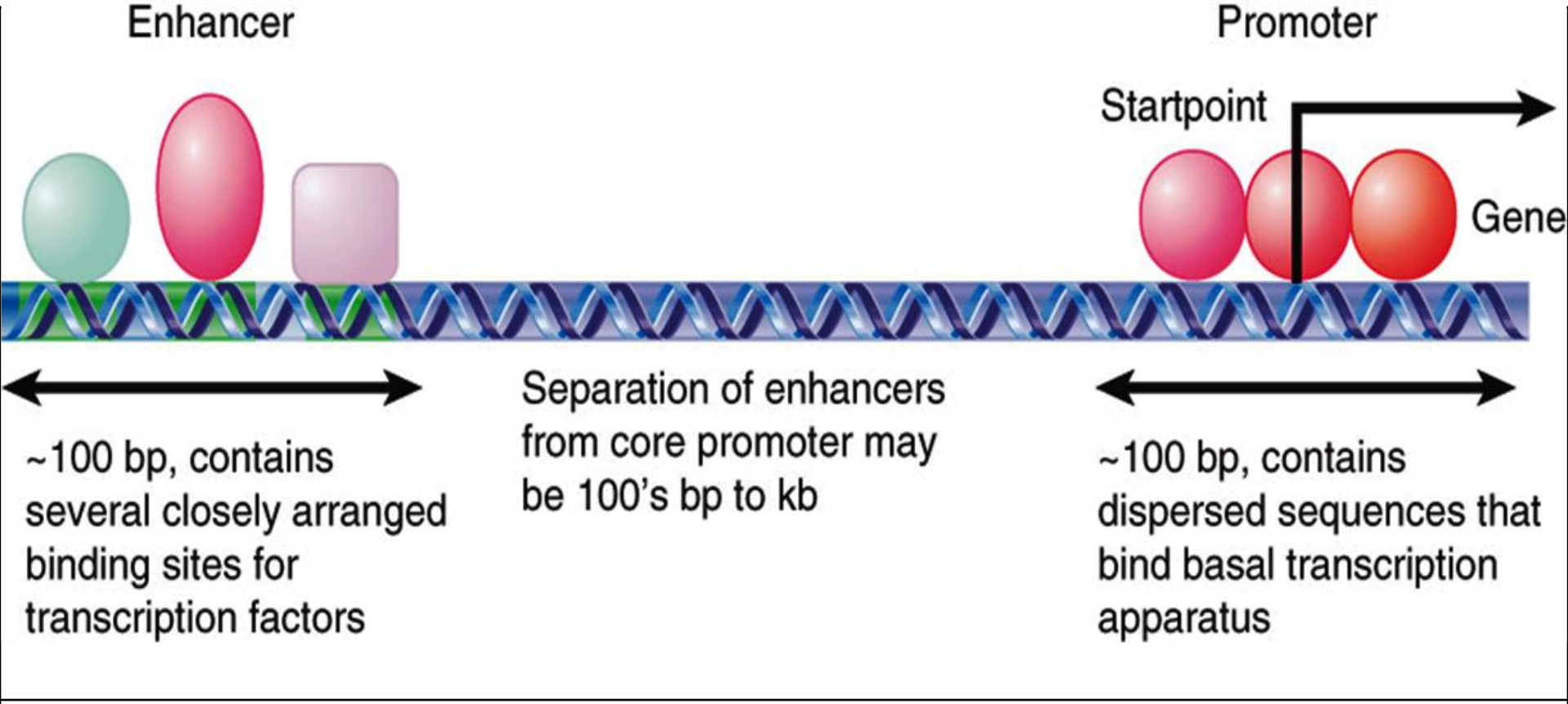
Regulatory Sequences Beyond The Core Promoter

- **CpG islands** are seen in the promoters of constitutively expressed genes & some tissue-regulated genes where they are unmethylated.
- If they are methylated gene expression is stopped. There are ~29,000 CpG islands in the human genome.
- All these DNA elements **bind** to regulatory proteins (activators and repressors), which increase or decrease rate of transcription from the core promoter.
- Some of these regulatory sequences are located many 10s or even 100s of Kb from the core promoters on which they act.

Regulatory Sequences Beyond The Core Promoter

- **P**romoter **P**roximal **E**lements present within 150bp from transcriptional initiation site. Ex. CAAT box / CAT box, GC Box**.**
- **U**pstream **A**ctivator **S**equences (UASs) – present in the upstream within 200bp Ex. – SP1 box.
- **E**nhancers - A *cis*-acting sequence that increases the utilization of (most) eukaryotic promoters, and can function in either orientation and in any location (upstream or downstream) relative to the promoter.

Regulatory Sequences - Enhancers



•Figure: A typical gene transcribed by RNA polymerase II has a promoter that extends upstream from the site where transcription is initiated. It also has an enhancer.

Regulatory Sequences Beyond The Core Promoter

- A series of gene expression regulating elements called **S**ilencers, **B**oundary **E**lements, and **I**nsulators.
- **Silencers** are DNA sequences which decrease or repress the rate of transcription of a gene by binding to the transcription factors.
- **Silencers** are located thousands of base pairs away from the gene they regulate.
- **Silencers** show opposite effect to that of **enhancers**.

Regulatory Sequences Beyond The Core Promoter

- **Boundary elements** or **Chromatin boundary elements** are cis-acting regulatory DNA signals which protect genes from the effects of the neighbouring heterochromatin.
- **Boundary elements** act by establishing barriers for heterochromatin spreading and are sufficient to protect a reporter gene from transcriptional silencing when inserted between the silencer and the reporter gene.

Regulatory Sequences Beyond The Core Promoter

- **Insulator** is a DNA sequence 40 bp or more that contributes to limitation of chromatin activation to distinct segments.
- An **insulator** is located between the enhancer(s) and the promoter, or between the silencer(s) and the promoter of adjacent genes or clusters of adjacent genes.

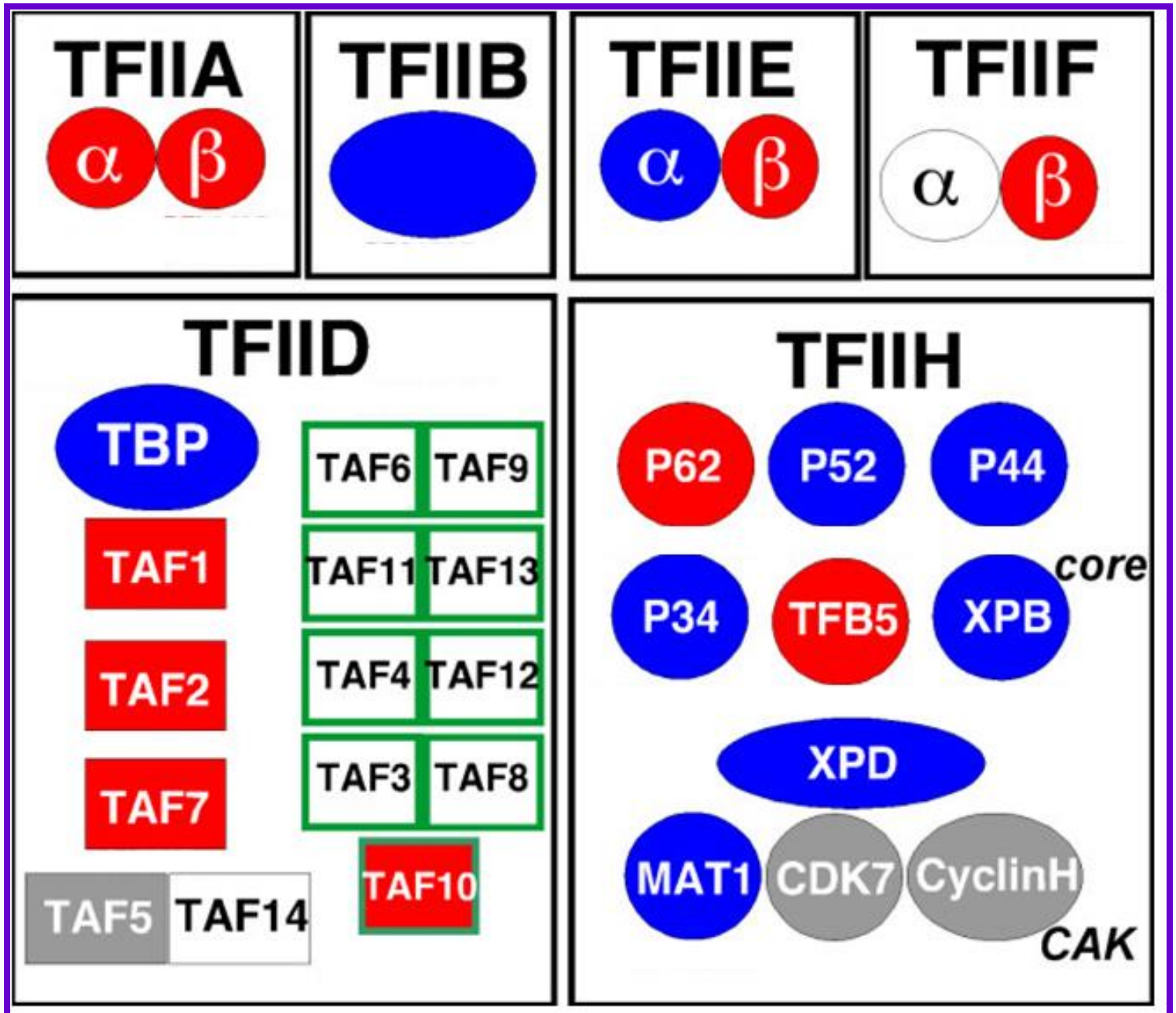
General Transcription factors of RNA Polymerase II

- A series of proteins required for basal transcription initiation from RNA polymerase II promoter sequences *in vitro* - TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIJ & TFIIH. These multi subunit proteins or factors are called transcription factors or *general transcription factors* (GTFs).
- However the DNA template in eukaryotic cells is incorporated into nucleosomes, hence additional factors are required, including the so called Mediator Complex, DNA binding regulatory proteins, and often chromatin-modifying enzymes.

General Transcription factors of RNA Polymerase II

- The general transcription factors collectively perform the functions of **σ** in bacterial transcription.
- The complete set of general transcription factors and polymerase, bound together at the promoter and poised for initiation, is called the **Pre-Initiation (PI) Complex**.

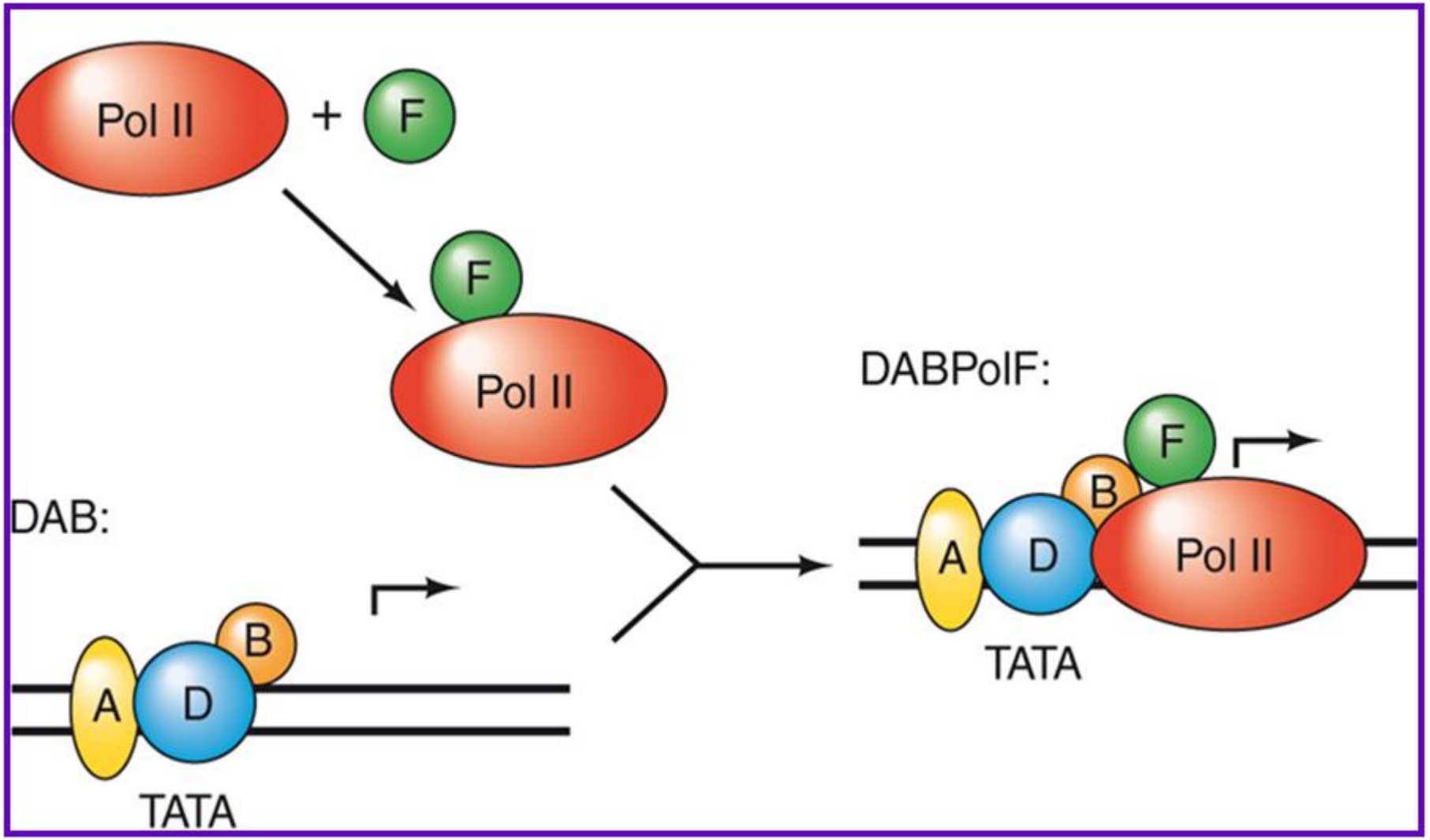
Transcription Factors of RNA Pol II



Transcription Factors of RNA Pol II

Factor	No. of subunits	Molecular mass (kDa)	Functions	Functions to Recruit:
TFIID: **TBP**	1	38	Recognize core promoter (TATA)	TFIIB
TFIID: **TAFs**	12	15-250	Recognize core promoter (non-TATA); Positive and negative regulation	RNA Pol II?
TFIIA	2	12, 19, 35	Stabilize TBP-DNA binding; Anti-repression	
TFIIB	1	35	Select start site for RNA Pol II	RNA PolII-TFIIF
RNA Pol II	12	10-220	Catalyze RNA synthesis	TFIIE
TFIIF	2	30, 74	Target RNA PolII to promoter; destabilize non-specific interactions between PolII and DNA	
TFIIE	2	34, 57	Modulate TFIIH helicase, ATPase and kinase activities; Directly enhance promoter melting?	TFIIH
TFIIH	9	35-89	Helicase to melt promoter; CTD kinase; promoter clearance?	
TFIIJ			Not yet understood	

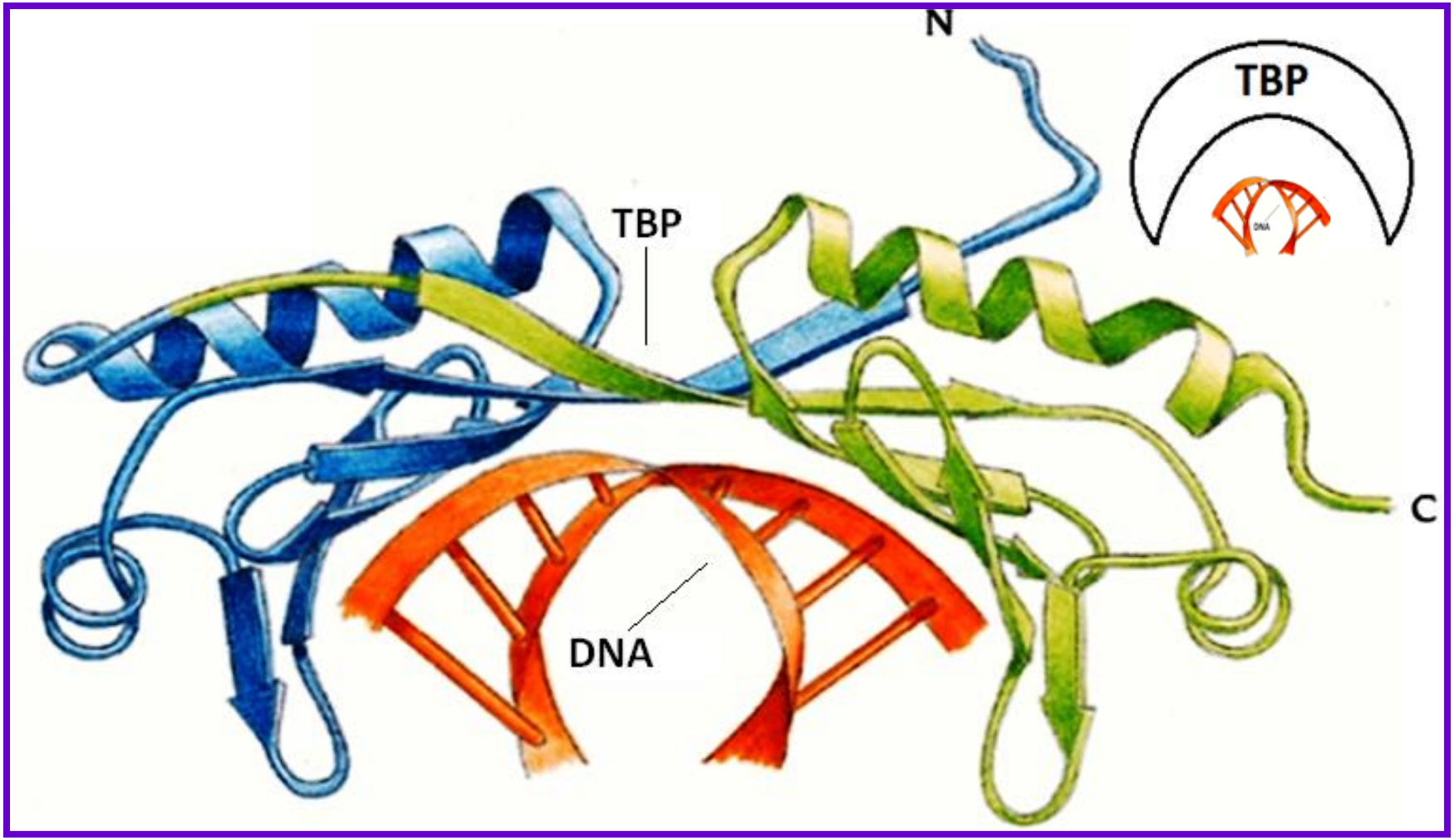
Model for Assembly of **Preinitiation Complex**



TFIID

- **TFIID** is a principal transcription factor in TATA box containing promoters. It has a molecular weight of ~800 kD.
- **TFIID** is a multiprotein complex consisting of TBP (TATA Binding Protein, which binds to TATA box) and atleast 14 other polypeptides known as TAFs (TBP associated factors).
- Some TAFs seem to be necessary for transcription initiation from promoters, lacking the TATA box.
- Other TAFs are tissue-specific coactivators.
- TAF subunits also interact with other GTFs therefore stabilyzing the complex.
- **TBP** (TATA Box Binding Protein) is present in all the 3 eukaryotic transcription complexes (in SL1 – RNA pol I; TFIID – RNA pol II and TFIIIB- RNA pol III.

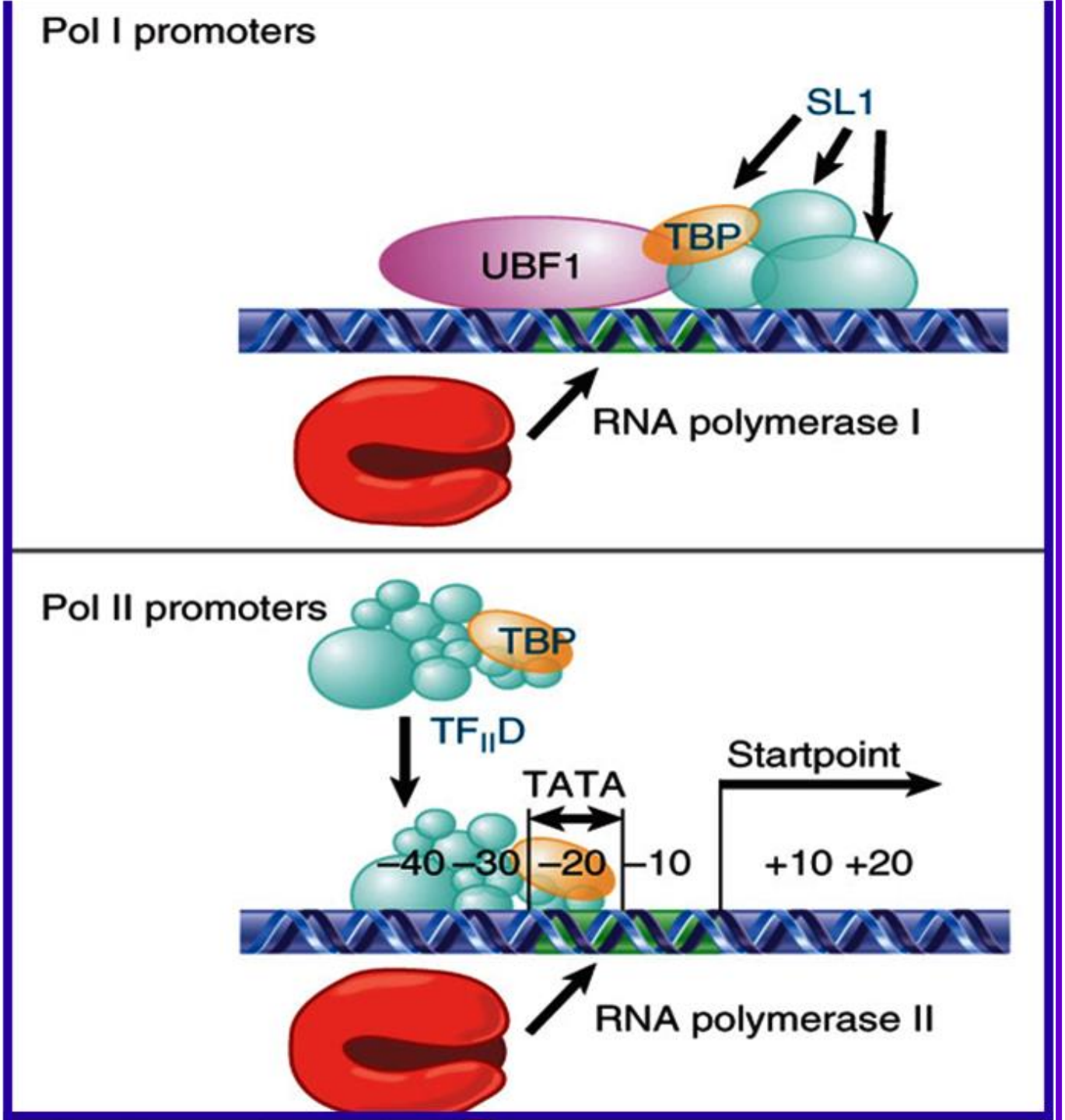
TBP



- Horse shoe shaped TBP (Monomeric Protein) is part of the transcription factors of all eukaryotic RNA polymerases

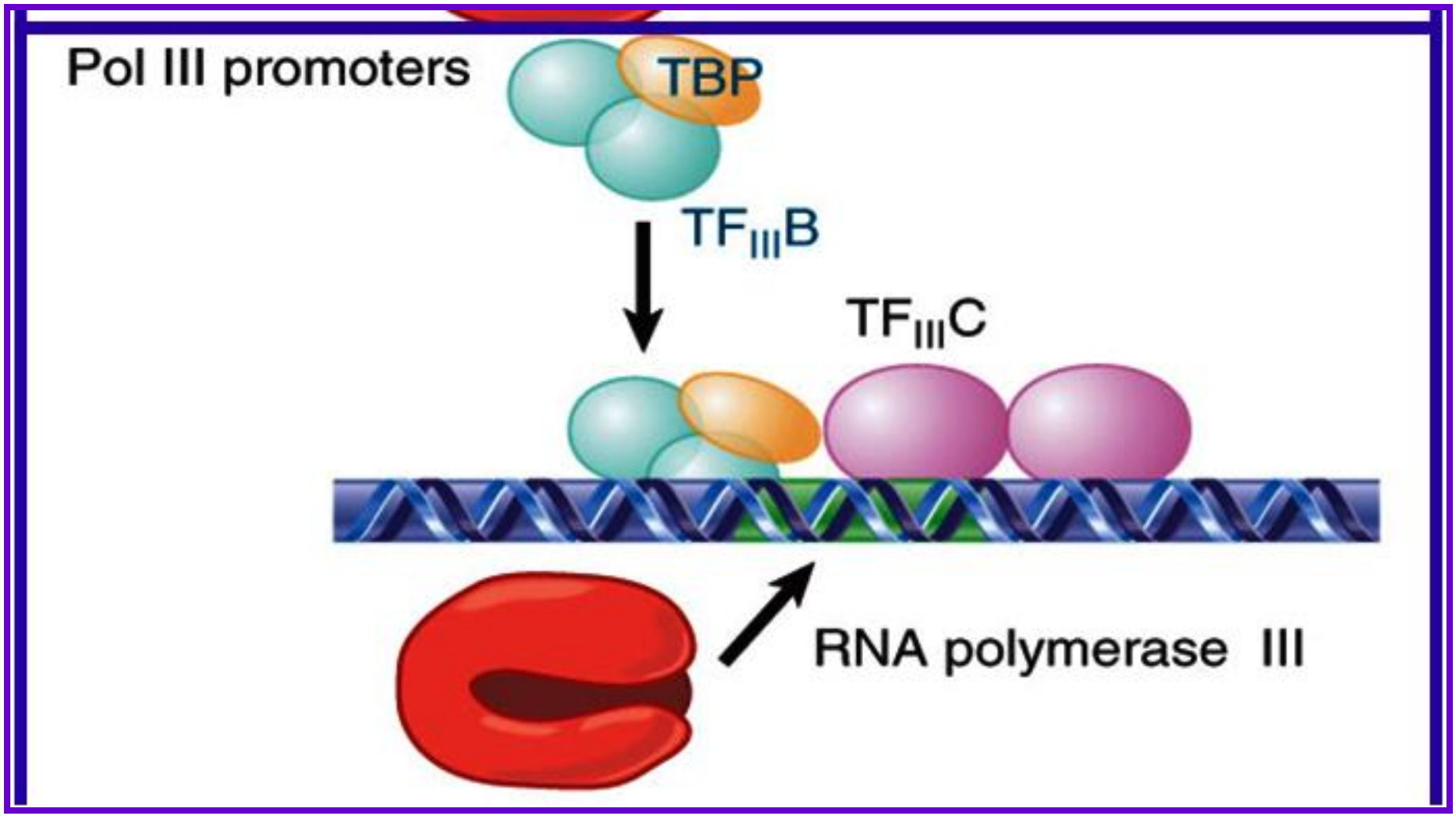
TFIID

- **TBP** is a saddle shaped Monomeric protein with overall dyad symmetry. The 2 halves are not identical.
- In case of RNA Pol II promoters TBP interacts with the minor groove of DNA, unwinds it and bends DNA.
- TATA box makes a pseudo-twofold sequence-specific interaction with two threonines and two asparagines of TBP.
- TBP bends DNA by stacking phenylalanines against DNA bases.
- The precise role of TBP in case of RNA pol I & pol III transcriptions is not known.



TBP Is a Universal Factor

TBP Is a Universal Factor



TFIIA

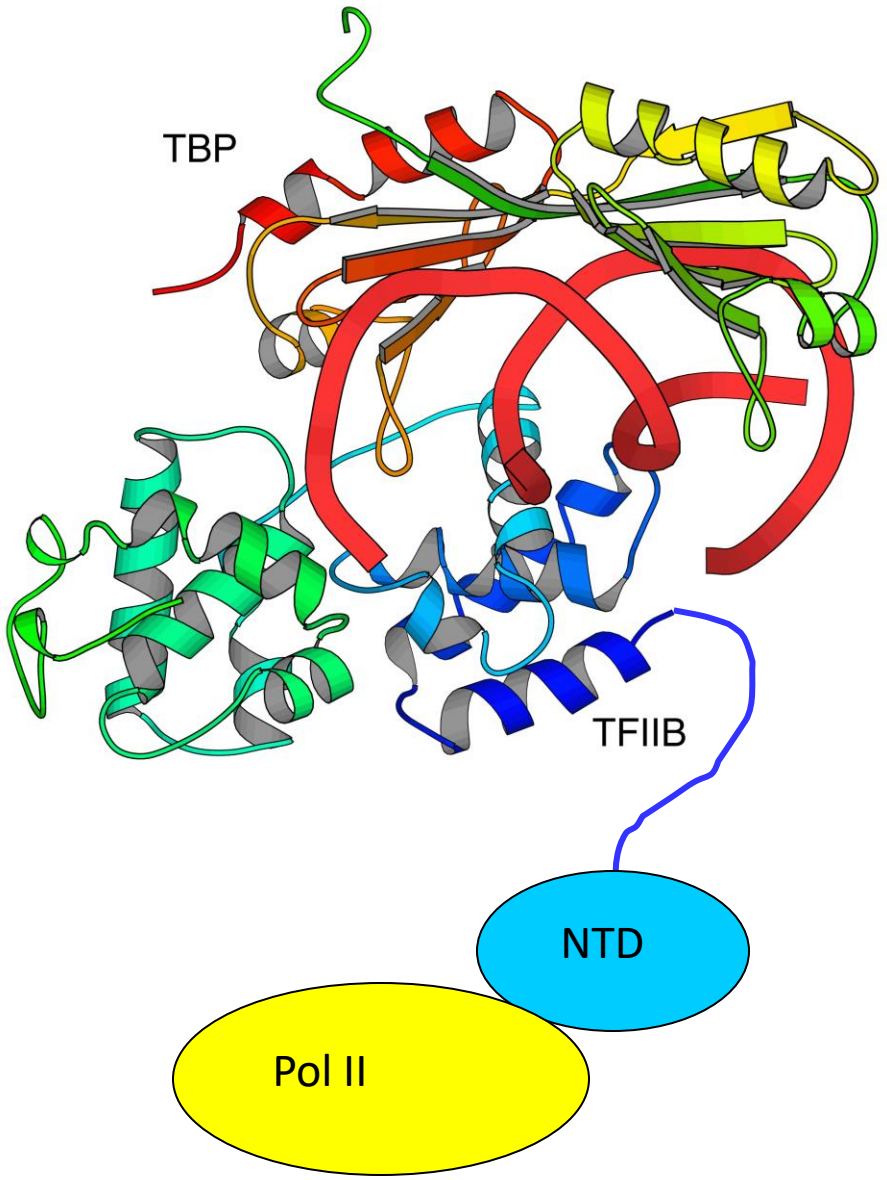
- **TFIIA** – It binds to TFIID and enhances the TFIID binding to TATA box, stabilizing the TFIID-DNA complex.
- DR1 & DR2 are inhibitory factors associated with TFIID.
- The binding of TFIIA to TFIID prevents the binding of inhibitory factors and allows the assembly process to continue.
- *In vitro* studies revealed that purified TFIID does not require TFIIA.

TFIIB

- **TFIIB** is a single polypeptide of 35 kDa.
- **TFIIB** is essential for RNA polymerase II binding.
- **TFIIB** has two domains. N-terminal domain binds RNA polymerase C-terminal domain contacts TBP and bent DNA.
- **TFIIB** is a bridging factor between TFIID and RNA polymerase.
- **TFIIB** binds to TFIID and then recruits RNA polymerase which is already bound to TFIIF.
- **TFIIB** is required before TFIIE and TFIIH are recruited to the pre-initiation complex.

C-Terminal Domain (CTD) of TFIIB

- CTD of TFIIB interacts with both TBP and DNA around the promoter – especially BRE element
- Rough positioning of Pol II is due to interaction of TFIIB CTD with TBP
- Fine positioning is due to interaction with DNA
- N-terminal domain of TFIIB interacts with Pol II



TFIIF Associates tightly with RNA Polymerase

- TFIIF is associated with RNA polymerase II and then recruited on to promoter.
- The binding of TFIIF to RNA pol II stabilizes the DNA-TBP-TFIIB complex.
- TFIIF has two subunits.
- TFIIF functions in start site selection.
- TFIIF binds to the non-template DNA strand.
- TFIIF also reduces non-specific binding of RNA pol II to DNA.

TFIIE

- **TFIIE** is a hetero**tetra**meric protein (a_2b_2) encoded by the GTF2E1 and GTF2E2 genes.
- **TFIIE** contains a zinc ribbon motif that can bind single stranded DNA.
- **TFIIE** promotes DNA **melting** at the promoter.
- **TFIIE** helps in the binding of TFIIH to the complex and also modulates or lowers TFIIH enzymatic activities.

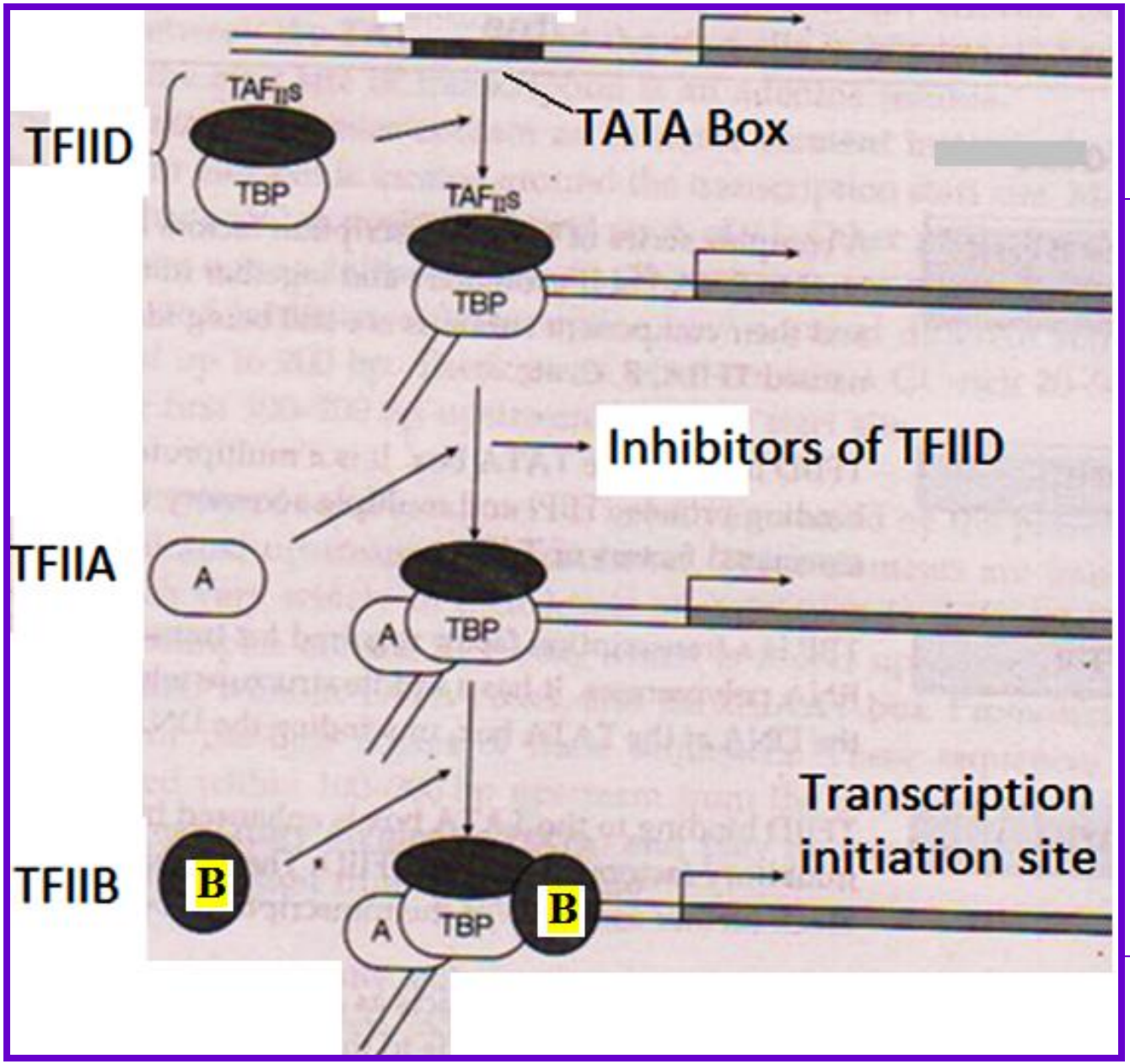
TFIIH

- TFIIH binds to the complex after RNA polymerase is bound.
- TFII H has **9** subunits. Out of them **five** are associated with **helicase** activity which winds DNA at initiation site to create '**transcription bubble**' allowing Pol II to bind to the template strand.
- Rest of the subunits **(4)** of TF IIH are associated with kinase which Phosphorylates the C-terminal domain of RNA polymerase in the begining of elongation.
- TFIIH subunits also recruit DNA-repairing enzymes if polymerase reaches damaged region in DNA and gets stalled.

TFIIJ

- TFIIJ details are not clearly understood.
- But this is associated along with TFIIH.

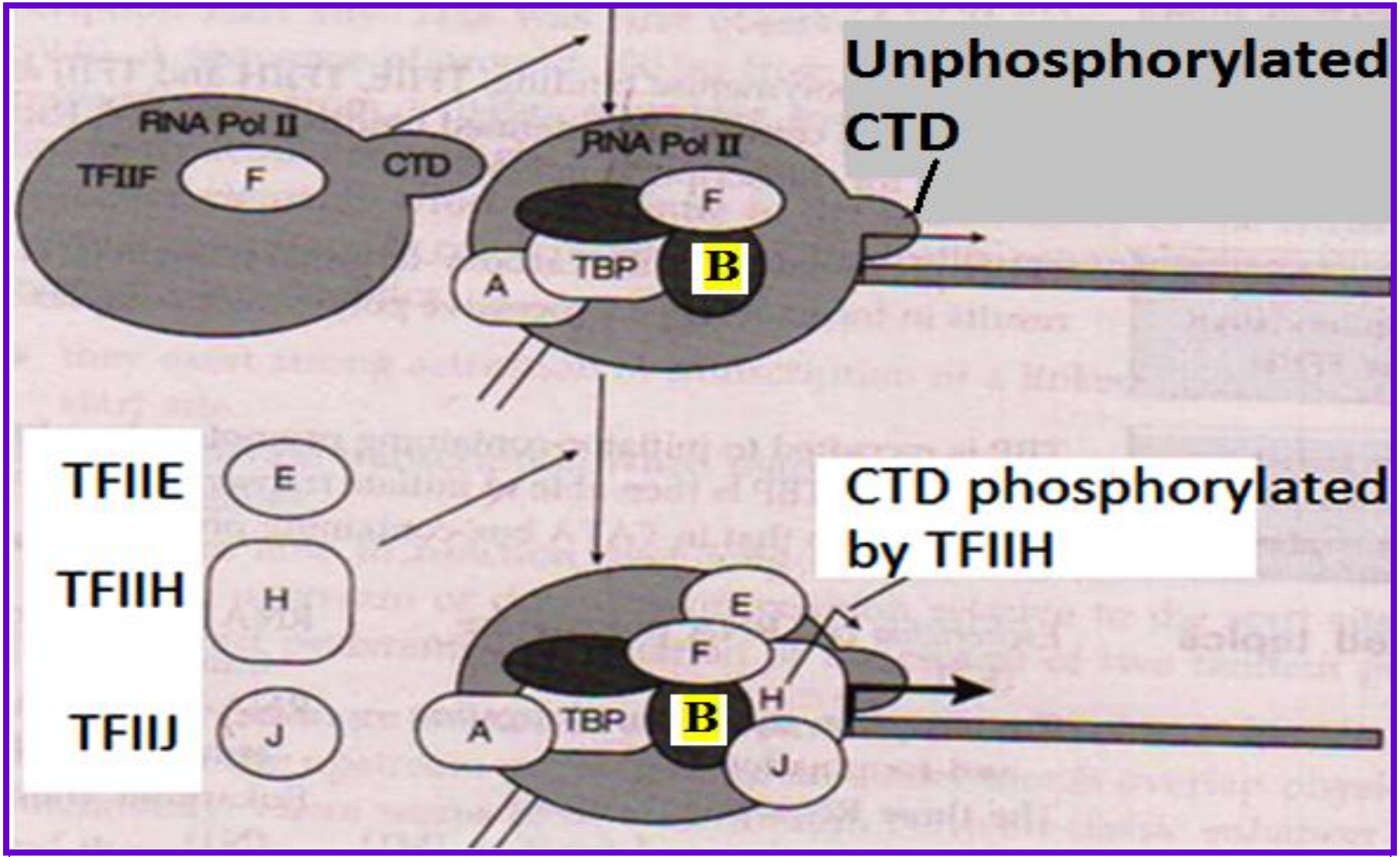
Assembly of RNA pol II Transcription Factors



A schematic diagram of assembly of RNA pol II transcription factors at TATA Box

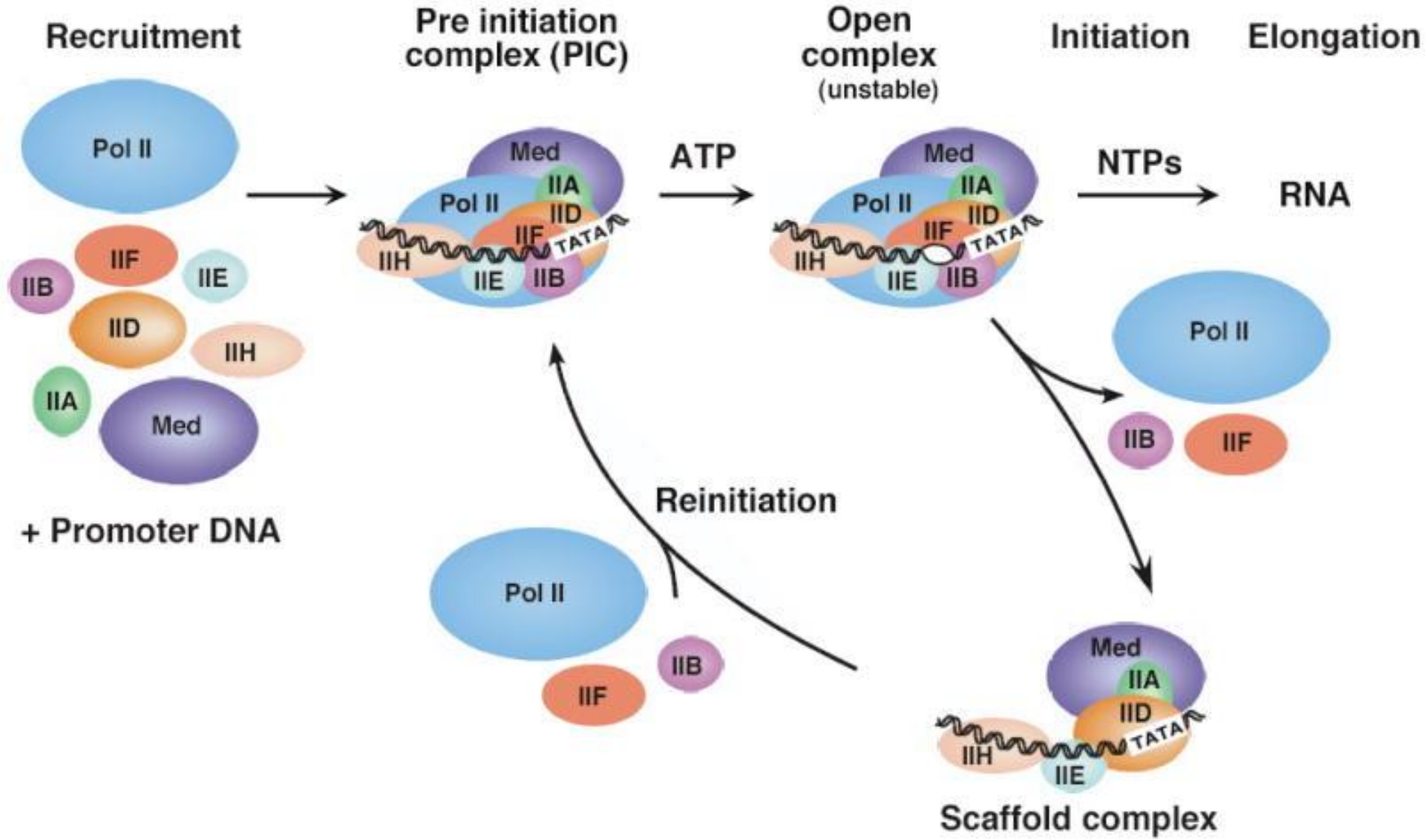
1

Assembly of RNA pol II Transcription Factors



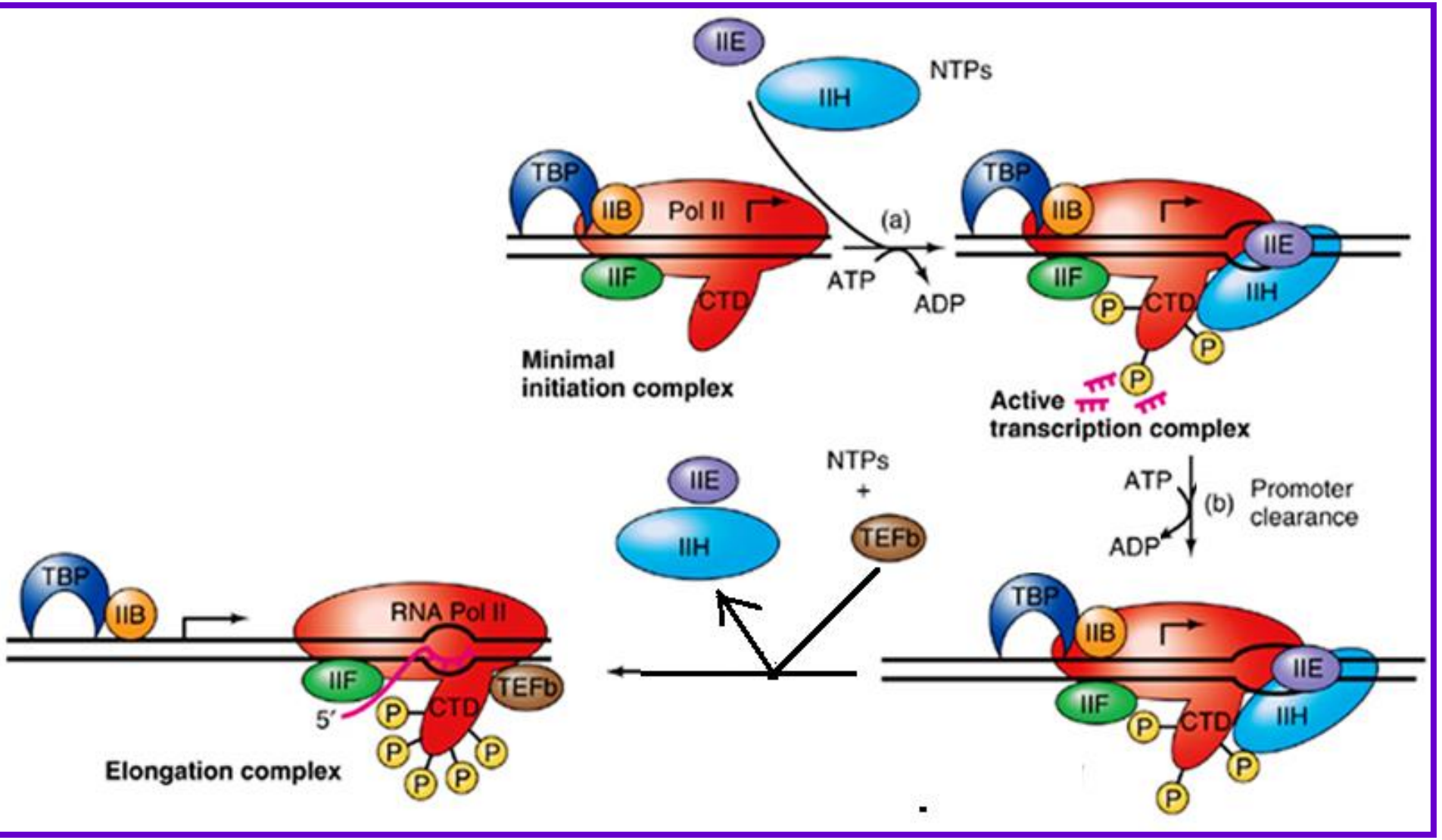
A schematic diagram of assembly of RNA pol II transcription factors at TATA Box.

Retention of a ‘Scaffold Complex’ Allows Efficient Re-initiation



Subsequent initiation events can be more efficient because much of initiation complex is already formed

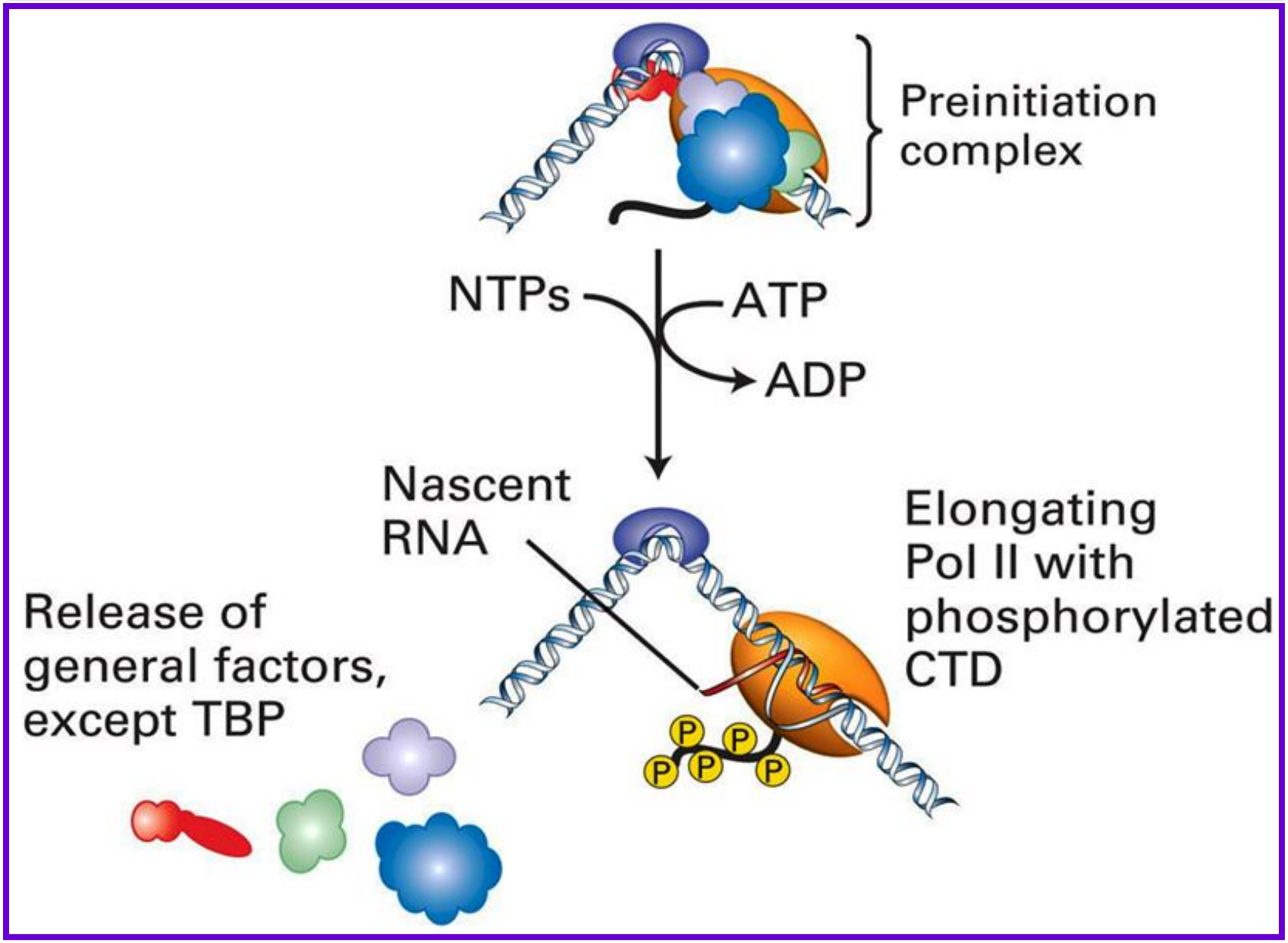
Model for General Transcription Factors: Initiation, Promoter Clearance and Elongation



A New Set of Factors Stimulate Pol II Elongation and RNA Proofreading

- During the transition from initiation to elongation, TFIIH phosphorylates the Pol II CTD, which results in promoter escape or clearance.
- Most of the initiation factors of RNA Pol II are replaced by elongation factors such as P-TEFb, TFIIS & hSPT5 (in yeast).
- The elongation factors stimulate elongation & some of them are also required for RNA processing.
- The enzymes involved in these processes are recruited to C-terminal tail of Pol II's large subunit.

Early events in Elongation

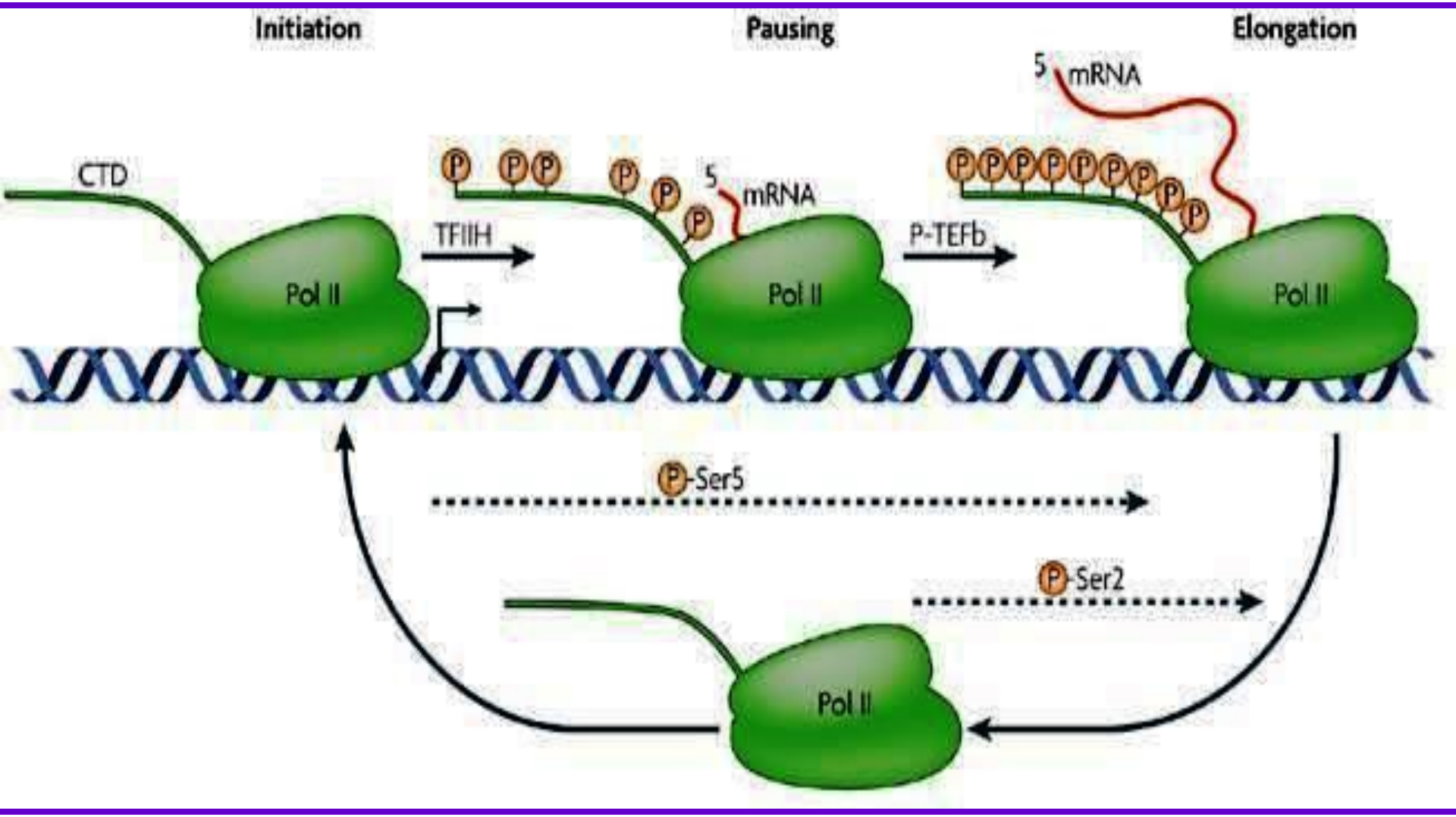


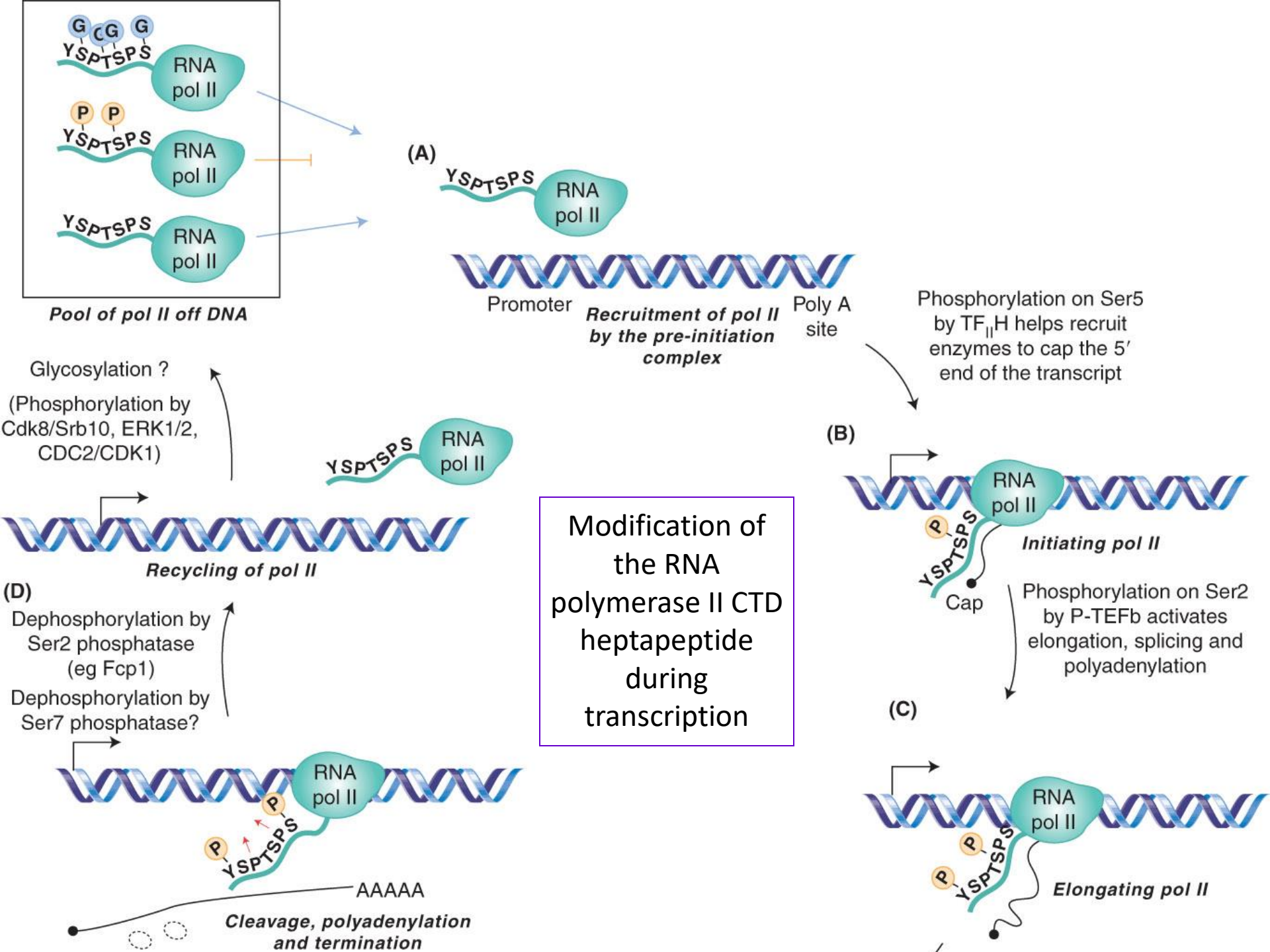
P-TEFb

- **P-TEFb (Transcription Elongation Factor b)** is a kinase that stimulates elongation in three separate ways.

1) P-TEFb phosphorylates **hSPT5** (another elongation factor) and thereby activates it.
2) P-TEFb phosphorylates **RNA Pol II** (serine residue at position 2 of the CTD repeats) which correlates with **splicing** in elongation.
3) P-TEFb recruits another elongation factor - **TAT-SF1**.

RNA ELONGATION



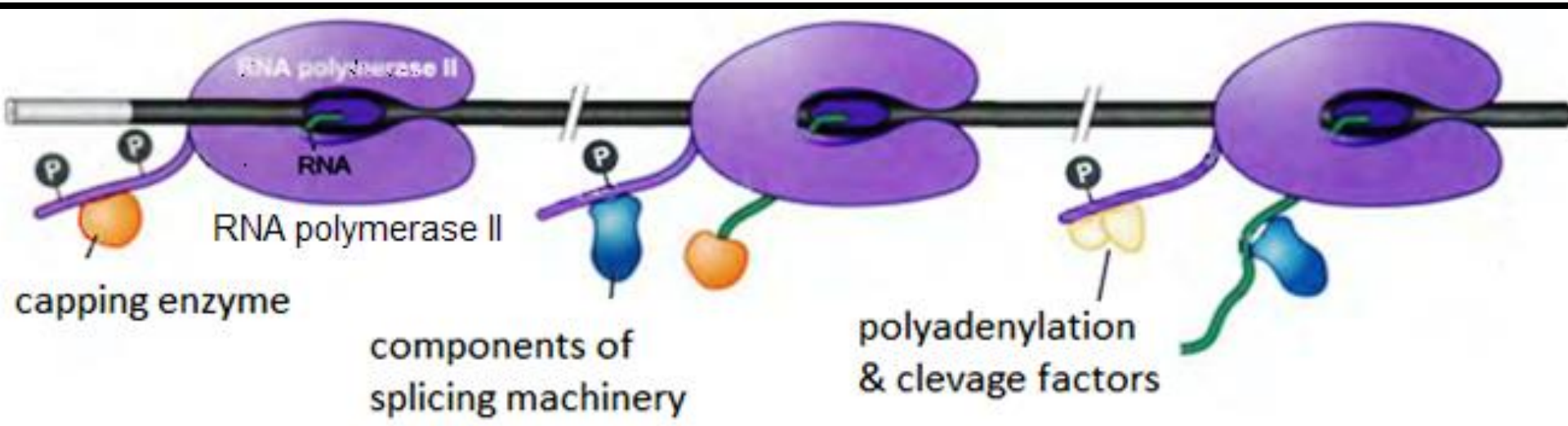

G
C
G
G
YSPTSPS
RNA pol II
P
P
YSPTSPS
RNA pol II
YSPTSPS
RNA pol II
Pool of pol II off DNA
(A)
YSPTSPS
RNA pol II
Promoter
Recruitment of pol II by the pre-initiation complex
Poly A site
Phosphorylation on Ser5 by TF$_{II}$H helps recruit enzymes to cap the 5′ end of the transcript
(B)
RNA pol II
P
YSPTSPS
Cap
Initiating pol II
Phosphorylation on Ser2 by P-TEFb activates elongation, splicing and polyadenylation
(C)
RNA pol II
P
P
YSPTSPS
Elongating pol II
AAAAA
Cleavage, polyadenylation and termination
(D)
Dephosphorylation by Ser2 phosphatase (eg Fcp1)
Dephosphorylation by Ser7 phosphatase?
RNA pol II
P
P
YSPTSPS
Glycosylation ?
(Phosphorylation by Cdk8/Srb10, ERK1/2, CDC2/CDK1)
YSPTSPS
RNA pol II
Recycling of pol II
Modification of the RNA polymerase II CTD heptapeptide during transcription

TFIIS

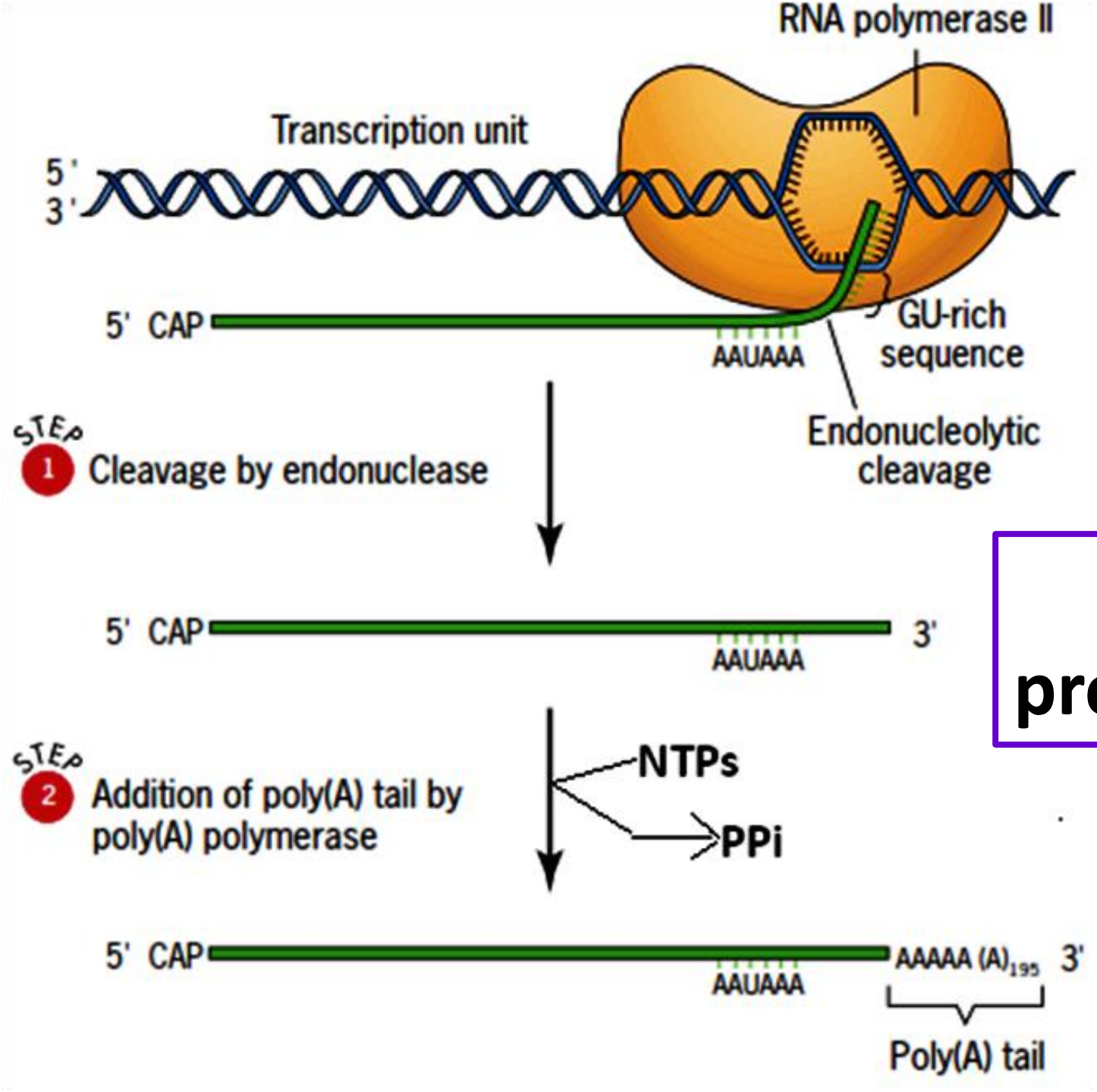
- RNA polymerase does not transcribe all sequences at a **constant** rate. It **slows** down while transcribing certain DNA sequences.
- The elongation factor **TFIIS reduces** polymerase **pausing** time when it encounters such sequences.
- TFIIS also **removes** misincorporated bases by performing the reverse reaction to nucleotide incorporation.
- In addition, TFIIS stimulates an inherent RNAse activity in polymerase (not part of the active site), allowing an alternative approach to remove misincorporated bases through local limited RNA degradation.

- Once transcribed, eukaryotic RNA is processed in various ways which also includes termination of RNA.
  1) Capping of the 5' end of the RNA;
  2) Splicing; and
  3) Polyadenylation of the 3' end of the RNA.
- The elongation factor **hSPT5** recruits and stimulates the 5' **capping** enzyme machinery.
- The elongation factor TAT-SF1 recruits components of the splicing machinery.

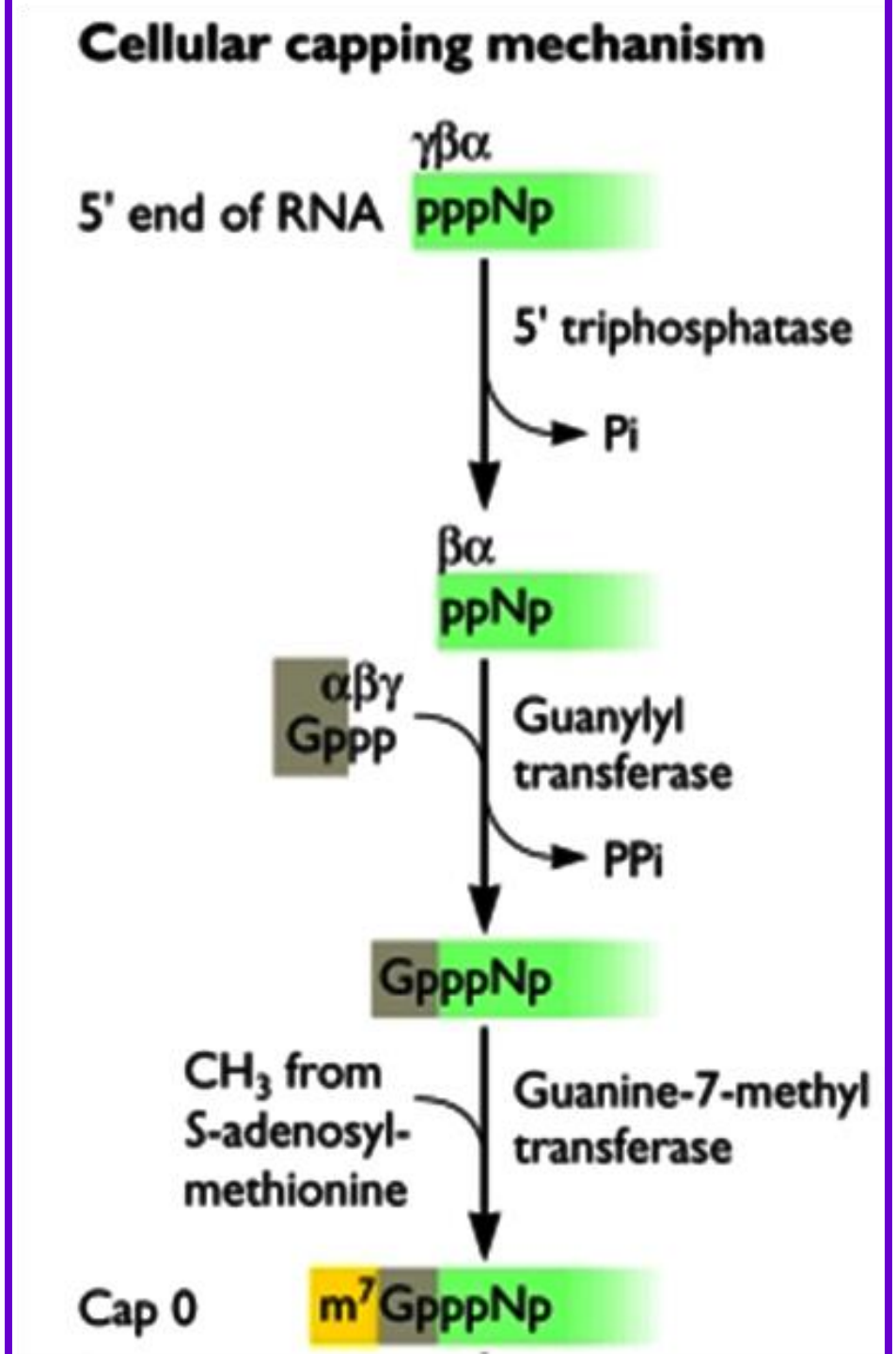
RNA processing enzymes are recruited by the tail of RNA Polymerase



Various enzymes involved in RNA processing recruited by the "tail" of polymerase

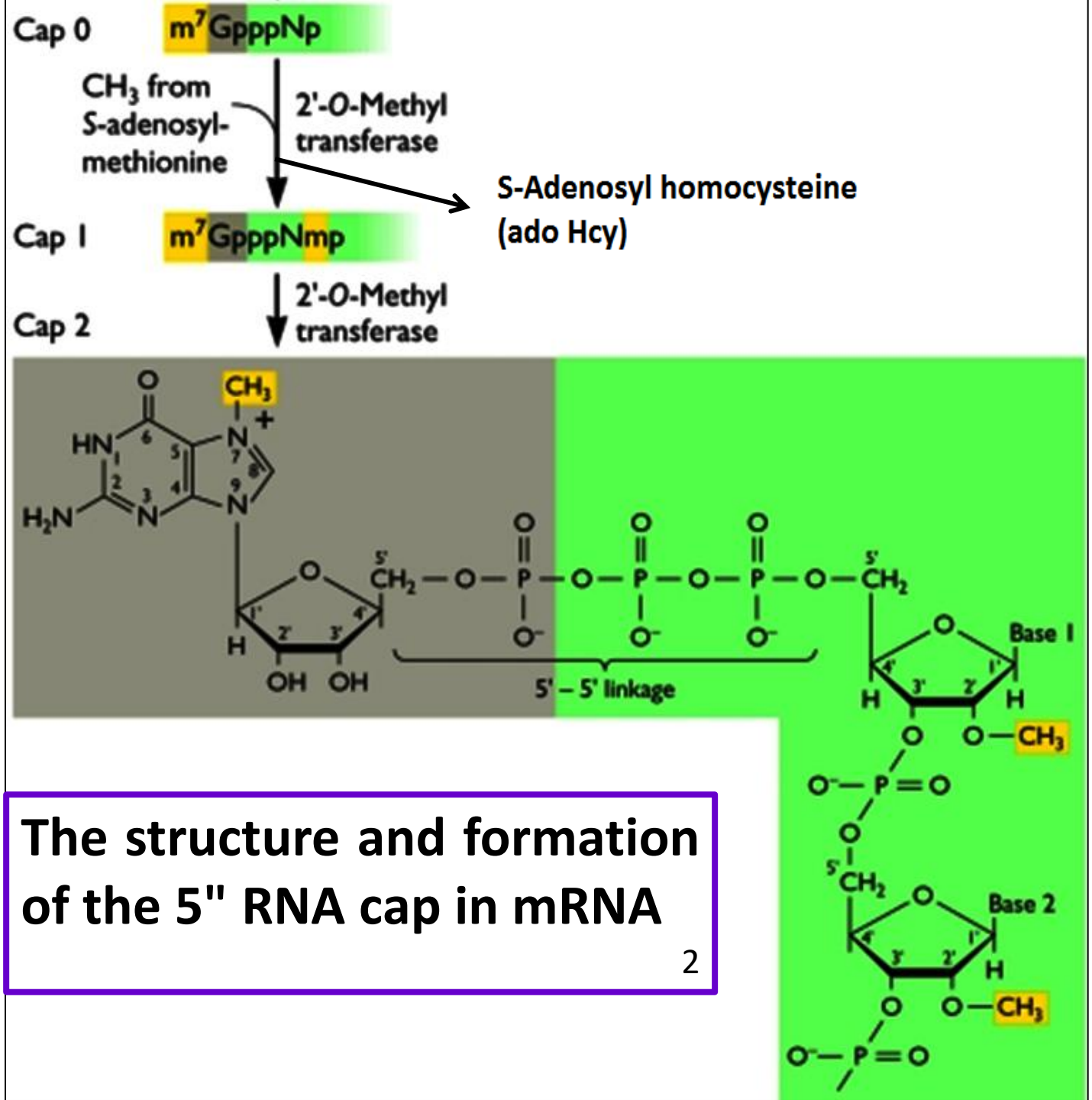


RNA processing

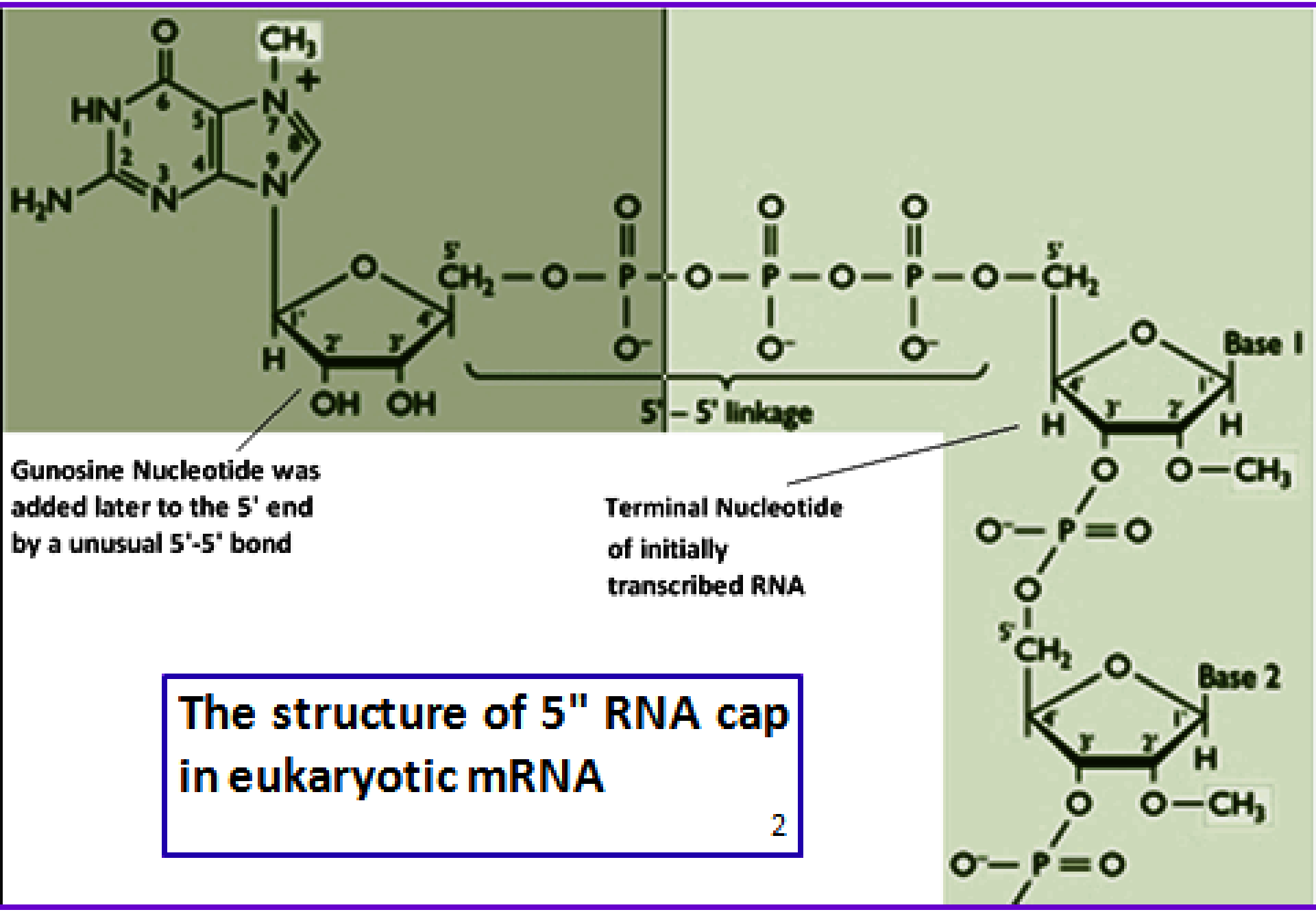


The structure and formation of the 5" RNA cap in **mRNA**

1



The structure and formation of the 5" RNA cap in mRNA



The structure of 5" RNA cap in eukaryotic mRNA

Capping

- The first RNA processing event, capping is created in three enzymatic steps.
  1) Phosphate group is removed from 5' end of transcript.
  2) GTP is added to 5' end of transcript by an unusual 5'-5' linkage involving three phosphates.
  3) A methyl group is added to the Guanosine nucleotide. The 2 & 3 nucleotides also may be methylated.

Capping

- The RNA is capped only after some 20-40 nucleotides are synthesized and transcription cycle has progressed to the transition between the initiation and elongation phases.
- After capping, **dephosphorylation** of Ser5 within the tail repeats leads to dissociation of the capping machinery, and further phosphorylation (Ser2 of tail repeats) causes recruitment of the machinery needed for **RNA splicing**.

Splicing

- Splicing is the next RNA processing.
- Splicing is the joining of exons by the removal of intervening intron sequences in a hnRNA (primary transcript) resulting in mature mRNA which is exported to cytoplasm through nuclear pores & undergoes translation.
- Note: Splicing literally means ‘’**joining”**.

Polyadenylation

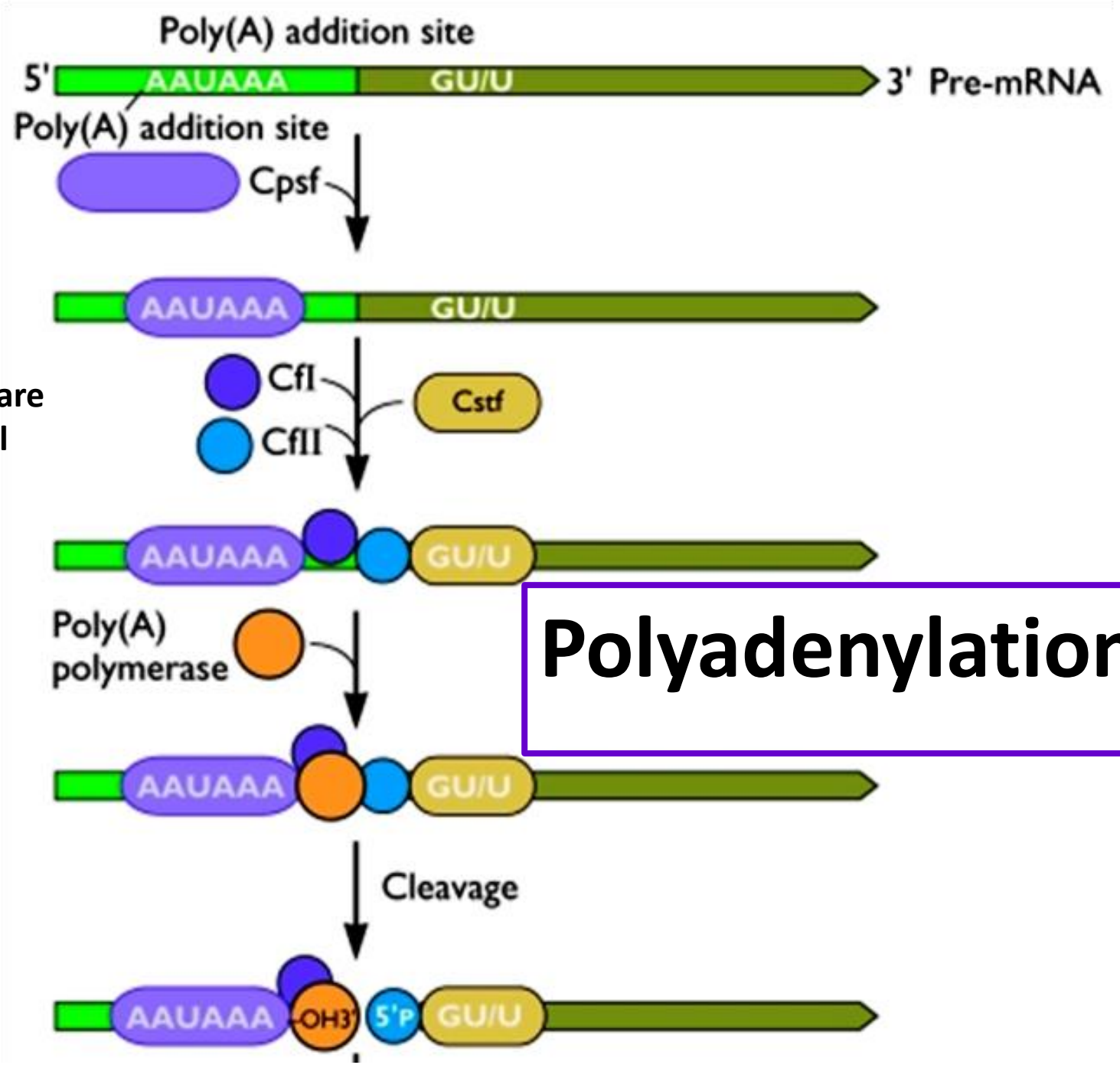
- The final RNA processing event is polyadenylation of the 3' end of the mRNA.
- Specific sequences, present at the end of genes are transcribed into RNA and trigger the transfer of the polyadenylation enzymes to that RNA, leading to three events:
  1) Cleavage of the **transcript**;
  2) Addition of many adenine residues to **transcript's** 3' end; and
  3) Subsequent termination of transcription by polymerase.

Polyadenylation - Process

- Two protein complexes are carried by the CTD of polymerase as it approaches the end of the gene: Cpsf (Cleavage and Polyadenylation Specificity Factor) and Cstf (Cleavage Stimulation Factor).
- Poly-A signals present in DNA are transcribed & trigger the transfer of CPSF and CstF to the RNA.
- Once CPSF and CstF are bound to the RNA, other proteins are recruited leading initially to RNA cleavage and then polyadenylation.

Polyadenylation

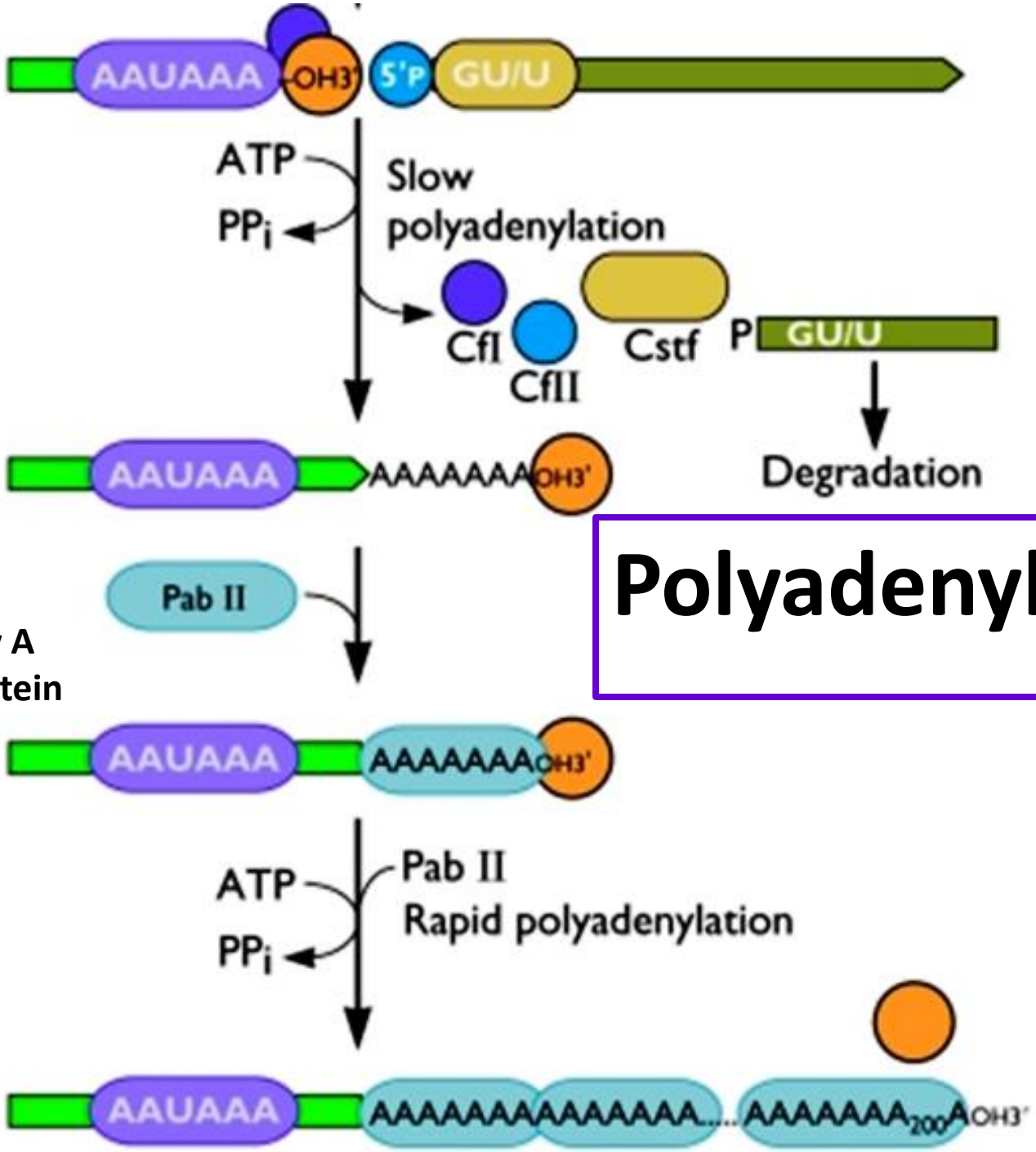
- Polyadenylation is mediated by an enzyme called poly-A polymerase, which adds about 200 adenines to the RNA's 3' end produced by the cleavage.
- This enzyme uses ATP and adds the nucleotides without a template. Thus, a long tail of '**A**'s is found in the RNA but not the DNA.
- It is not clear what determines the length of the poly A tail, but that process involves other proteins that bind specifically to the poly-A sequence.
- The mature mRNA is then transported from the nucleus.
- This poly '**A**' tail is unique to the transcripts made by RNA Pol II.



CfI & CfII are additional clevage factors

Polyadenylation

1



Polyadenylation

2

PabII – Poly A binding Protein

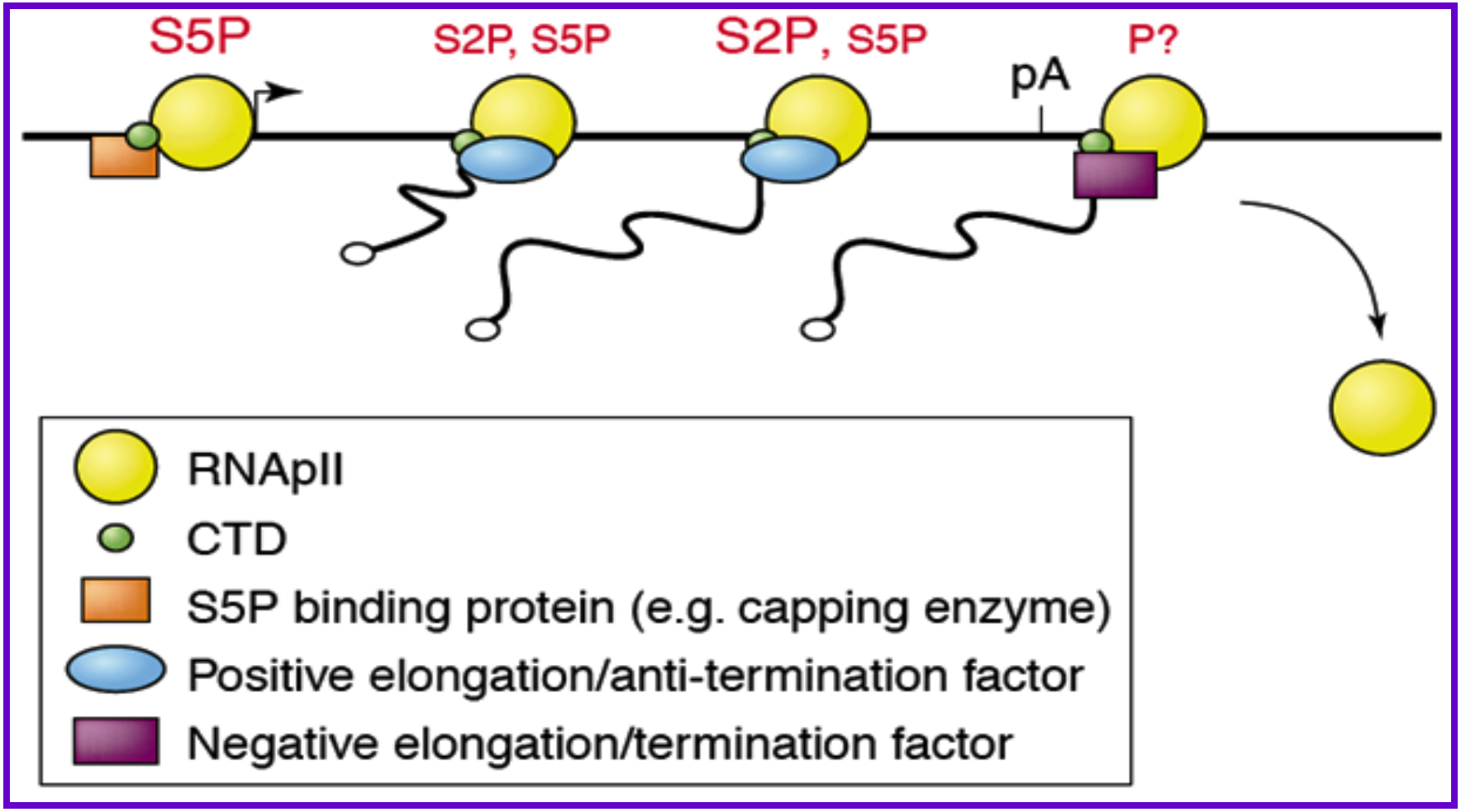
Termination

- RNA polymerase enzyme does not terminate immediately when the RNA is cleaved and polyadenylated.
- Rather, it continues to move along the template, generating a second RNA molecule (several hundred nucleotides) before terminating.
- The polymerase dissociates from the template, releasing the new RNA, which is degraded without ever leaving the nucleus.
- It is not understood what links polyadenylation to termination, but it is clear that the polyadenylation signal is required for termination (interestingly, RNA cleavage is not).

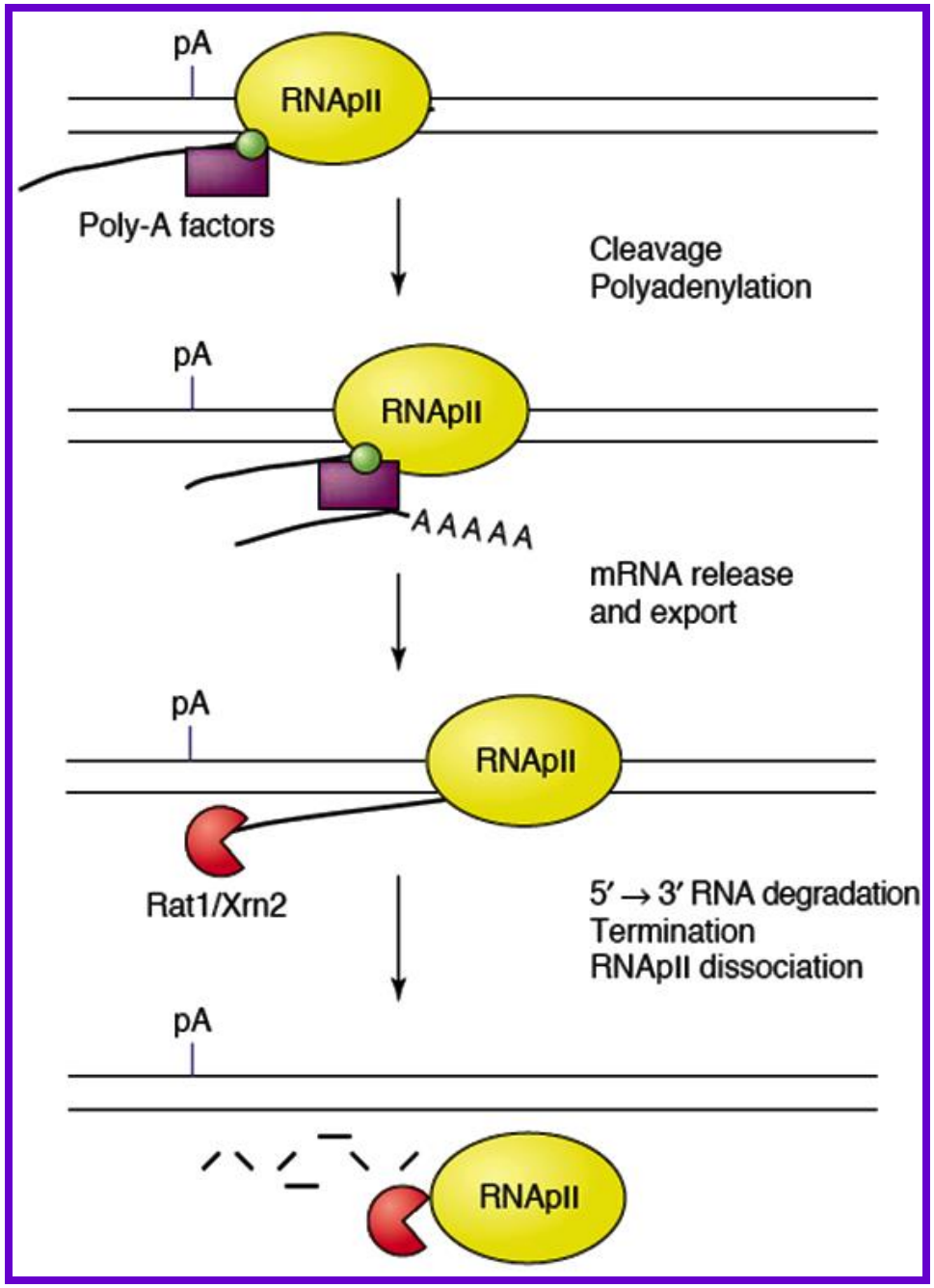
Termination

- Two basic models have been proposed to explain the link between polyadenylation and termination:
- (1) The transfer of 3' processing enzymes from the polymerase CTD tail to the RNA triggers a conformational change in the polymerase that reduces processivity of the enzyme, leading to spontaneous termination soon afterward.
- (2) The second model proposes that the absence of a 5' cap on the second RNA molecule is sensed by the polymerase, which, as a result, recognizes the transcript as improper and terminates.

Transcription Termination Model 1



Transcription Termination Model 1: PolyA signal leads to changes in composition of RNAP



Transcription Termination Model 2

Transcription Termination Model 2: mRNA is cleaved at polyA signal, generating new 5’-end that is rapidly degraded (torpedo model)

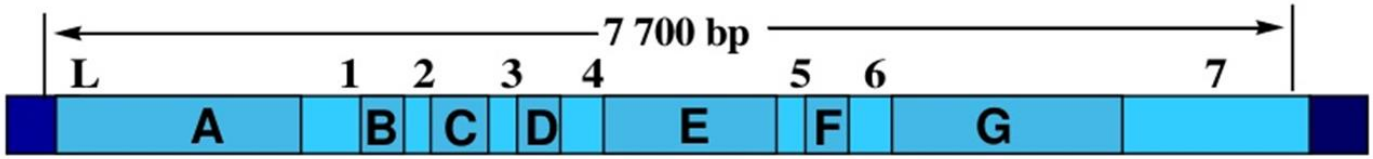
Introns

- An **intron** is a portion of a RNA coded by gene that do not code for amino acids.
- Heterogeneous nuclear RNA consists of coding sequences (exons) and non-coding sequences (introns); introns are removed and exons code for the amino acid product are joined to form mature mRNA.
- Introns – 4 classes

1) **Group I introns** are found in some nuclear, mitochondrial, and chloroplast genes coding for rRNAs, mRNAs, tRNAs and bacteria.
2) **Group II introns** - mitochondrial or chloroplast mRNAs in fungi, algae, plants and bacteria.
3) **Spliceosomal introns** and
4) **Fourth class of introns** are seen in certain tRNAs.

Splicing

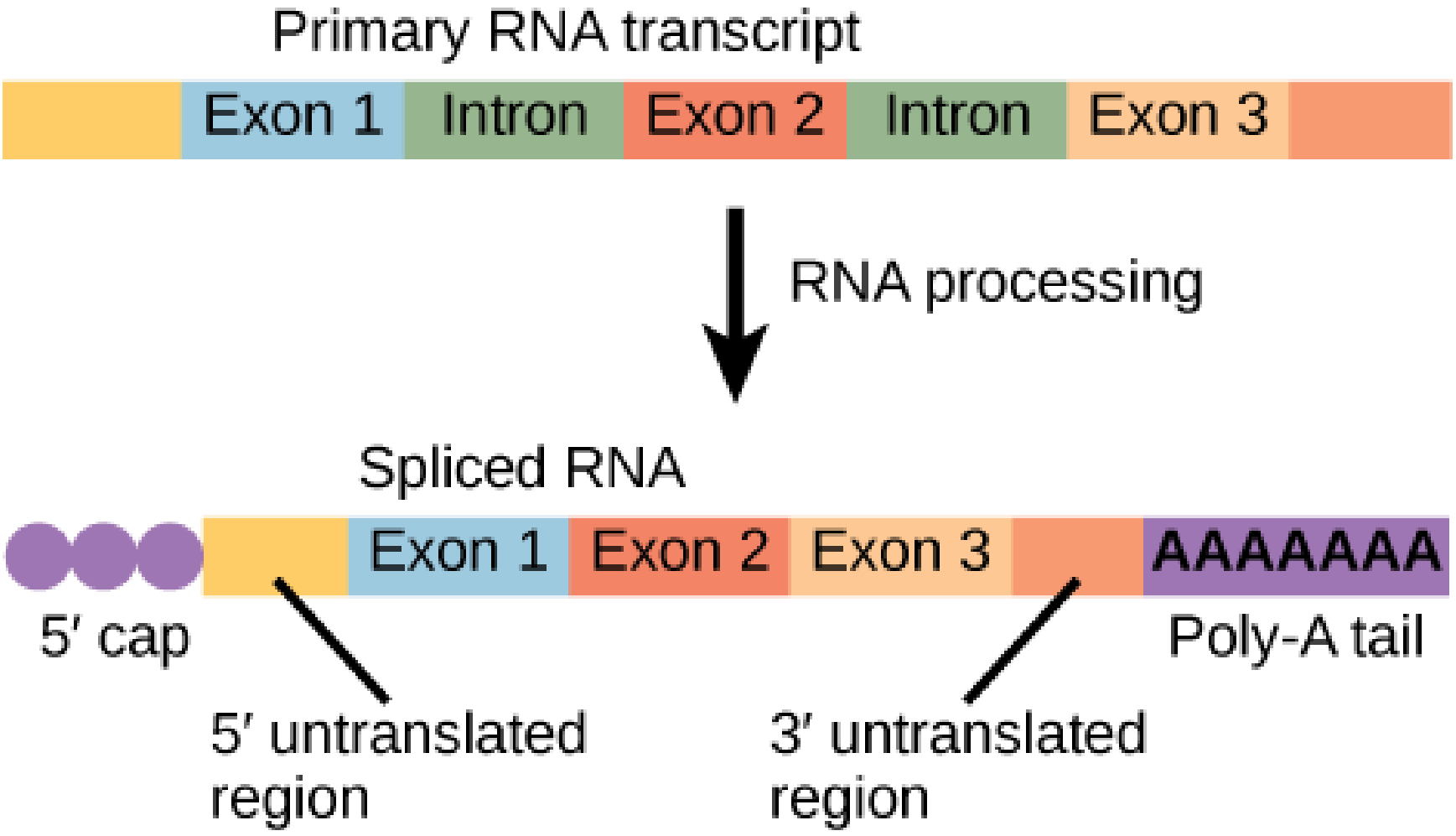
The structural genes are composed of coding and non-coding regions that are alternatively separated.



A-G non-coding region 1-7 coding region

Noncoding sequences called intervening sequences or Introns & coding sequences called Exons

Splicing



Eukaryotic mRNA contains introns that must be spliced out. A 5′ cap and 3′ poly-A tail are also added.

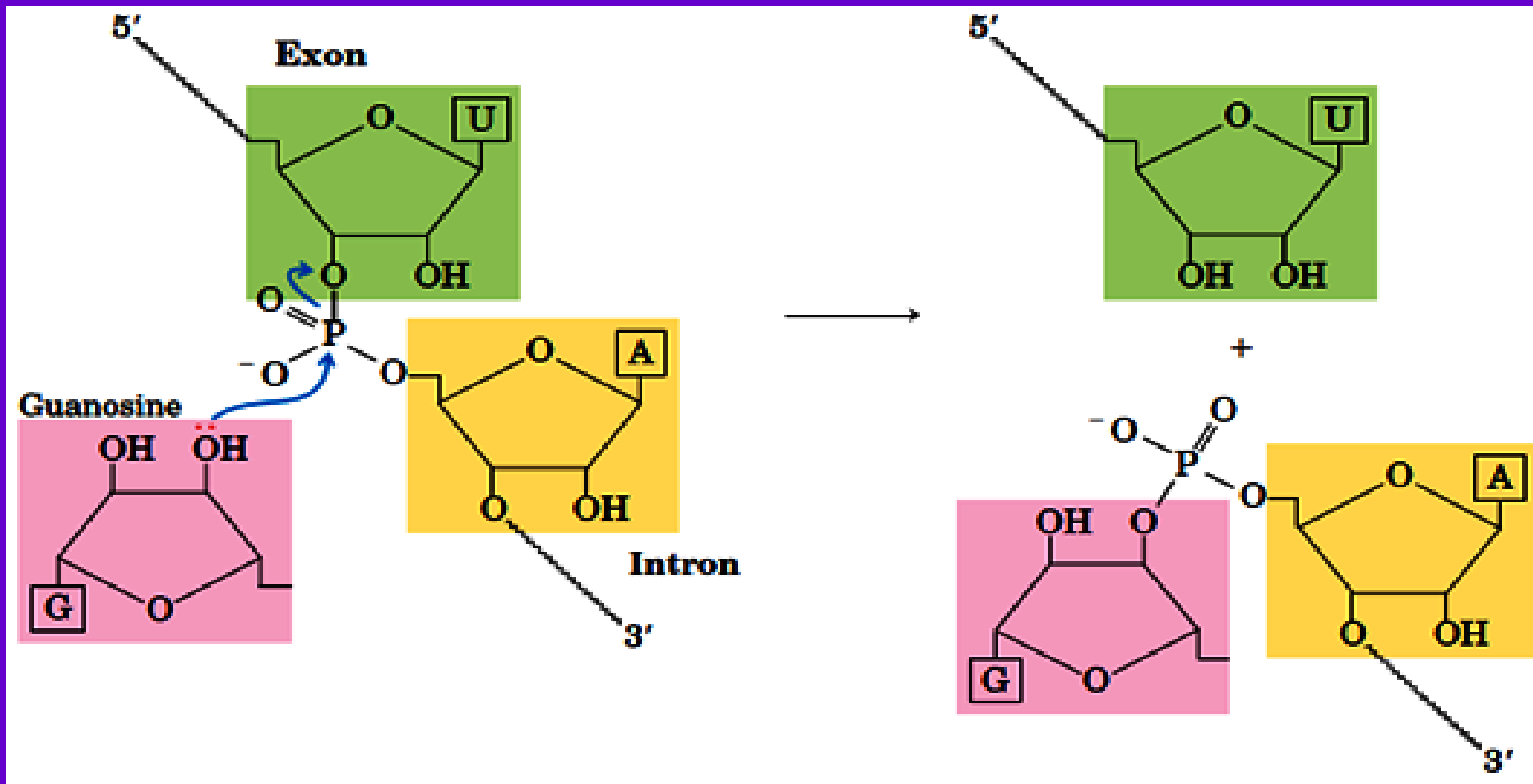
Group I Introns

- Group I & group II introns are *self-splicing*—i.e. no **protein enzymes** are involved and do not require a high energy cofactor (such as ATP) for splicing.
- The splicing mechanisms in both groups involve two transesterification reaction steps as shown in the figure given below.
- A ribose 2'- or 3'- hydroxyl group makes a nucleophilic attack on "phosphorus" and, in each step, a new phosphodiester bond is formed at the expense of the old, maintaining the balance of energy.
- These reactions are very similar to the DNA breaking and rejoining reactions promoted by topoisomerases and site-specific recombinases.

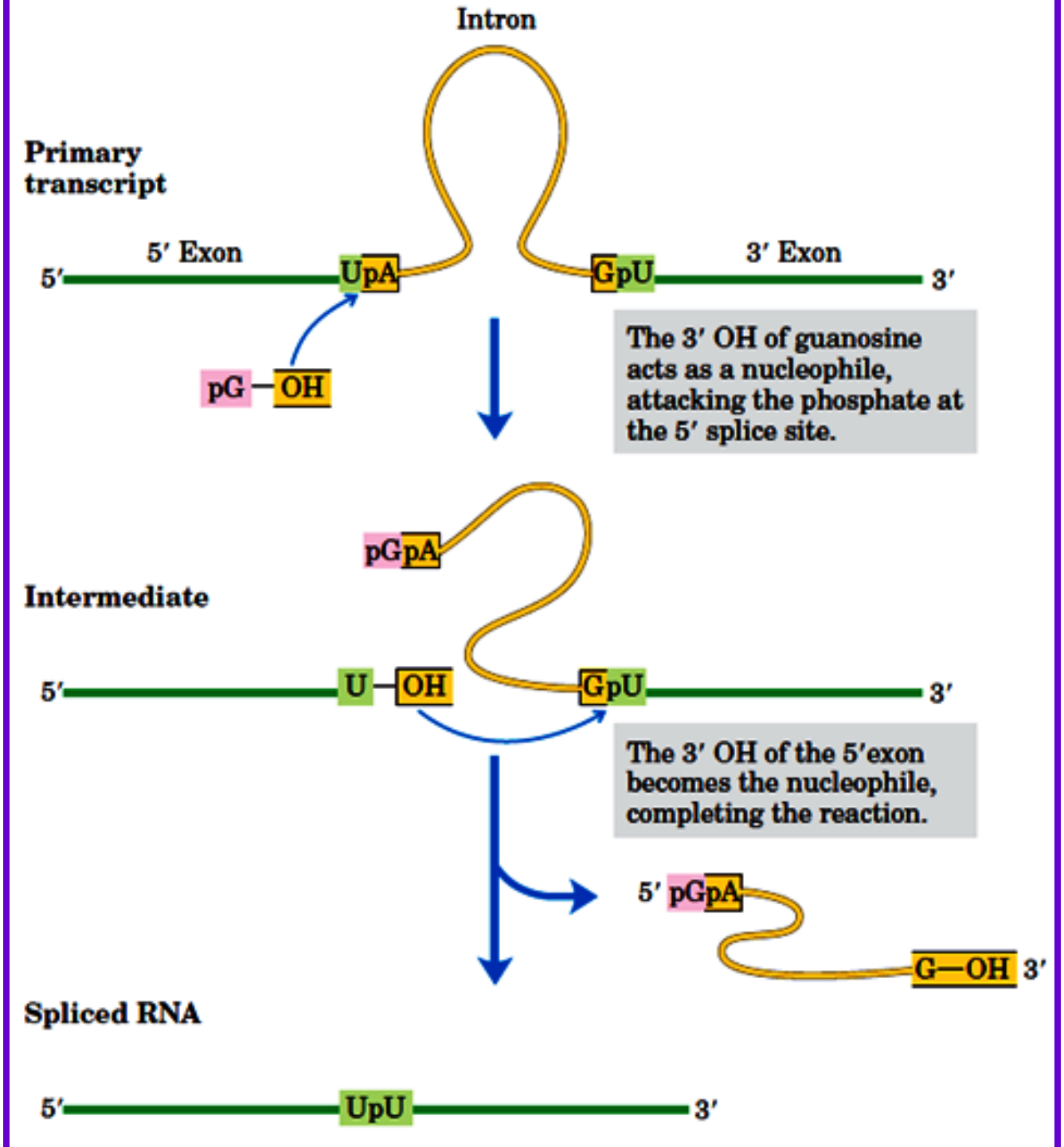
Splicing reaction of Group I Introns

- The splicing reaction of **group I introns** require a guanine nucleoside or nucleotide cofactor.
- The 3’-hydroxyl group of guanosine is used as a nucleophile and forms a normal 3→5 -phosphodiester bond with the 5’ end of the intron.
- The 3’ hydroxyl group of the displaced exon then acts as a nucleophile at the 3’ end of the intron resulting in precise excision of the intron and ligation of the exons.

Transesterification Reaction



This is the first step in the removal of Group I introns.
3’ OH group of Guanosine acts as a nucleophile

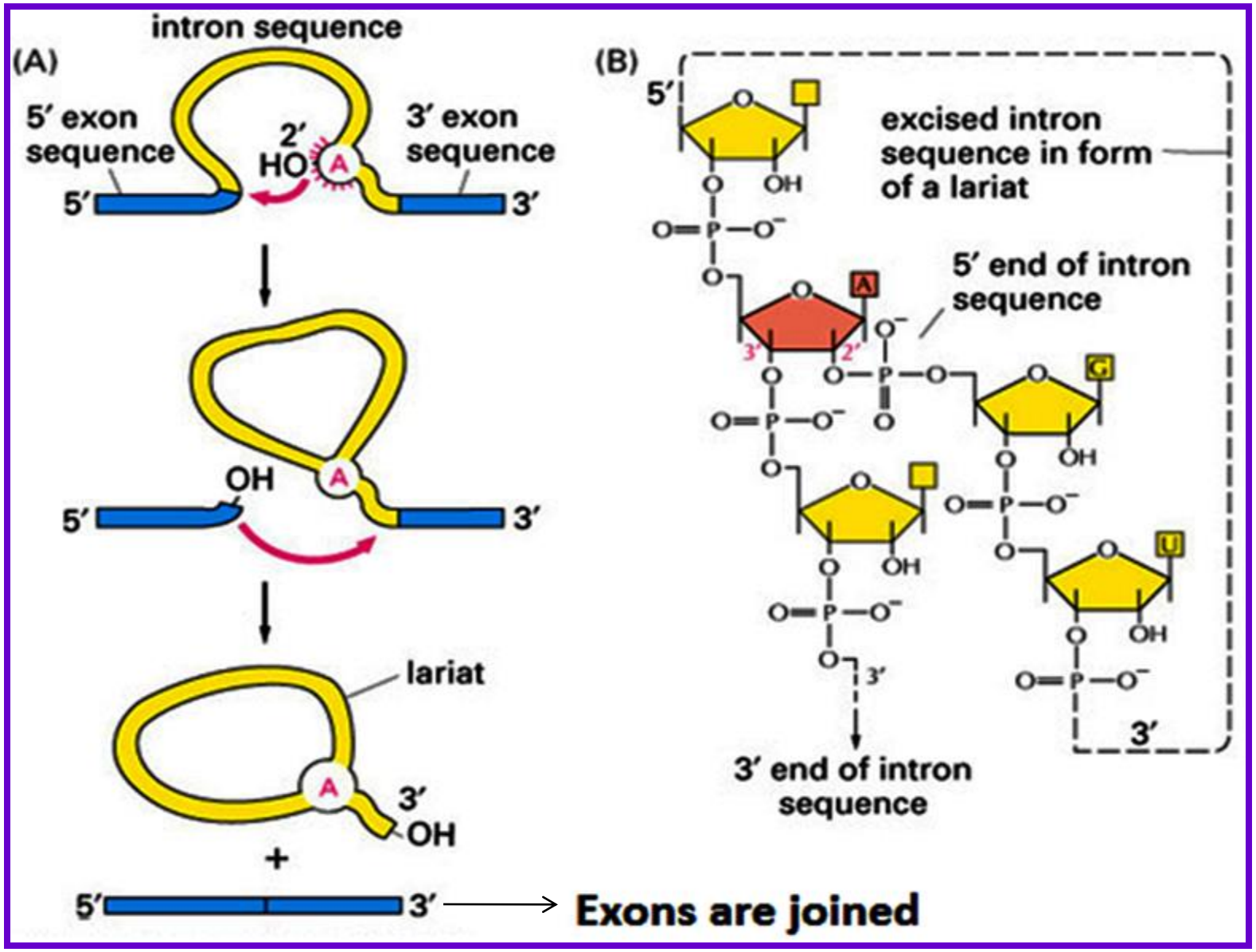


Splicing reaction of Group I Introns

Group II Introns

- **In group II introns** the reaction pattern is similar to group I introns except for the nucleophile in the first step.
- In group II introns the 2'-hydroxyl group of an "**A**" residue *within* the acts as a nucleophile and a branched lariat structure is formed as an intermediate (shown in figure).
- The resulting RNA splices itself accurately without any protein enzymes.

Splicing reaction of Group II Introns



Spliceosomal Introns

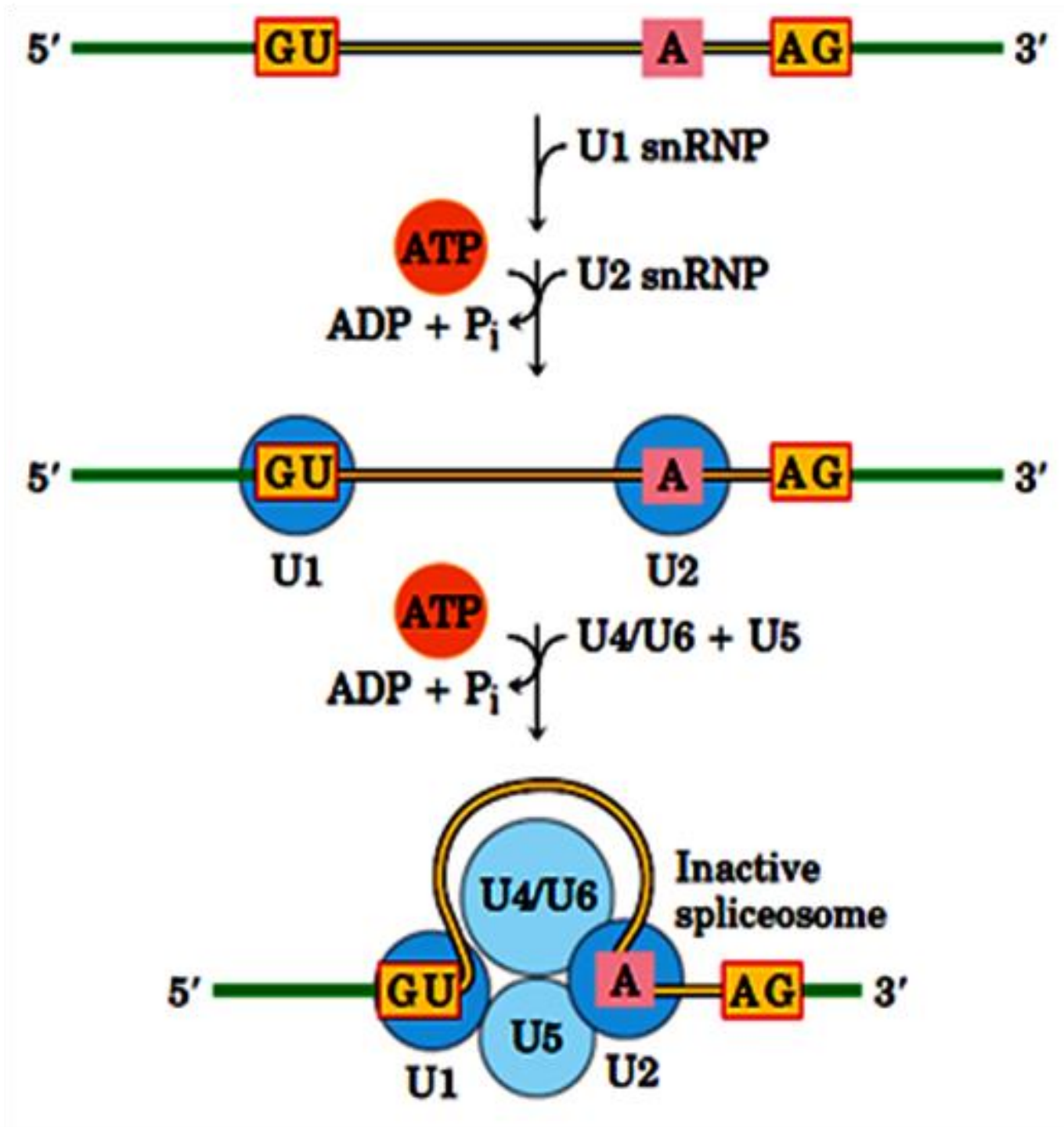
- However most introns are *not* self-splicing, and are not designated with a group number.
- These are called **spliceosomal introns,** because their removal occurs within and is catalyzed by a large protein complex called a **spliceosome.**
- They form the third and largest class of introns and include those found in nuclear mRNA primary transcripts.

Spliceosomal Introns

- Splicing is catalyzed by the spliceosome which is a large RNA-protein complex composed of five small nuclear ribonucleoproteins (snRNPs, pronounced 'snurps').
- Each snRNP contains one of a class of eukaryotic noncoding RNAs, 100 to 600 nucleotides long, known as **small nuclear RNAs (snRNAs).** They bind to proteins to form snRNPs.
- **Major snRNPs – 5 Types -U1, U2, U4, U5, UR6**
- **Other snRNPs- U7, U14, U15**

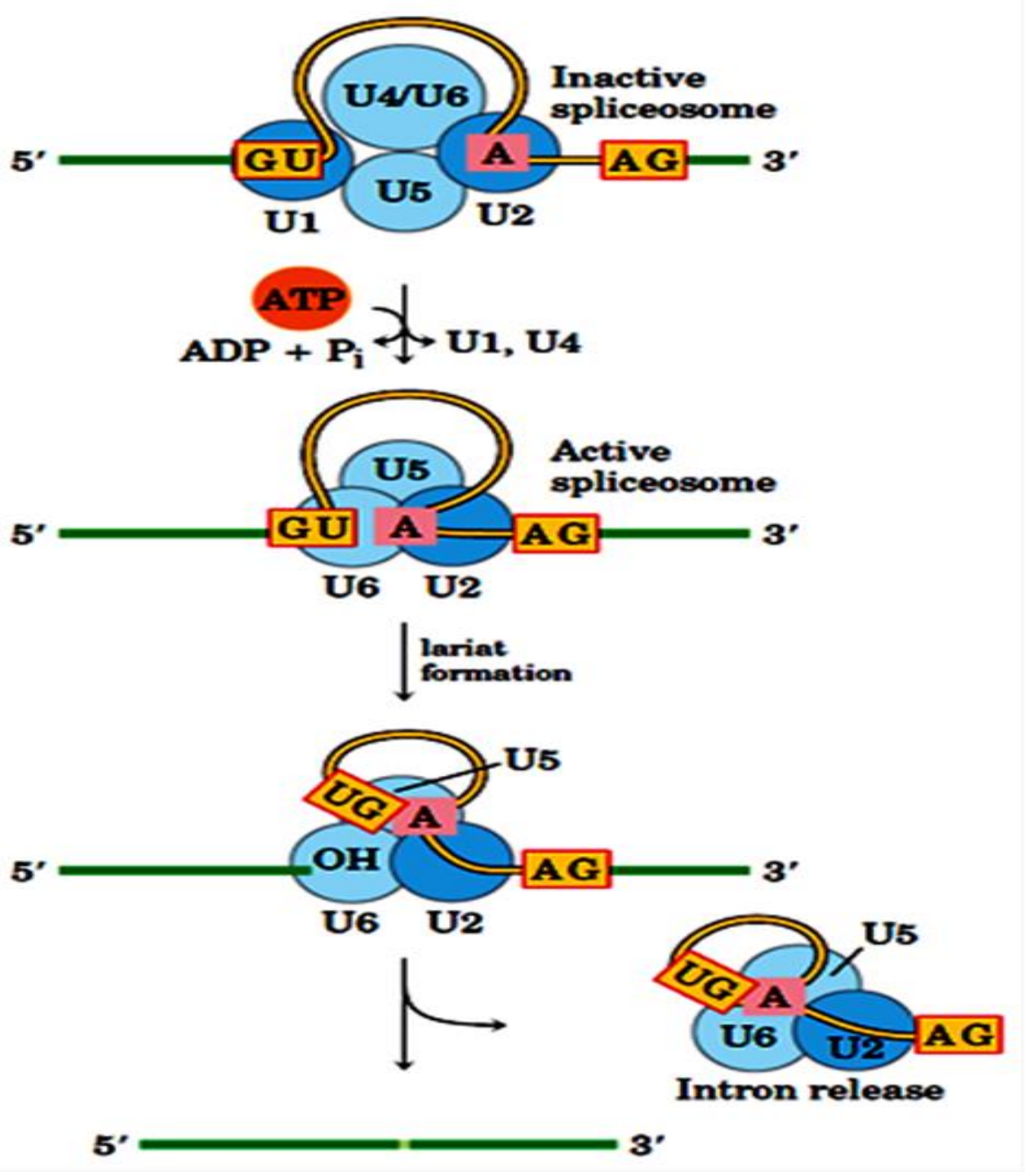
Small nuclear ribonucleoproteins

- Additional snRNPs are functionally diverse, but in many cases the RNA component of snRNPs can base-pair with a substrate for precise alignment and possible catalysis.
- The U7 snRNP directs 3′-end mRNA formation for histone transcripts, and the 7SK snRNP regulates transcription.
- Two special groups of snRNPs, small nucleolar RNPs (snoRNPs) and small Cajal-body RNPs (scaRNPs), are restricted to their named subnuclear compartments in order to direct post-transcriptional modification of ribosomal and splicing RNAs, respectively.
- Certain herpesviruses express high levels of novel snRNPs involved in the regulation of gene expression.
- Due to their important biological roles, there are many diseases associated with snRNPs.



Assembly of Spliceosomes

1 - Formation of an inactive spliceosome



Assembly of Spliceosomes

2 - Formation of an inactive spliceosome

Fourth Class Of Introns

- The **fourth class of introns**, found in certain tRNAs, is distinguished from the group I and II introns in that the splicing reaction requires ATP and an endonuclease.
- The splicing endonuclease cleaves the phosphodiester bonds at both ends of the intron, and the two exons are joined by a mechanism similar to the DNA ligase reaction.
- **Trans-splicing** is a form of splicing that joins two exons that are not within the same RNA transcript.

Alternative Splicing

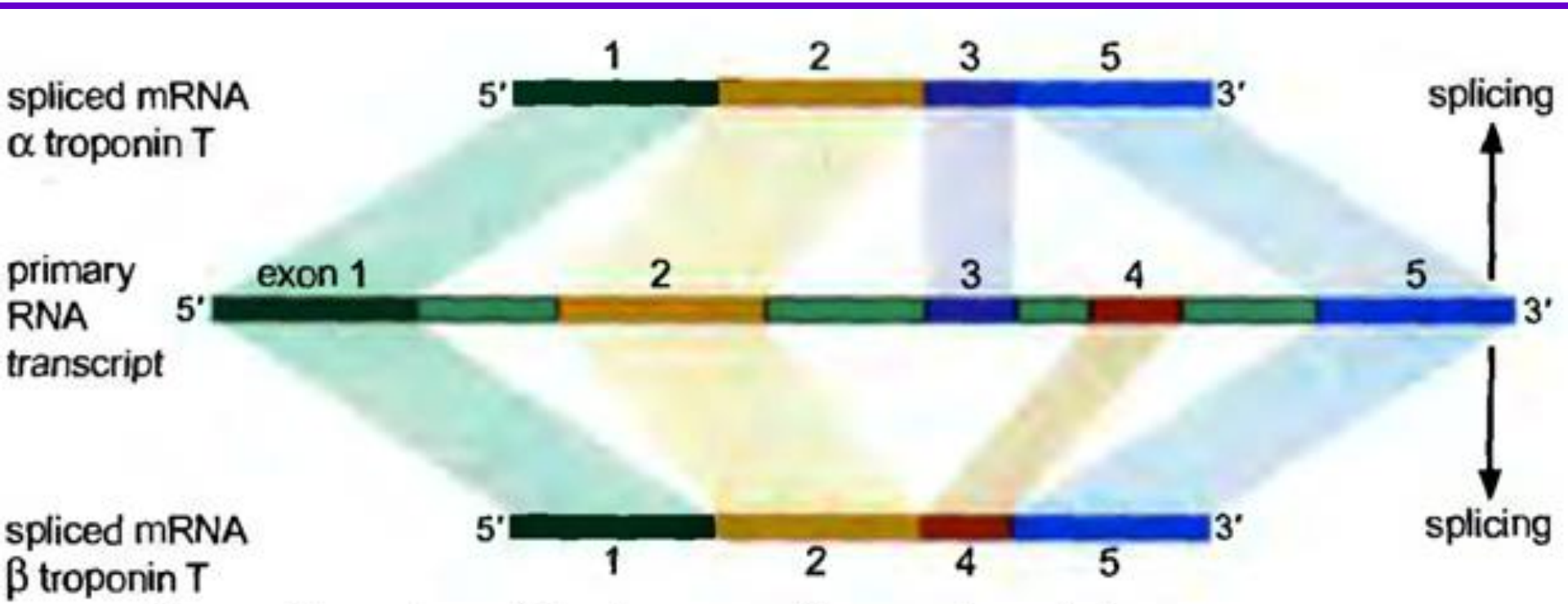


Figure: Alternative splicing in mammalian muscle protein

Alternative Splicing

- Single Genes Can Produce Multiple Products by Alternative Splicing.
- Many genes in higher eukaryotes encode RNAs that can be spliced in alternative ways to generate two or more different mRNAs and, thus, different protein products.

  Ex. (1) Mammalian muscle protein αTroponin T & βTroponin T are produced from same gene by alternative splicing.

  (2) Drosophila DSCAM gene produces 38,016 different mRNAs and proteins.

RNA Editing

- **RNA Editing** is another way of altering the sequence of an mRNA. RNA editing, like RNA splicing, can change the sequence of an RNA after it has been transcribed.
- The protein produced upon translation is different from that predicted from the gene sequence.
- There are two mechanisms that mediate editing: site-specific deamination and guide RNA-directed uridine insertion or deletion.

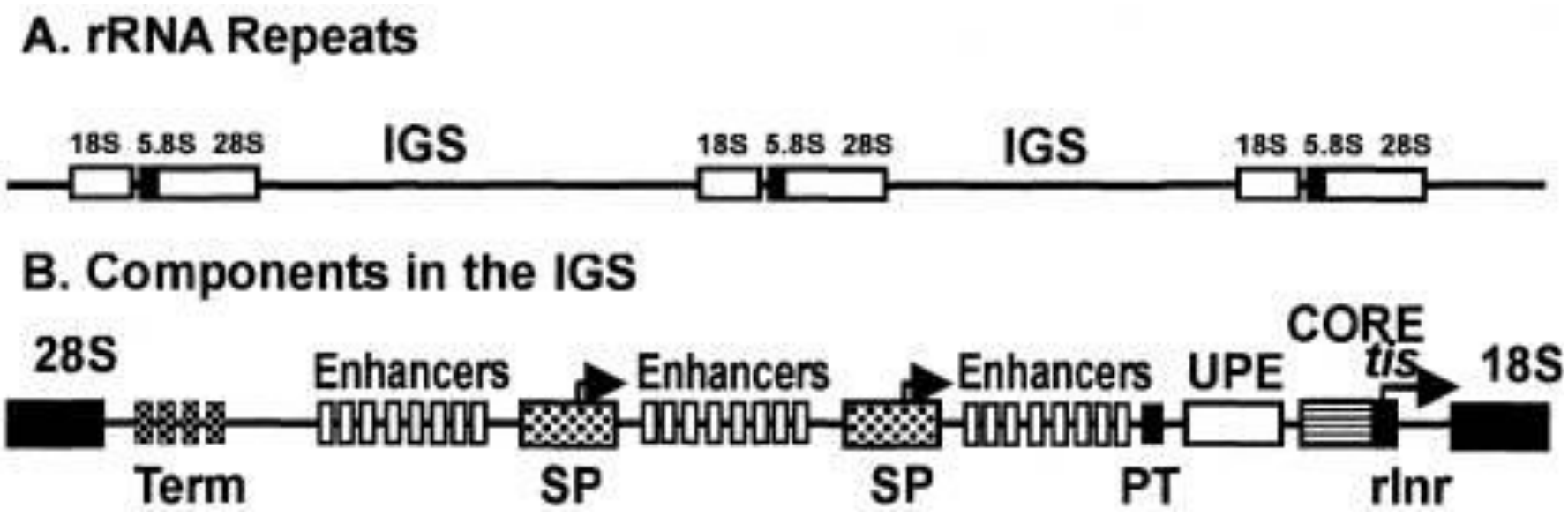
RNA Polymerase I

- **RNA polymerase 1** (**Pol I**), a 590 kDa enzyme that consists of 14 protein subunits (polypeptides).
- It only transcribe ribosomal RNA (but not 5S rRNA) which accounts for over 50% of the total RNA synthesized.

RNA Polymerase I

- About 400 copies of the 42.9-kb rDNA gene are arranged as tandem repeats in nucleolus organizer regions in a cell and these alone will be transcribed by RNA pol1.
- Each copy contains a ~13.3 kb sequence encodes a 45S pre-rRNA which is then post-transcriptionally cleaved into the 18S, the 5.8S, and the 28S RNA molecules.

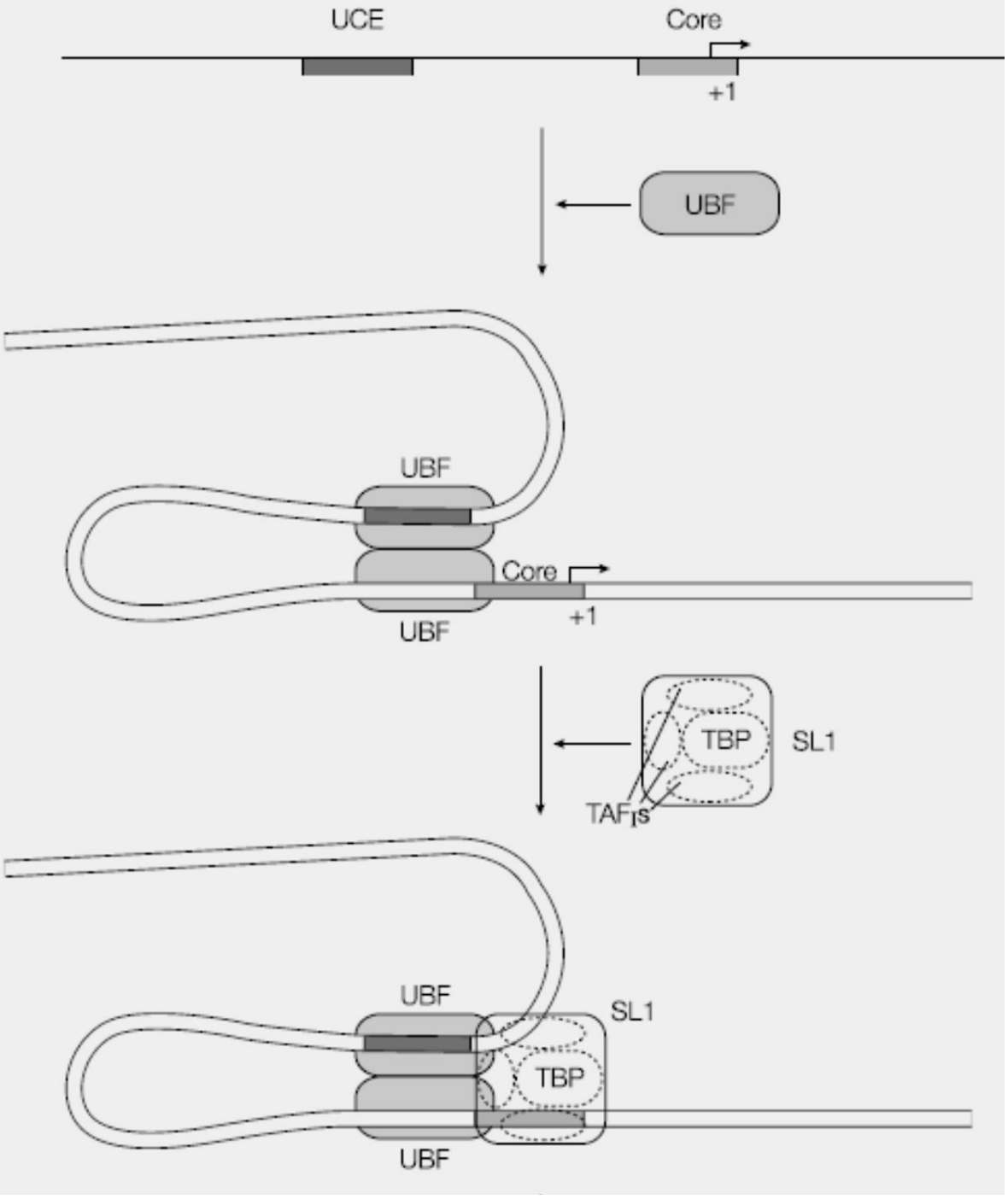
Generic organisation of pol I transcription units

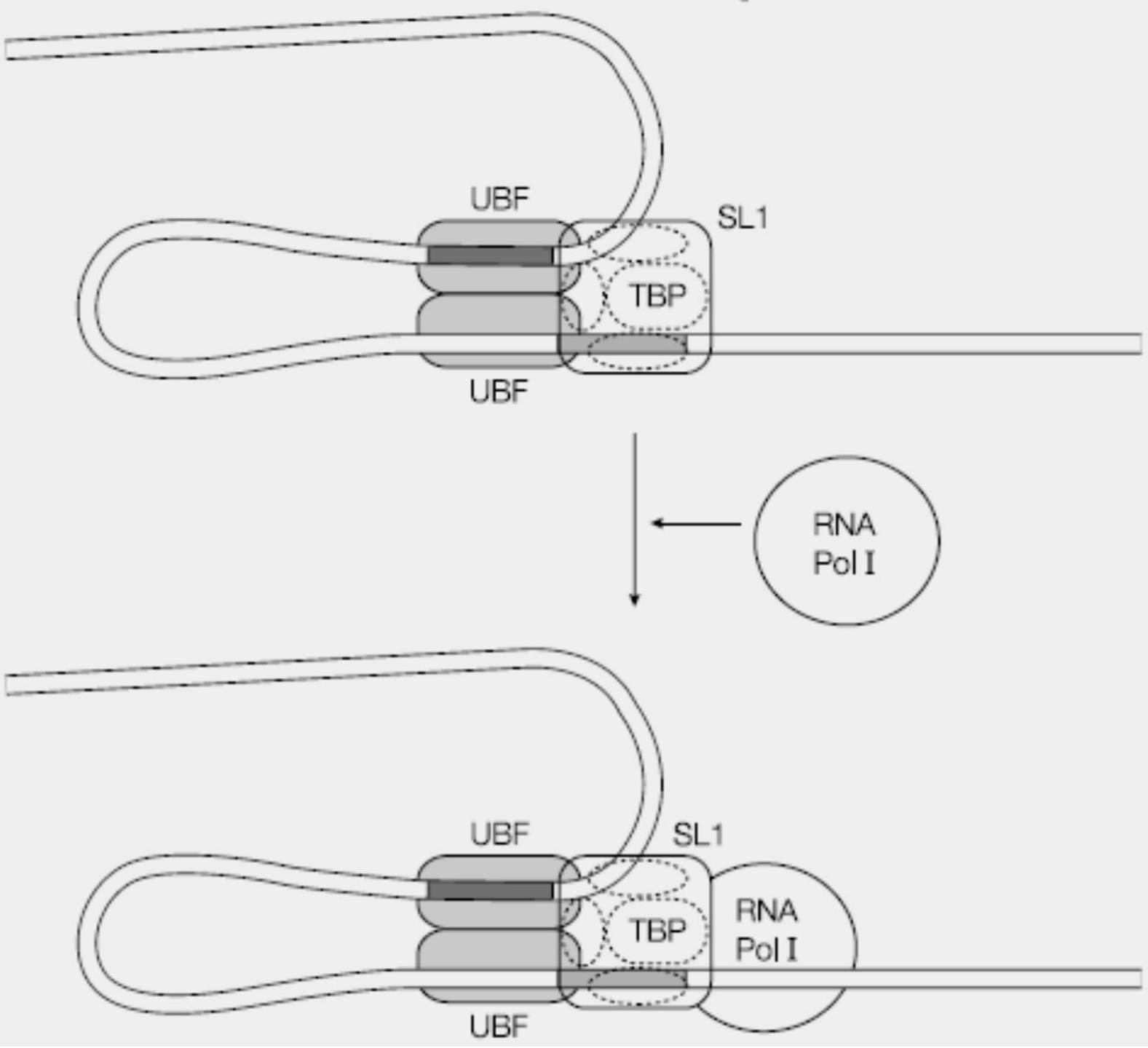


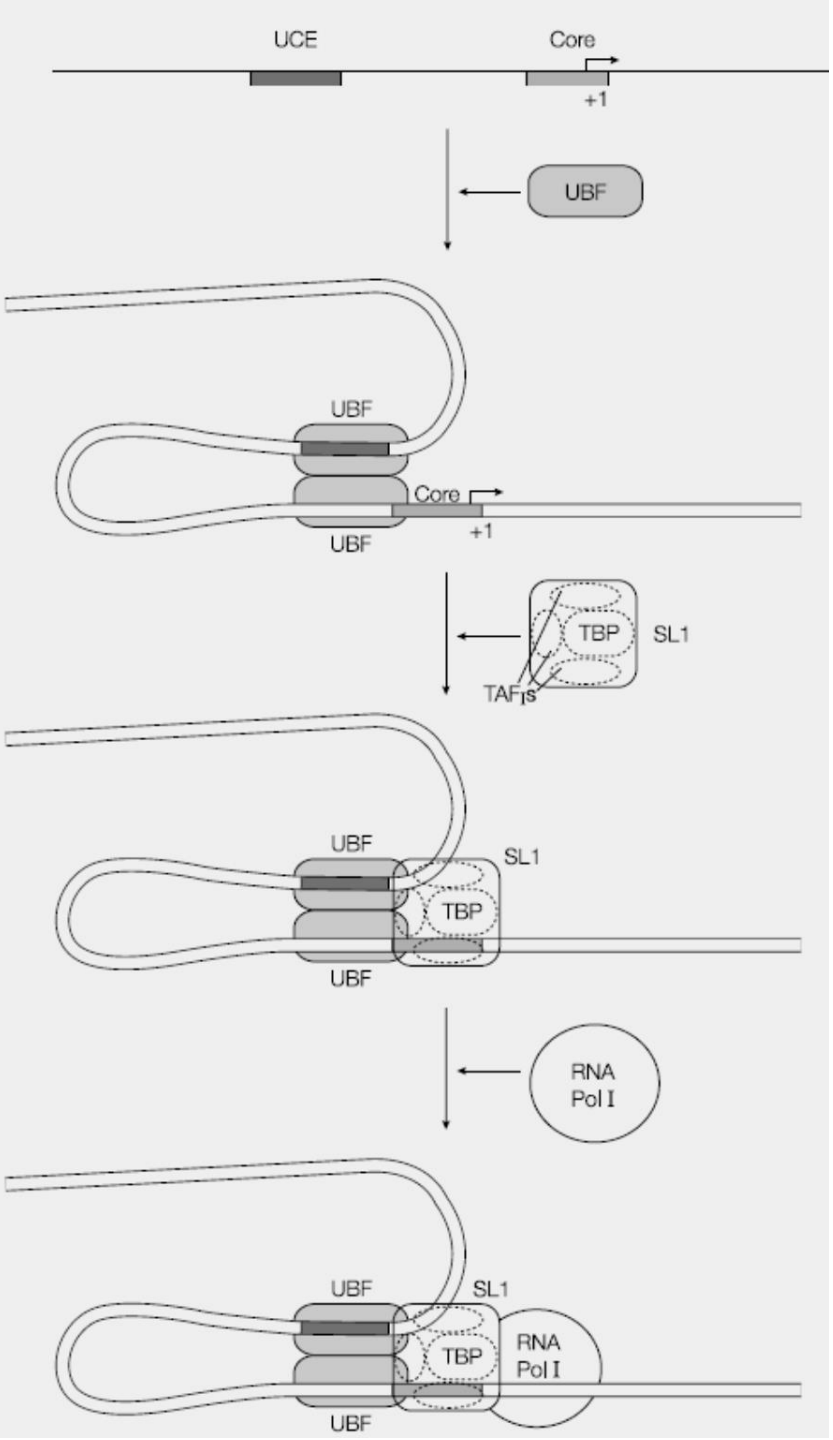
Generic organisation of pol I transcription units. (A) The rRNA coding units are separated by **intergenic spacers** (IGS). (B) The IGS contains a series of **terminators** (term), **enhancers**, a **spacer** promoter (SP), a proximal terminator (PT), the upstream promoter element (UPE) and the promoter core, which includes the ***rInr***. The sites of transcription initiation are indicated by tis and/or the bent arrows.

RNA Polymerase I Has a Bipartite Promoter

- The promoter of RNA polymerase I consists of a core promoter and an upstream promoter element (UPE).
- The core binding factor SL1 binds to the core promoter. **TATA-binding protein (TBP)** is also a part of SL1.
- The factor UBF1 wraps DNA around a protein structure to bring the core and UPE into proximity.
- RNA polymerase I binds to the UBF1-SL1 complex at the core promoter.


UCE
Core
+1
UBF
UBF
Core
+1
UBF
TBP
SL1
TAFIs
UBF
SL1
TBP
UBF


UBF
SL1
TBP
UBF
RNA
Pol I
UBF
SL1
TBP
RNA
Pol I
UBF


UCE
Core
+1
UBF
UBF
Core
+1
UBF
TBP
SL1
TAFIs
UBF
SL1
TBP
UBF
RNA
Pol I
UBF
SL1
TBP
RNA
Pol I
UBF

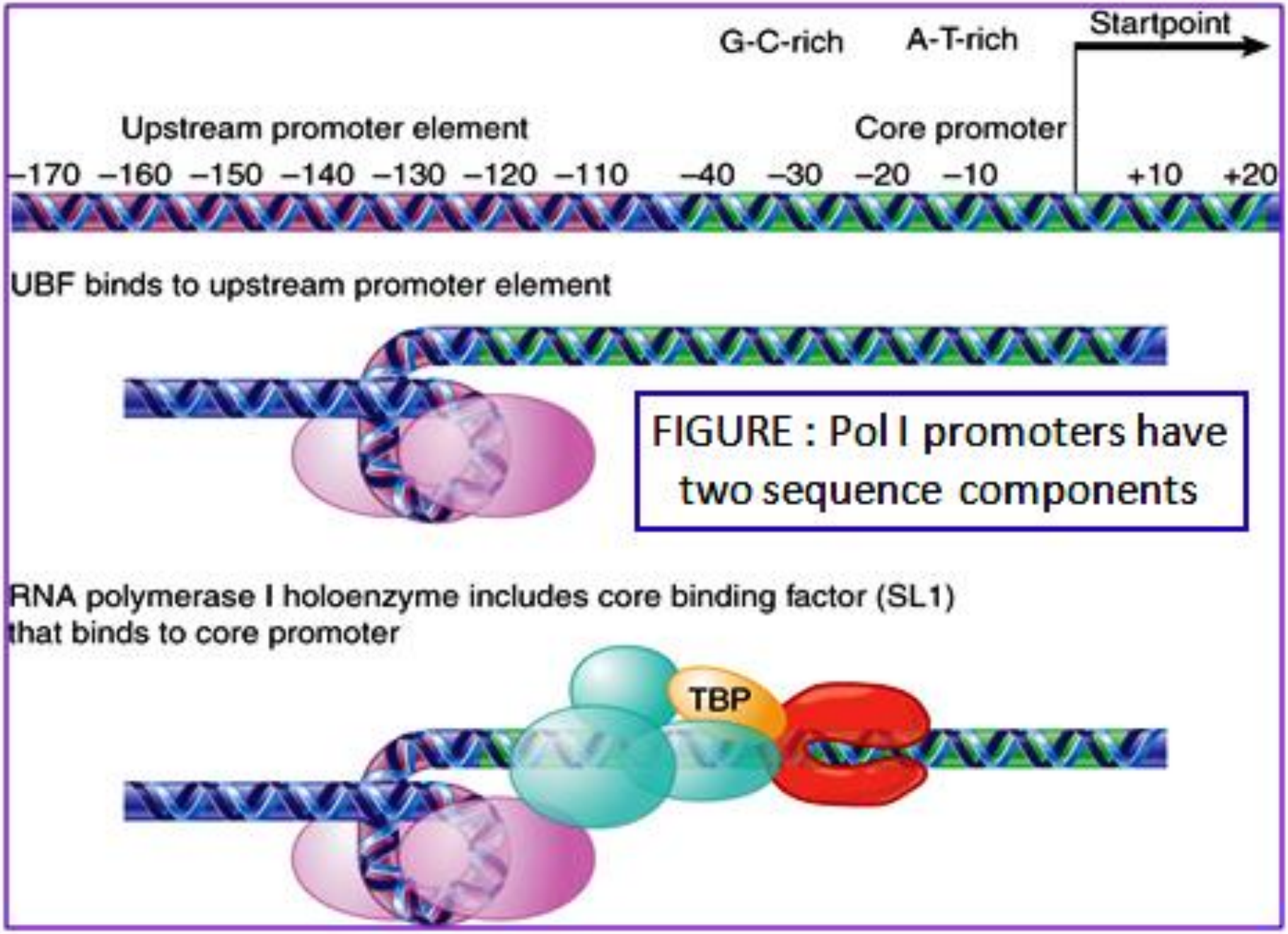


FIGURE : Pol I promoters have two sequence components

RNA Polymerase III

RNA Polymerase III

- RNA polymerase III (RNA Pol III) has at least 17 subunits.
- The enzyme is located in the nucleoplasm and it synthesizes the precursors of 5S rRNA, the tRNAs and other small nuclear and cytosolic RNAs.
- It is moderately sensitive to alpha amanitin and is inhibited by high concentrations (100 mg/ml)

RNA Polymerase III Uses Both Downstream and Upstream Promoters

- RNA polymerase III has **two types** of promoters.

1. **Internal promoters** have short consensus sequences located within the transcription unit and cause initiation to occur at a fixed distance upstream.
2. **Upstream promoters** contain three short consensus sequences upstream of the start point that are bound by transcription factors.

Promoters of RNA Polymerase III

- Two transcription control regions or promoters, called A box and B box, lie downstream from the transcription start site of t RNA genes and 2 such sequences called A and C boxes are seen in the downstream of 5srRNA genes.
- These sequences are therefore both conserved sequences in tRNAs but also conserved promoter sequences in the DNA.

RNA Polymerase III Uses Both Downstream and Upstream Promoters

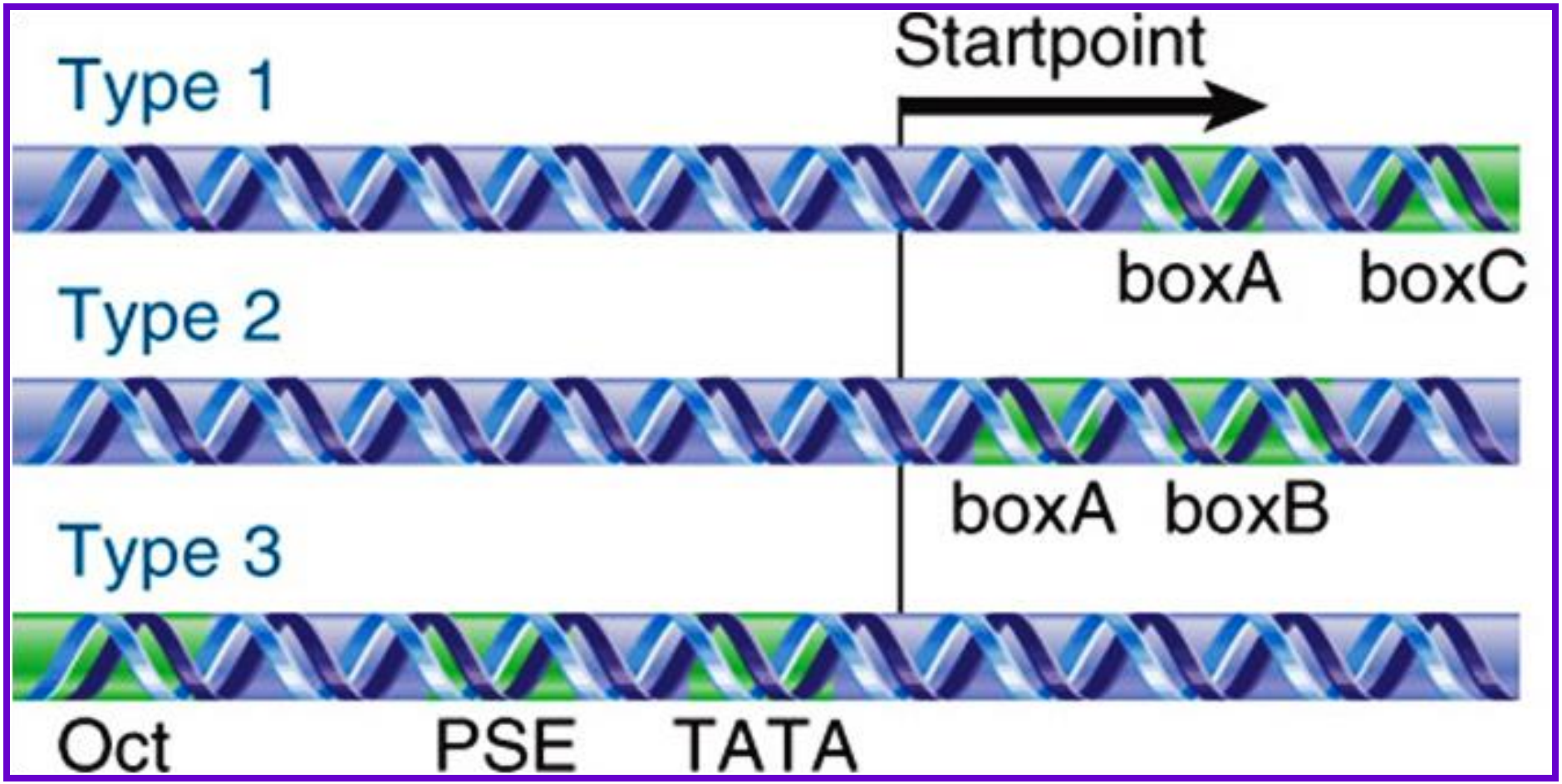


FIGURE: There are three types of pol III promoters

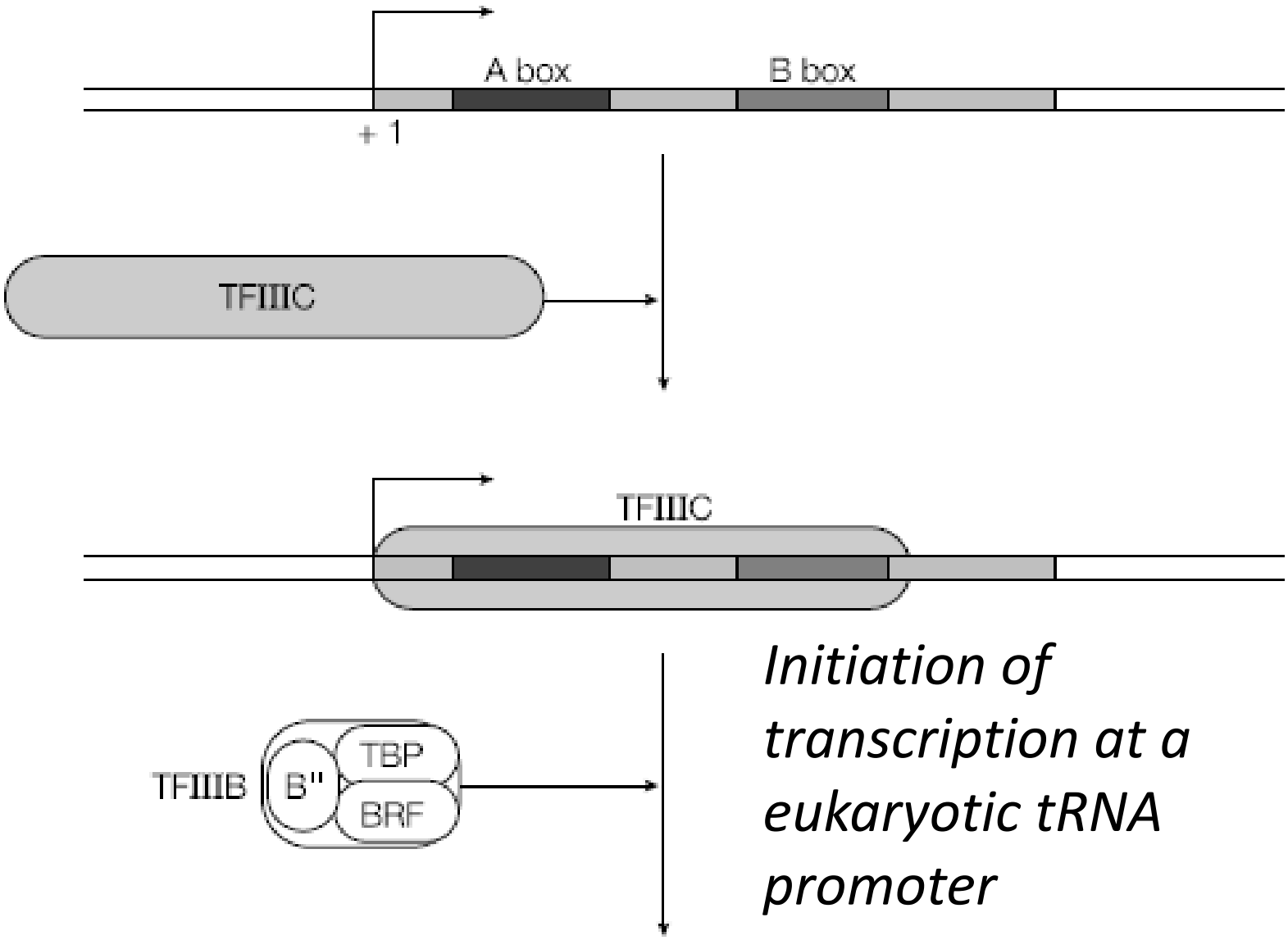
RNA Pol III Transcription Factors

- **TFIIIA**
  1) First eukaryotic transcription factor identified
  2) First DNA-binding protein found to contain zinc fingers (has 9)
  3) Binds specifically to internal promoters of 5S rRNA genes
- **TFIIIB and TFIIIC**
  1) Both required for transcription and depend on each other
  2) TFIIIB contains TBP and a Brf1 (TFIIB-related factor 1) TBP enables RNA polymerase to bind.

RNA Polymerase III Transcription Factors

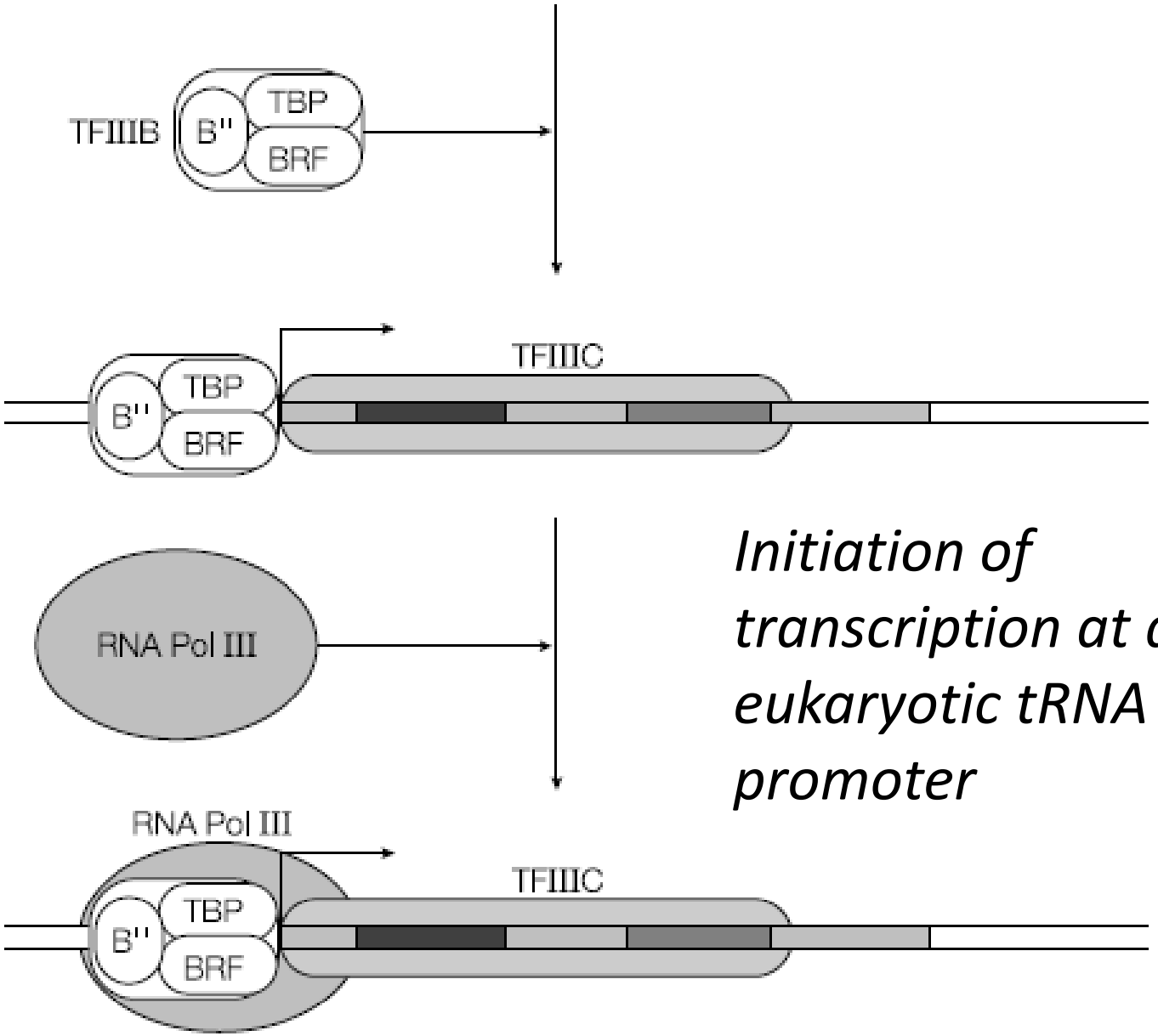
- **Preinitiation complex –** The assembly of transcription factors at the promoter before RNA polymerase binds in eukaryotic transcription.

RNA Polymerase III



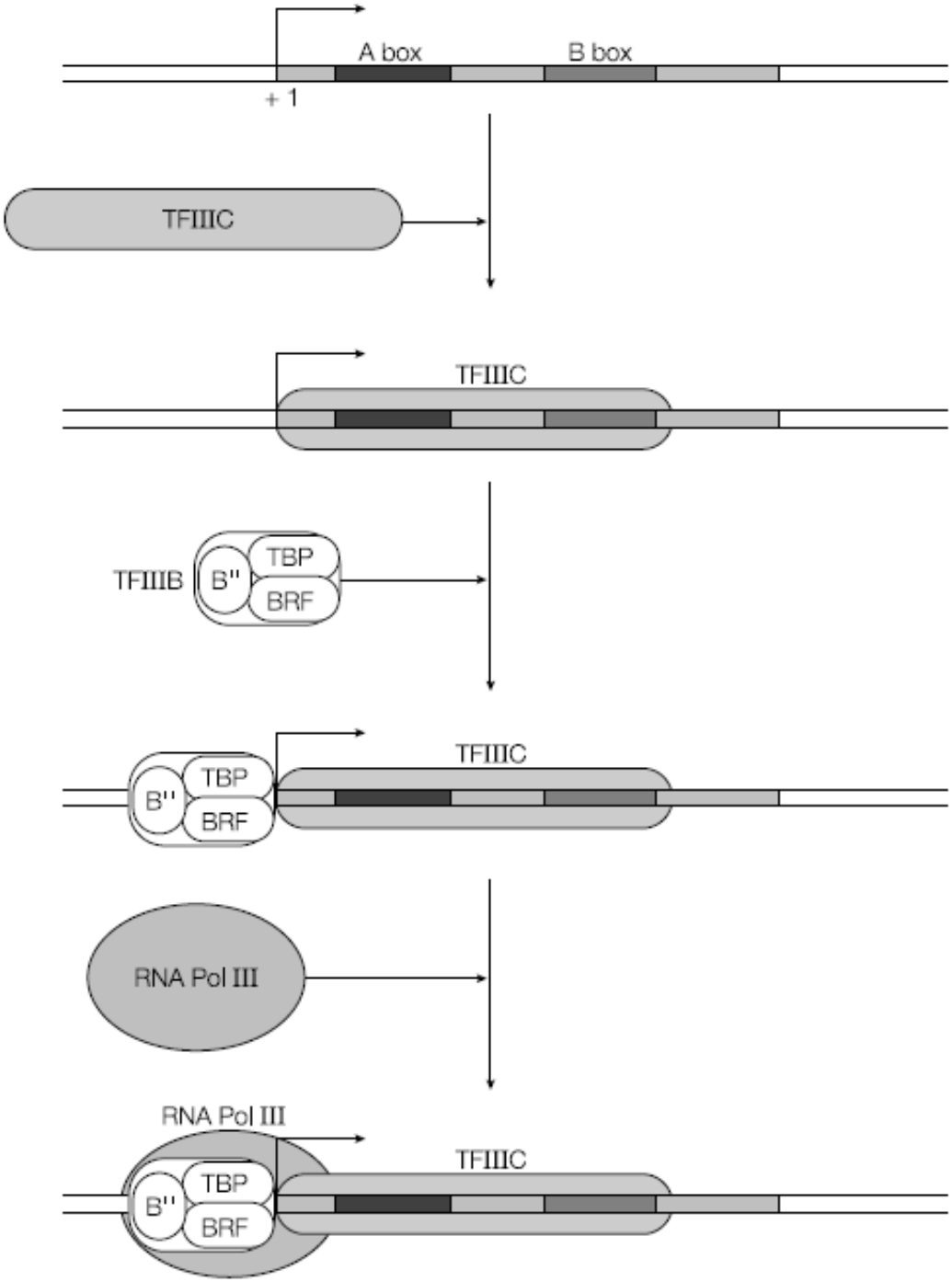
Initiation of transcription at a eukaryotic tRNA promoter

RNA Polymerase III



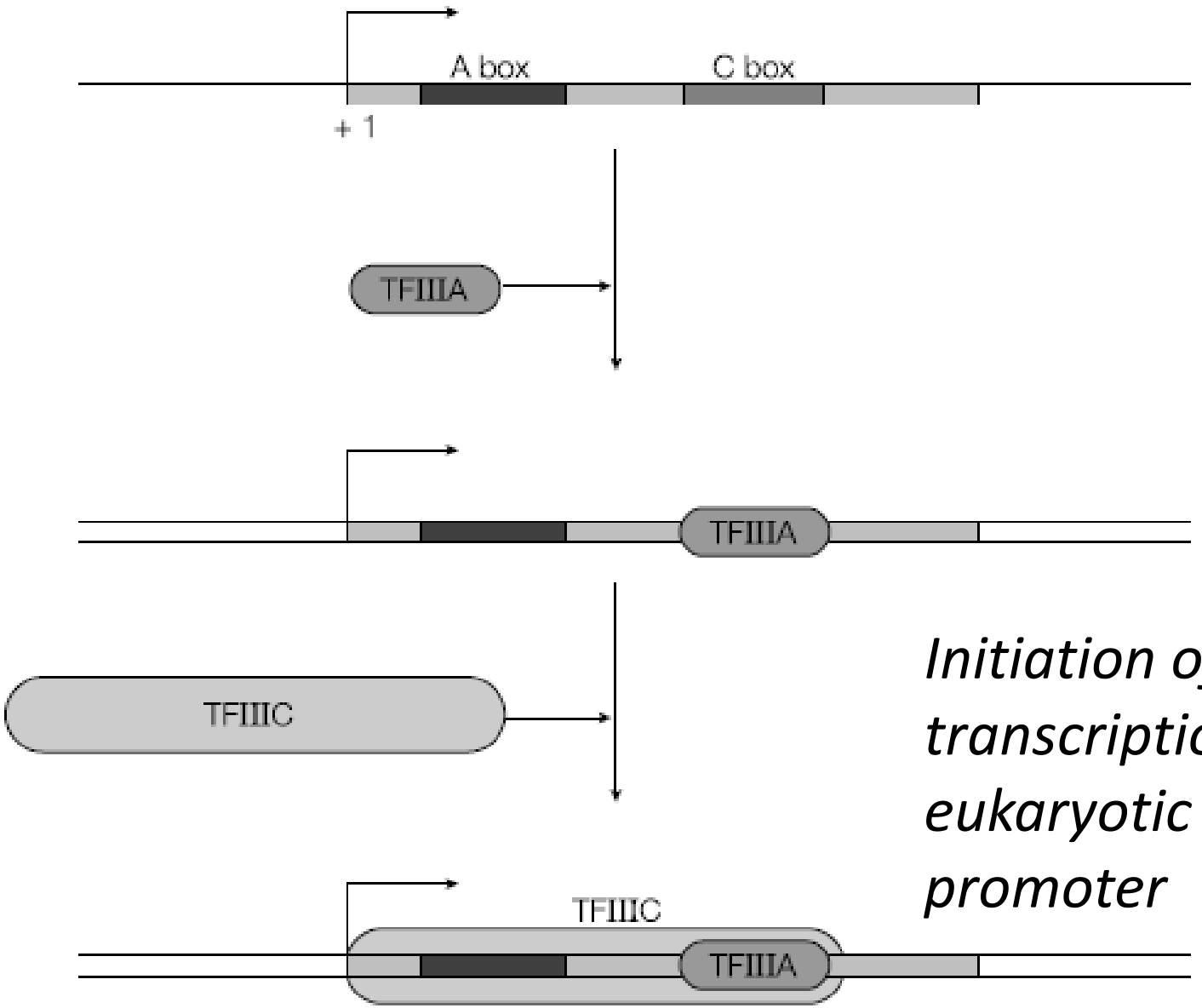
Initiation of transcription at a eukaryotic tRNA promoter

RNA Polymerase III



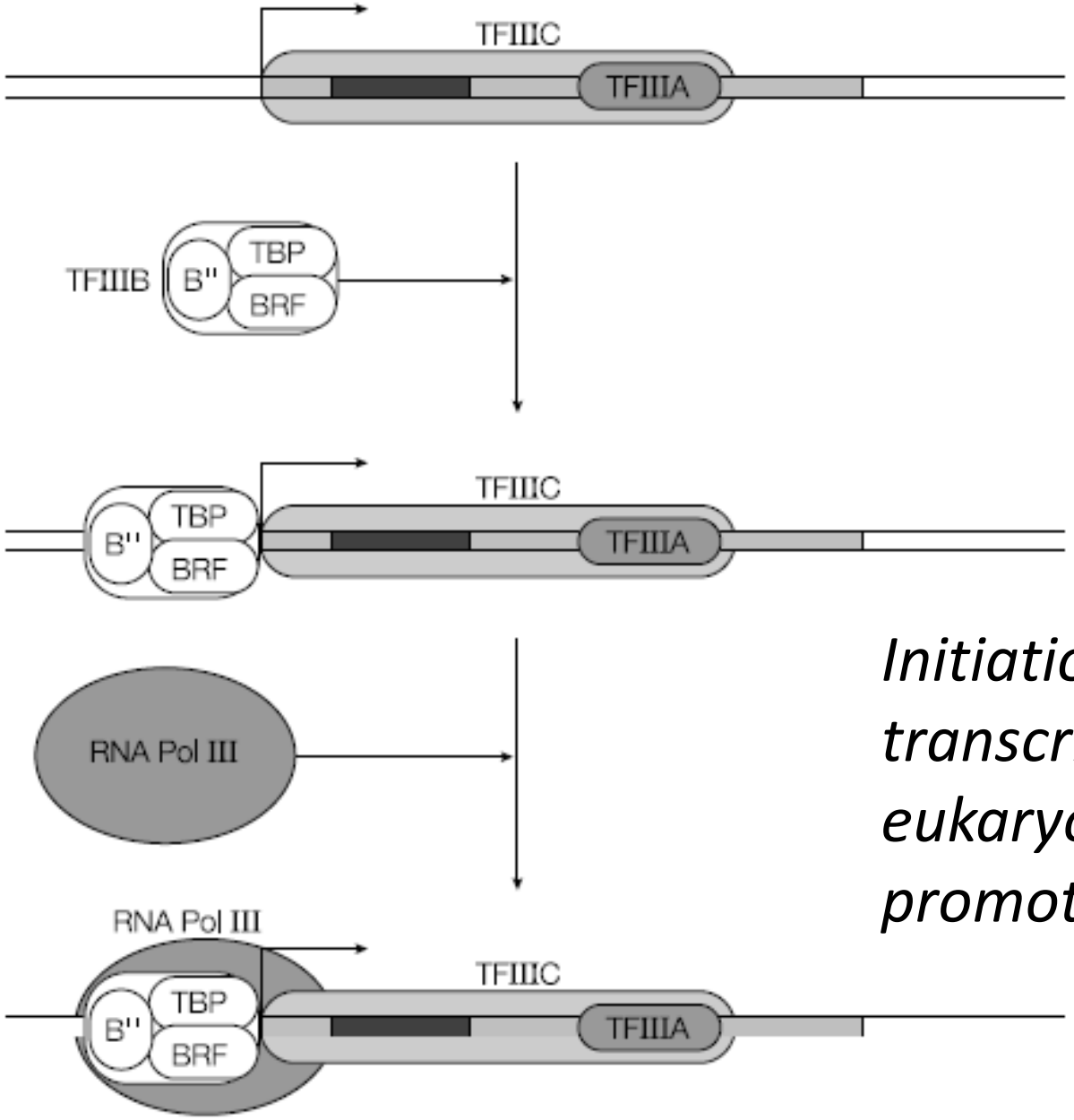
Initiation of transcription at a eukaryotic tRNA promoter

RNA Polymerase III



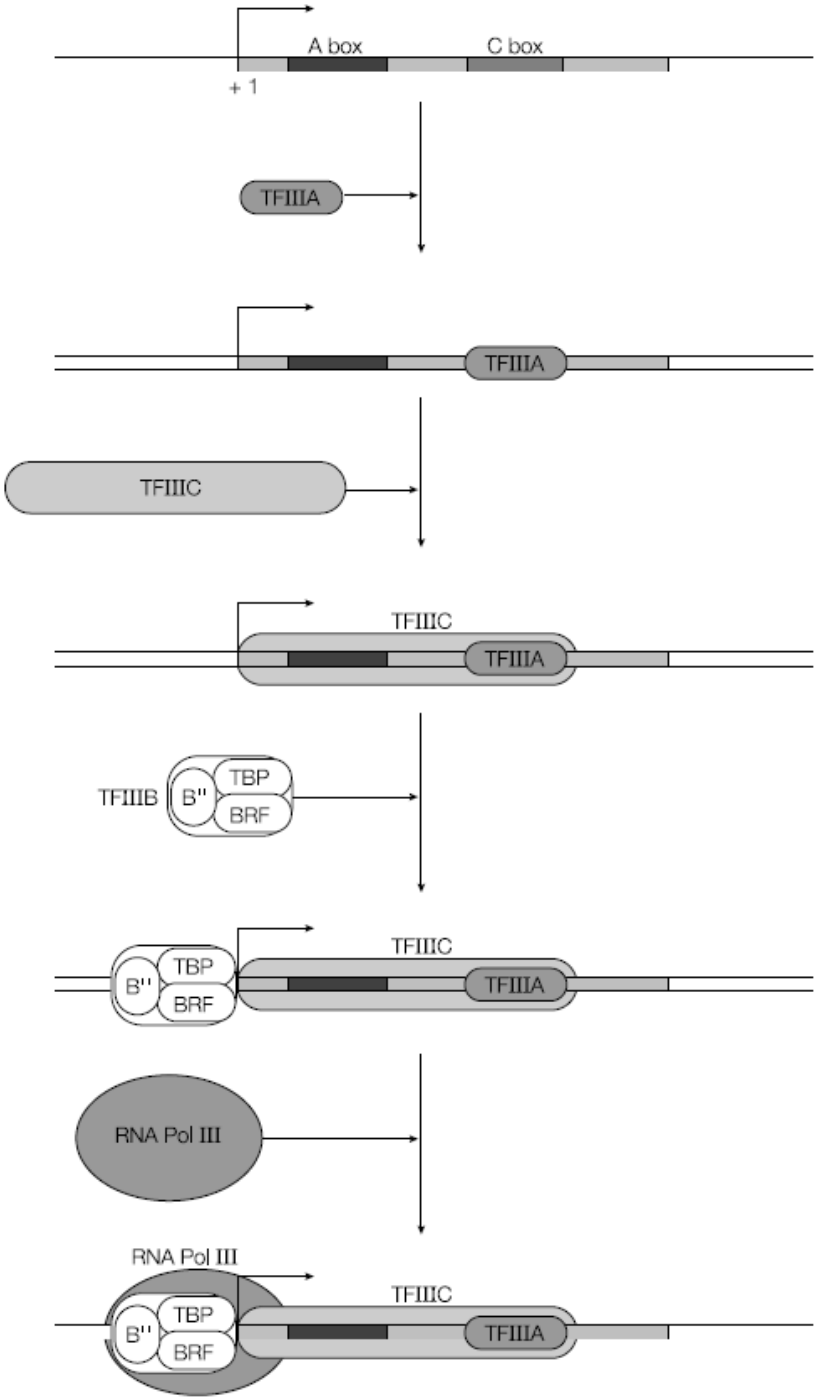
Initiation of transcription at a eukaryotic 5S rRNA promoter

RNA Polymerase III



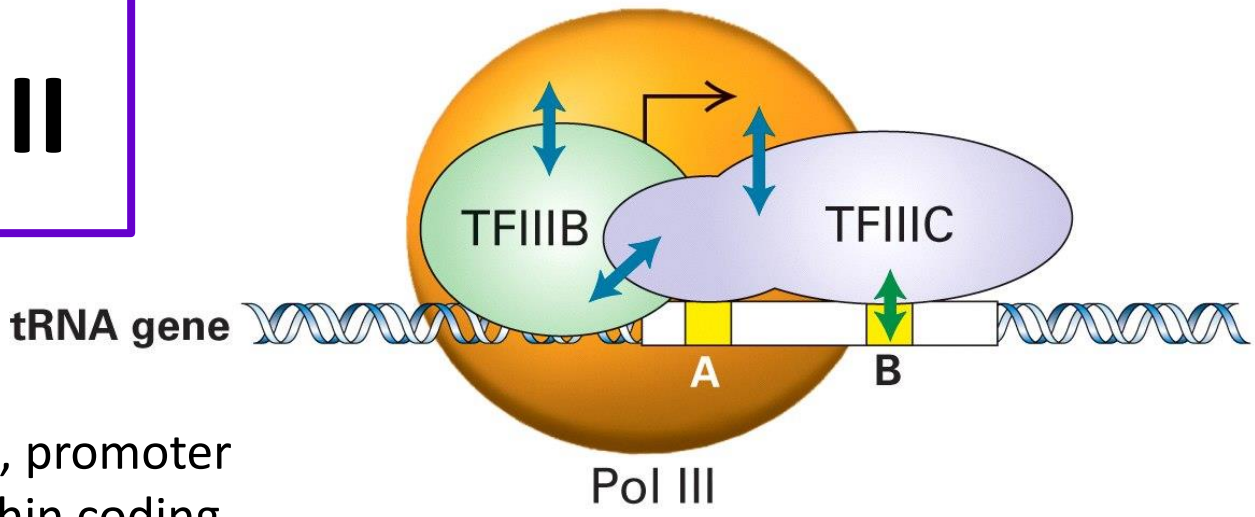
Initiation of transcription at a eukaryotic 5S rRNA promoter

RNA Polymerase III

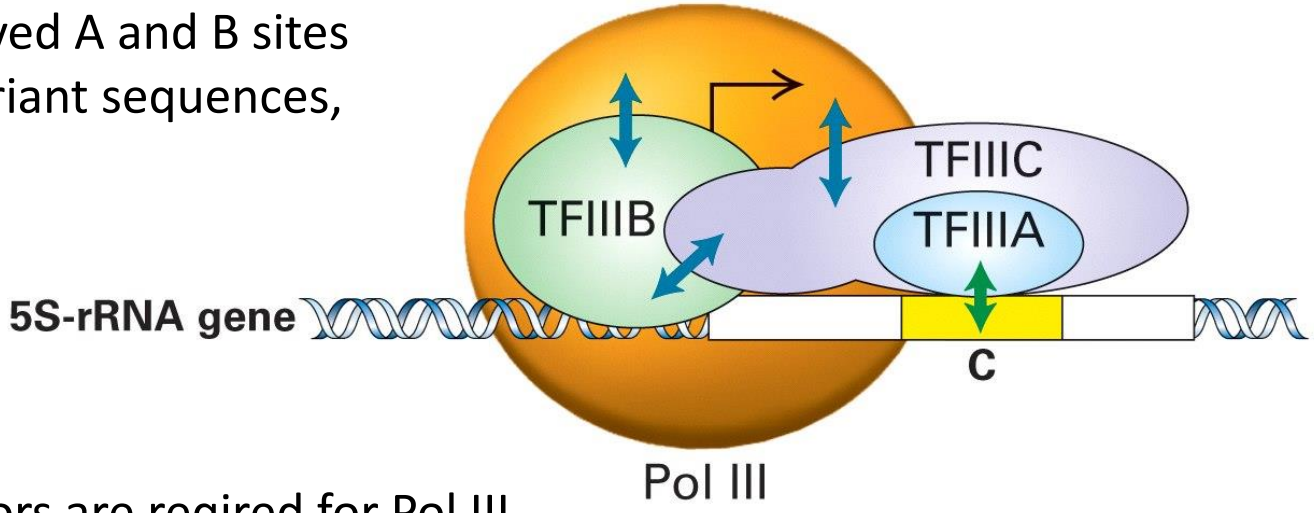


Initiation of transcription at a eukaryotic 5S rRNA promoter

RNA pol III



Unlike other polymerases, promoter sequences lie entirely within coding sequence. The conserved A and B sites also code for two invariant sequences, observed in all tRNAs

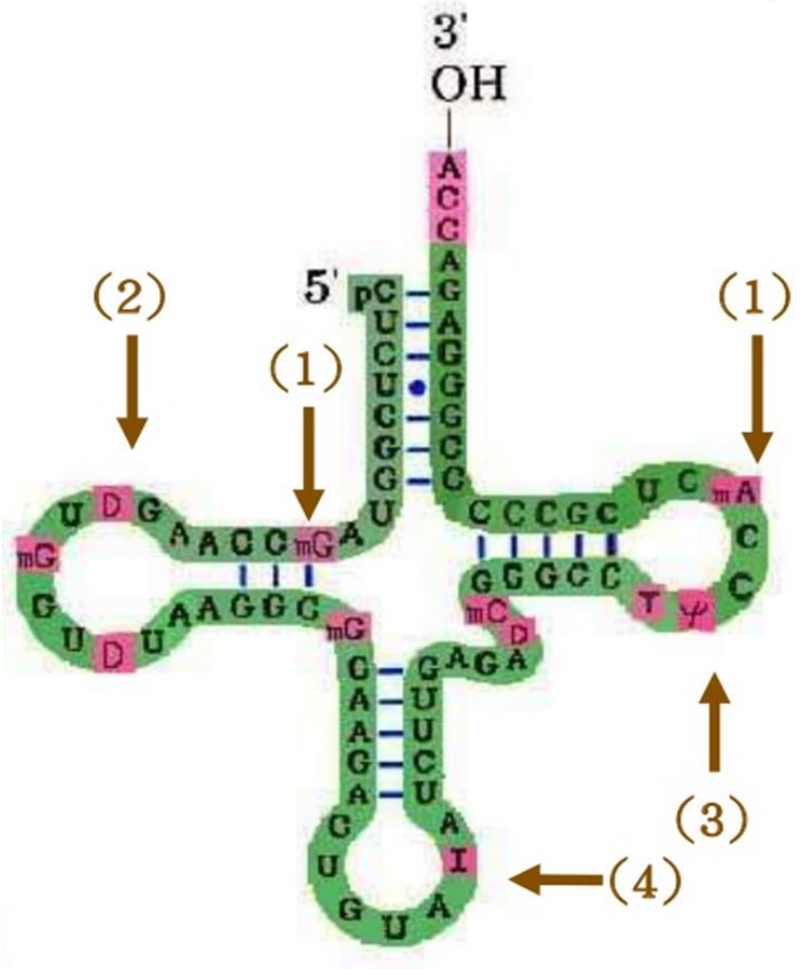


Three initiation factors are reqired for Pol III initiation. TFIIIA is required only for 5S rRNA transcription. TFIIIB has three subunits, one of them similar to TFIIB and another one being TBP (same as for Pol I and II)

Termination of transcription by **RNA pol III**

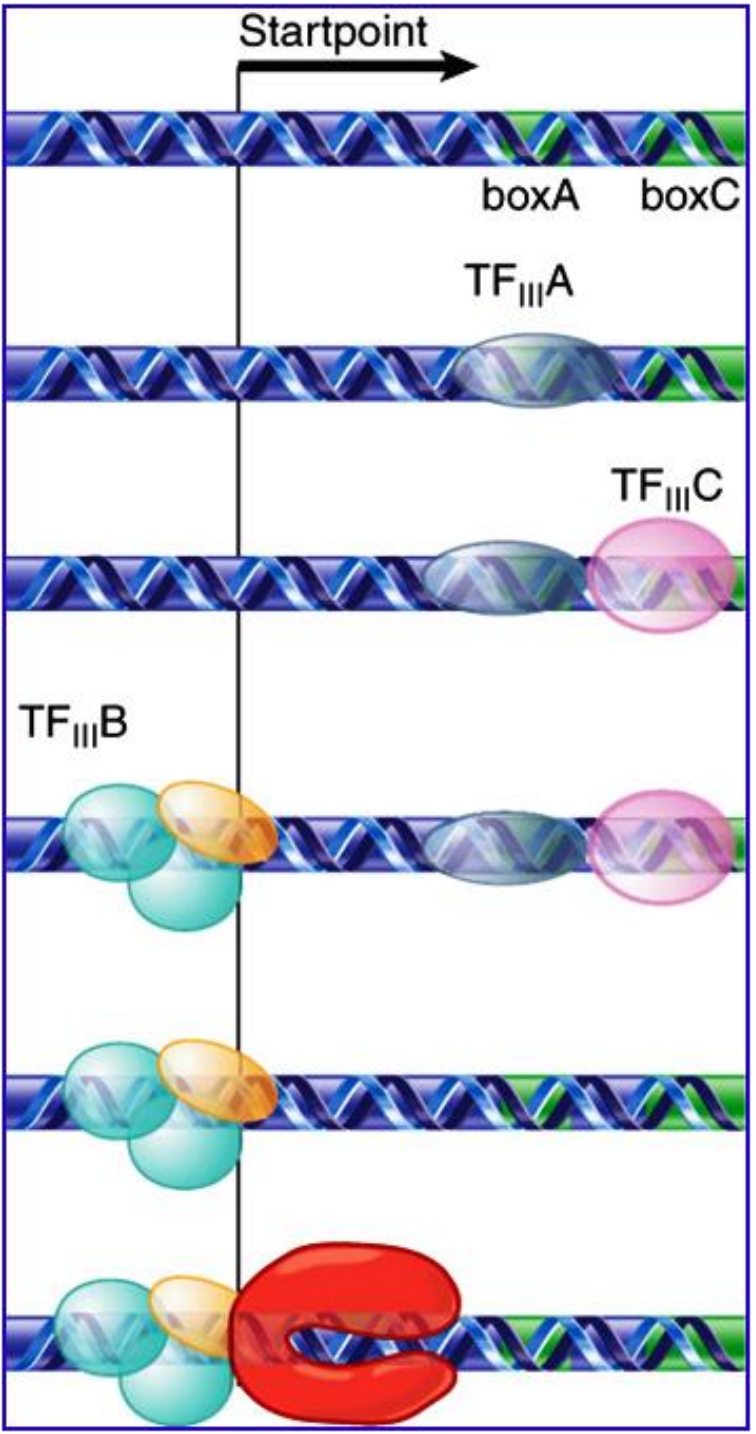
- RNA Pol III recognizes a simple nucleotide sequence of dA residues whose termination efficiency is affected by surrounding sequence.
- The sequence 5-GCAAAAGC-3 is an efficient termination signal in the *Xenopus borealis* somatic 5S rRNA gene.

Base modification



1. **Methylation**
   A→mA, G→mG
2. **Reduction**
   U→DHU
3. **Transversion**
   U→ψ
4. **Deamination**
   A→I

THANK YOU



RNA Polymerase III Uses Both Downstream and Upstream Promoters

- **Assembly factors** – Proteins that are required for formation of a macromolecular structure but are not themselves part of that structure.
- TF$_{III}$A and TF$_{III}$C bind to the consensus sequences and enable TF$_{III}$B to bind at the startpoint.

FIGURE : Type 1 pol III promoters use TFIIIA/C

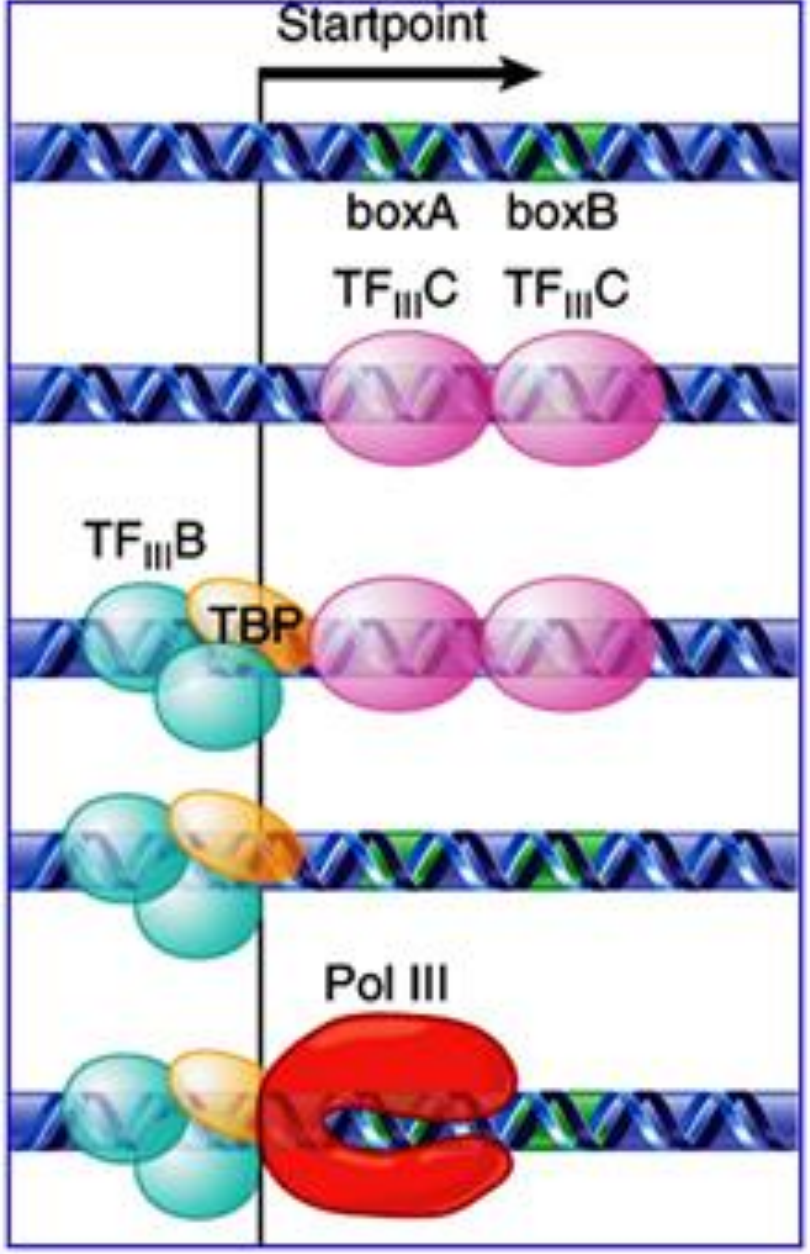
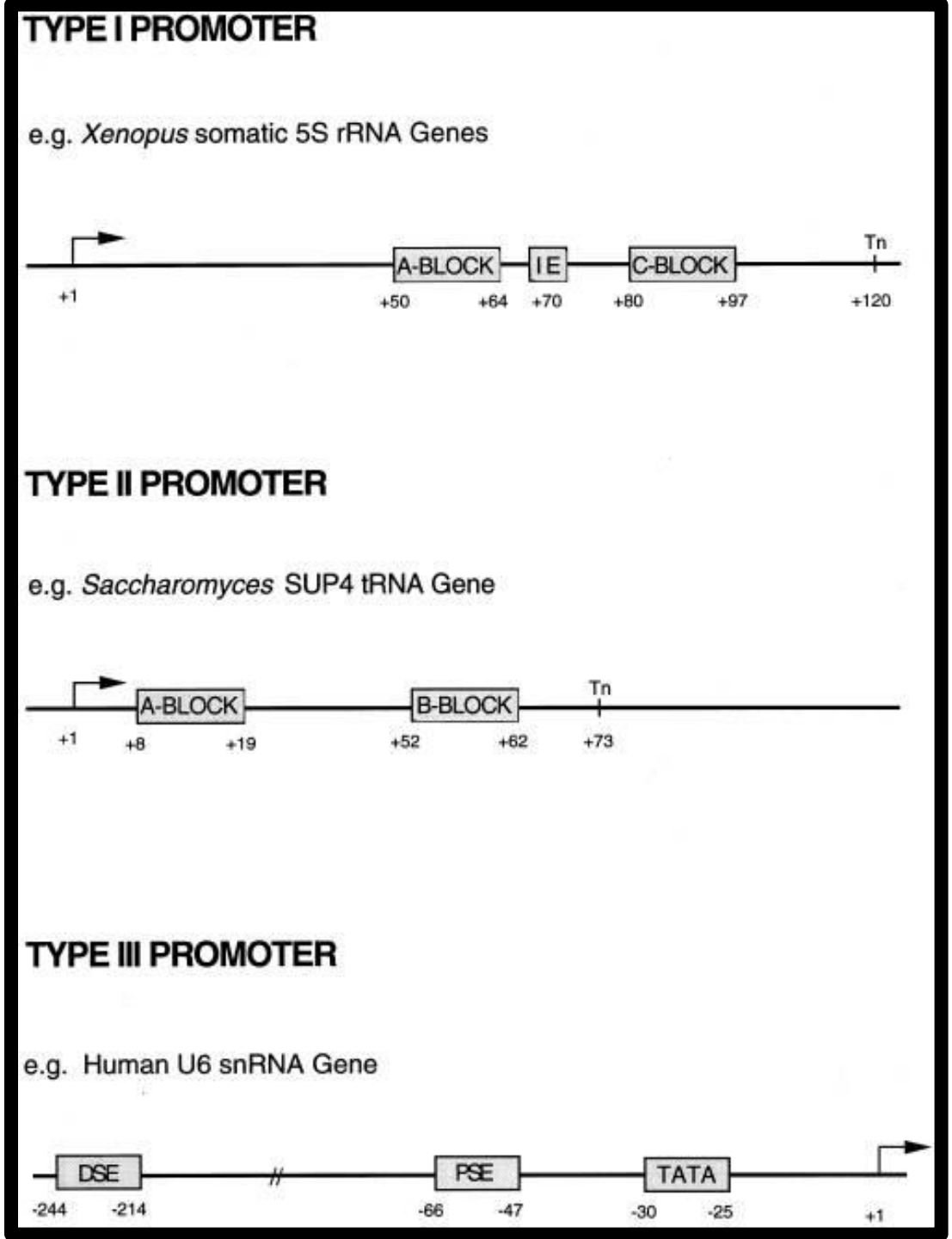


FIGURE : Type 2 internal promoters use TFIIIC

RNA Polymerase III Uses Both Downstream and Upstream Promoters

- TF$_{III}$A and TF$_{III}$C bind to the consensus sequences and enable TF$_{III}$B to bind at the start point.



Organisation of the three general types of promoter used by pol III. The site of transcription initiation is indicated by +1 and the site of termination is indicated by Tn. Also shown are the positions of various promoter elements, including the intermediate element (IE), proximal sequence element (PSE) and distal sequence element (DSE).

Small nuclear ribonucleoproteins

- Small nuclear ribonucleoproteins (snRNPs) are protein–ribonucleic acid (RNA) complexes defined by a core noncoding RNA of approximately 100–600 nucleotides and tightly bound proteins that together accumulate in the nucleus.
- The snRNPs are best known for their role in RNA splicing complexes, including U1, U2, U4, U5 and U6 snRNPs found in the spliceosome.
- Additional snRNPs are functionally diverse, but in many cases the RNA component of snRNPs can base-pair with a substrate for precise alignment and possible catalysis.
- The U7 snRNP directs 3’-end mRNA formation for histone transcripts, and the 7SK snRNP regulates transcription.
- Two special groups of snRNPs, small nucleolar RNPs (snoRNPs) and small Cajal-body RNPs (scaRNPs), are restricted to their named subnuclear compartments in order to direct post-transcriptional modification of ribosomal and splicing RNAs, respectively.
- Certain herpesviruses express high levels of novel snRNPs involved in the regulation of gene expression.
- Due to their important biological roles, there are many diseases associated with snRNPs.