Transcription in Eukaryotes

Transcription

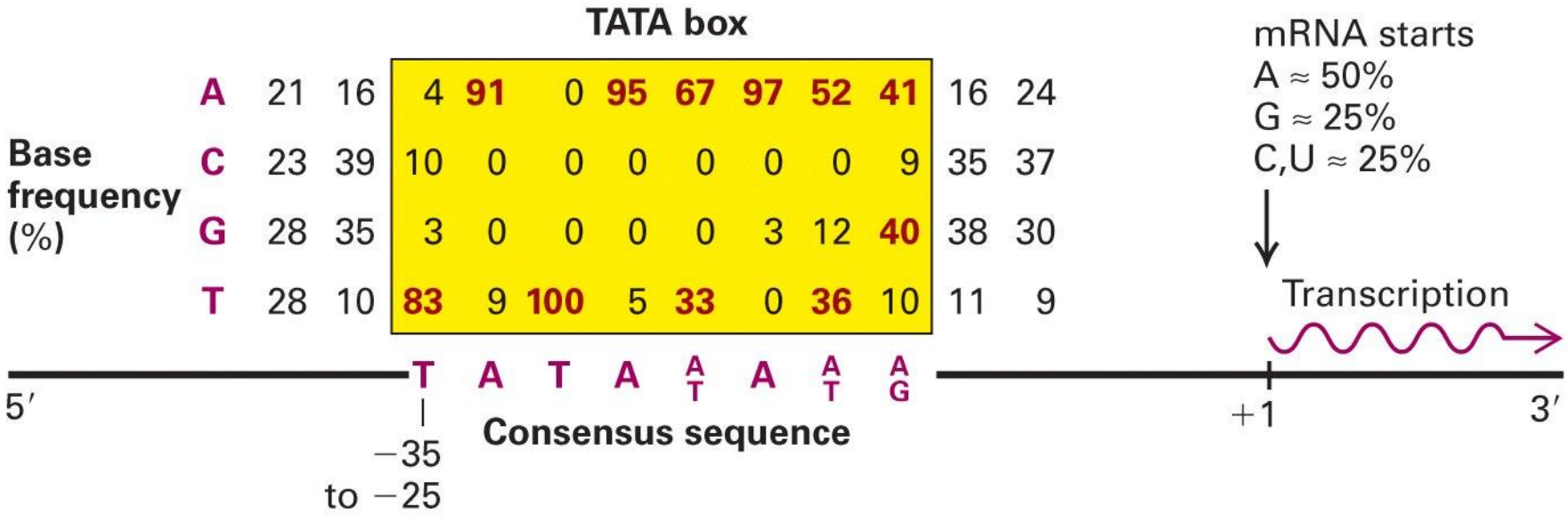
- RNA pol I, located exclusively in the nucleolus, catalyzes the synthesis of 3 of the RNAs found in ribosomes: the 28S, 18S, and 5.8S
- RNA pol II, found only in the nucleoplasm, synthesizes mRNAs and some snRNAs
- RNA pol III, found only in the nucleoplasm, synthesizes the tRNAs, 5S rRNA, and snRNAs not made by pol II

Promoters and Enhancers

- How promoters are studied
  - Through mutational analysis
  - Compare DNA sequences upstream of a number of protein-coding genes
- Found the following:
  - Promoters encompass about 200 bp upstream and contain two regions: the core promoter and promoter proximal elements

- Core promoter: set of cis-acting sequence elements
  - Inr (Initiator) which spans the initiation start site (+1)
  - TATA box or TATA element located at -30 = TATAAA
- Proximal Elements are further upstream (50 to 200 nucleotides); examples are as follows:
  - “cat” box (CAAT) -75; GC box -90
  - Enhancers (function upstream or downstream-1000 bp away)
- The binding of Activator proteins to sequence elements

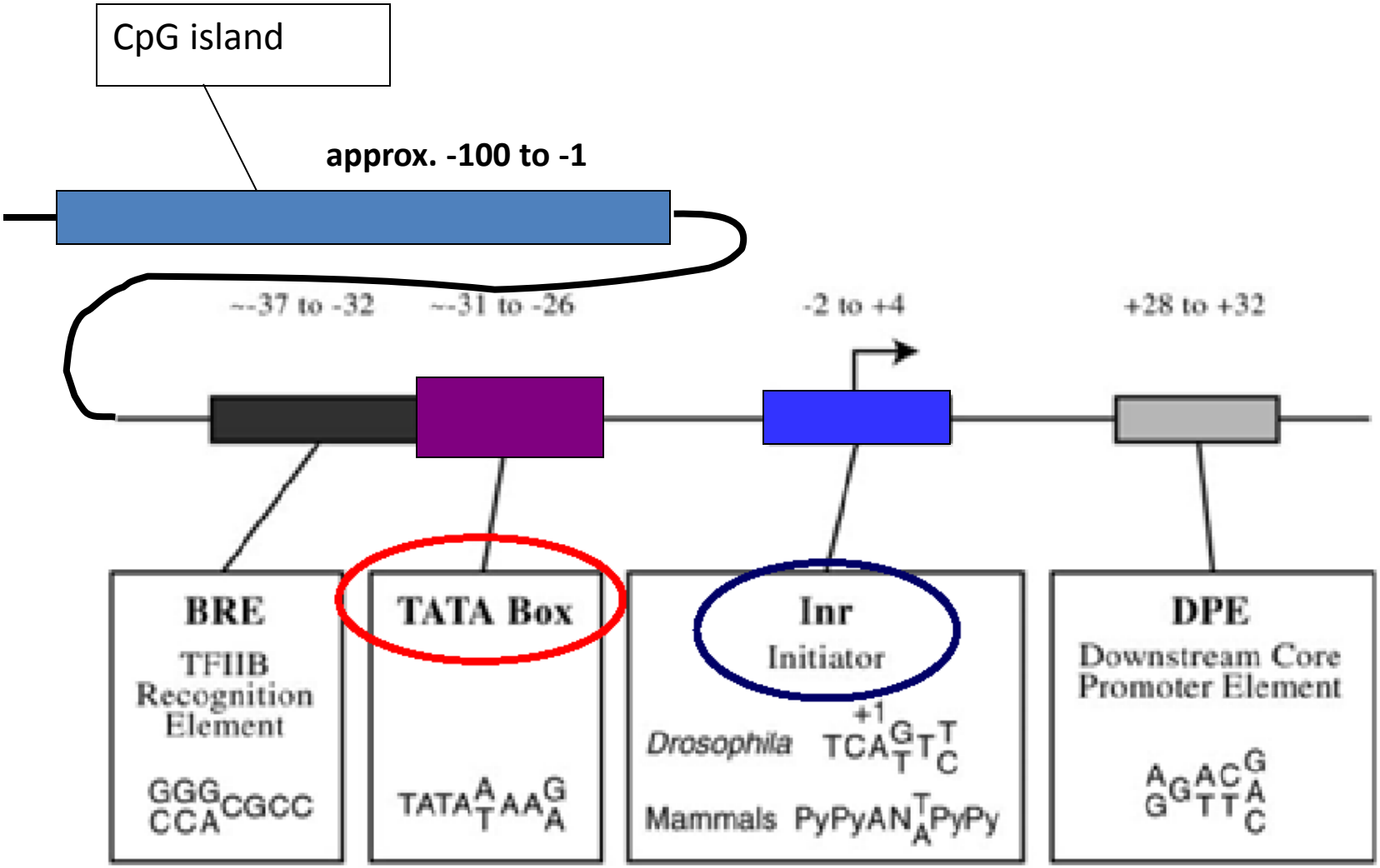
The TATA box

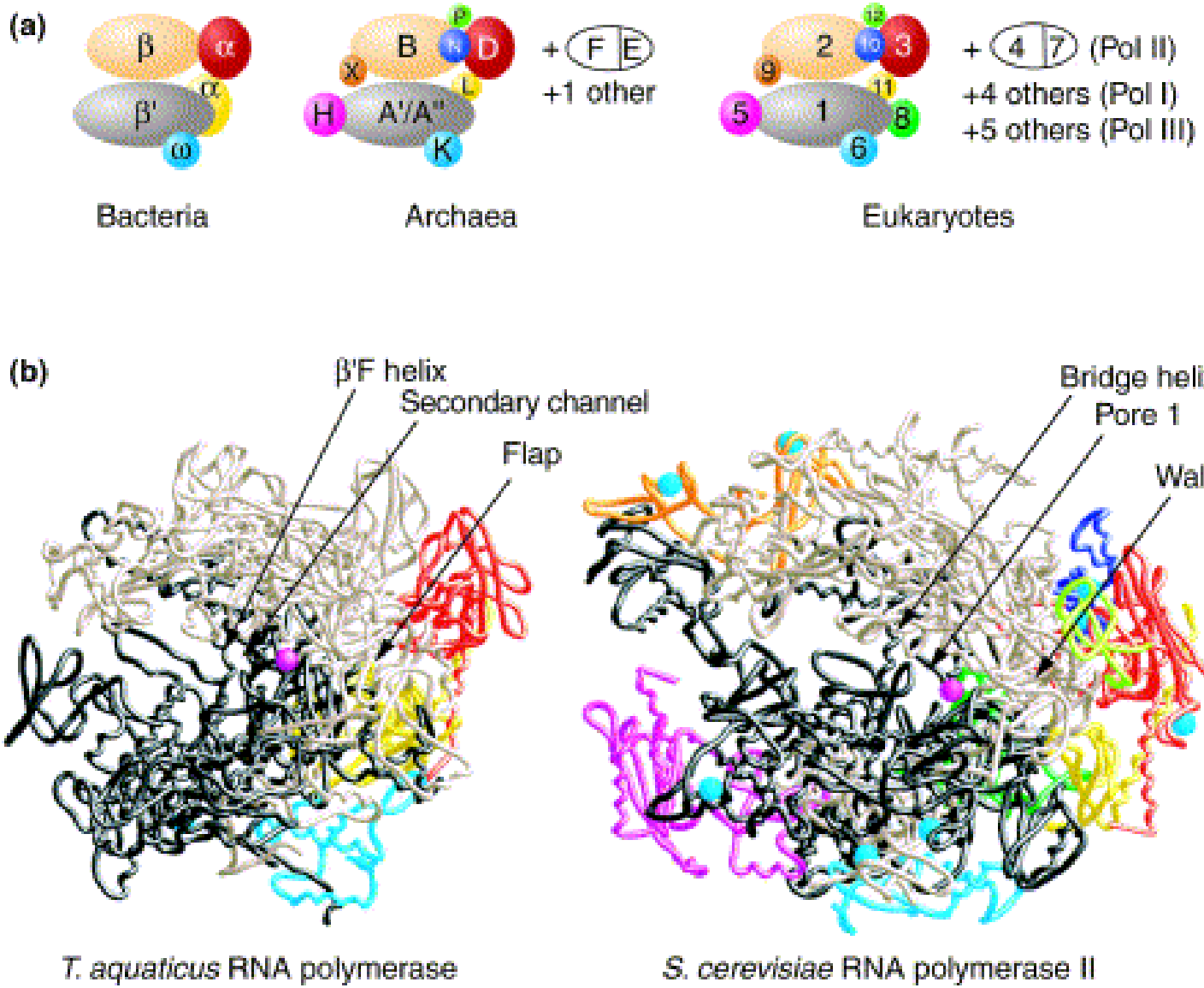

TATA box
Base frequency (%)
A 21 16 4 91 0 95 67 97 52 41 16 24
C 23 39 10 0 0 0 0 0 0 9 35 37
G 28 35 3 0 0 0 0 3 12 40 38 30
T 28 10 83 9 100 5 33 0 36 10 11 9
5′
T A T A A A A G
A T T G
Consensus sequence
−35 to −25
+1
3′
mRNA starts
A ≈ 50%
G ≈ 25%
C,U ≈ 25%
Transcription

Other promoter sequences

- **In**itiato**r** (Inr): Y-Y-A^{+1}-N-T/A-Y-Y-Y. Some promoters contain both initiator and TATA box, whereas others only one of them
- CpG islands: CG rich stretch of 20-50 nucleotides within ~100 base pairs upstream the start site.
- BRE (TFII**B** **r**ecognition **e**lement) sequence, which has the consensus G/C-G/C-G/A-C-G-C-C, is located immediately upstream of the TATA element of some promoters and increases the affinity of TFIIB for the promoter
- **D**ownstream **p**romoter **e**lement (DPE) is present in some genes with initiator promoter. It is located about 30 nucleotides downstream of the transcription start site and includes a common G-A/T-C-G sequence

Summary of promoter elements

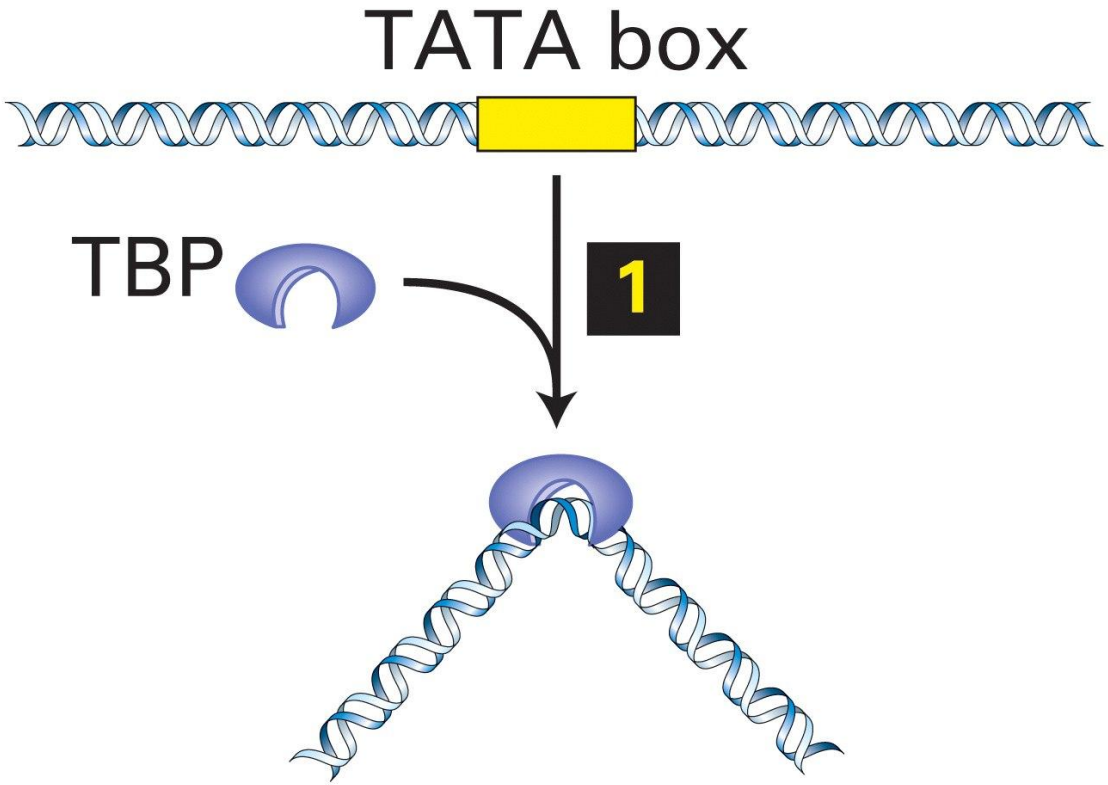



(a)
β
α
α
β'
ω
Bacteria
B
N
D
P
X
L
H
A'/A"
K
+ F E
+1 other
Archaea
2
12
10
3
9
11
5
1
8
6
+ 4 7 (Pol II)
+4 others (Pol I)
+5 others (Pol III)
Eukaryotes
(b)
β'F helix
Secondary channel
Flap
Bridge helix
Pore 1
Wall
T. aquaticus RNA polymerase
S. cerevisiae RNA polymerase II

General transcription factors

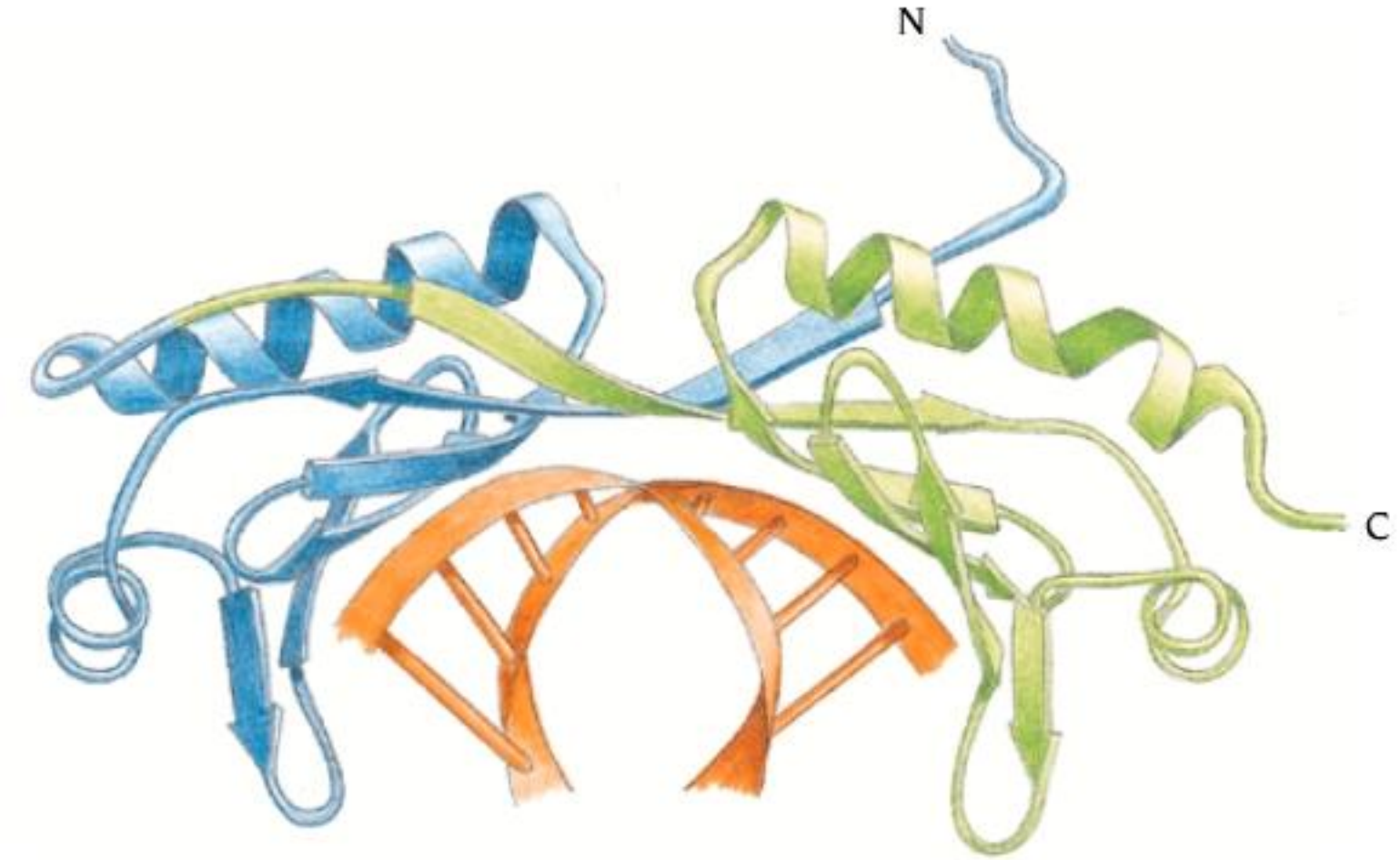
- **G**eneral **t**ranscription **f**actors are required for transcription in eukaryotes from all genes
- GTFs assist RNA Pol II in transcription initiation
- GTFs are designated TFIIA, TFIIB,... and most of them are multimeric proteins
- Equivalent GTFs are highly conserved among the eukaryotes
- In prokaryotes, only one general transcription factor, known as σ factor is required

TFIID

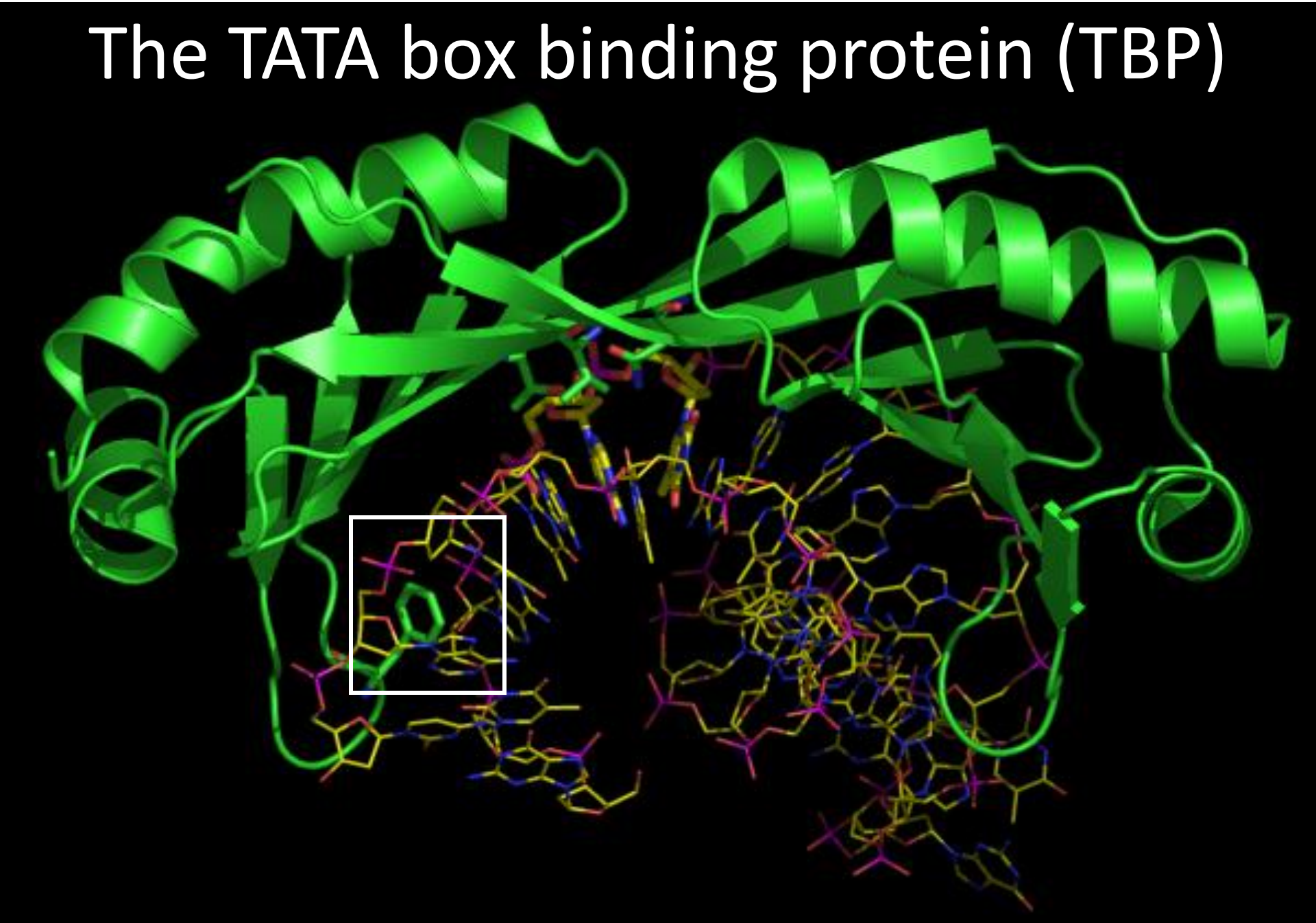


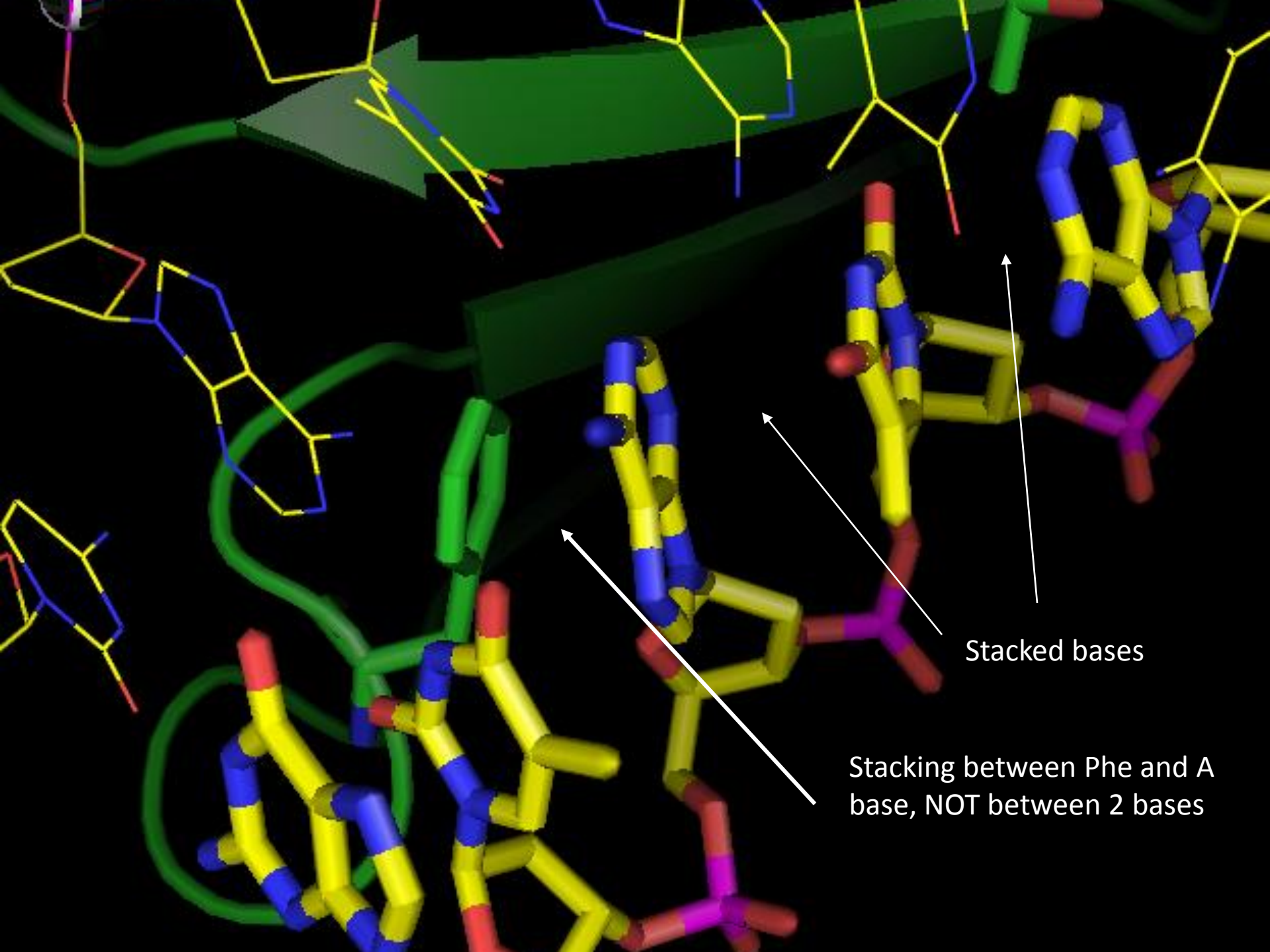
- TFIID is composed of 14 subunits, one of them being **T**ATA box **b**inding **p**rotein, TBP
- Functions: promoter recognition, TFIIB recruitment

The TATA box binding protein (TBP)

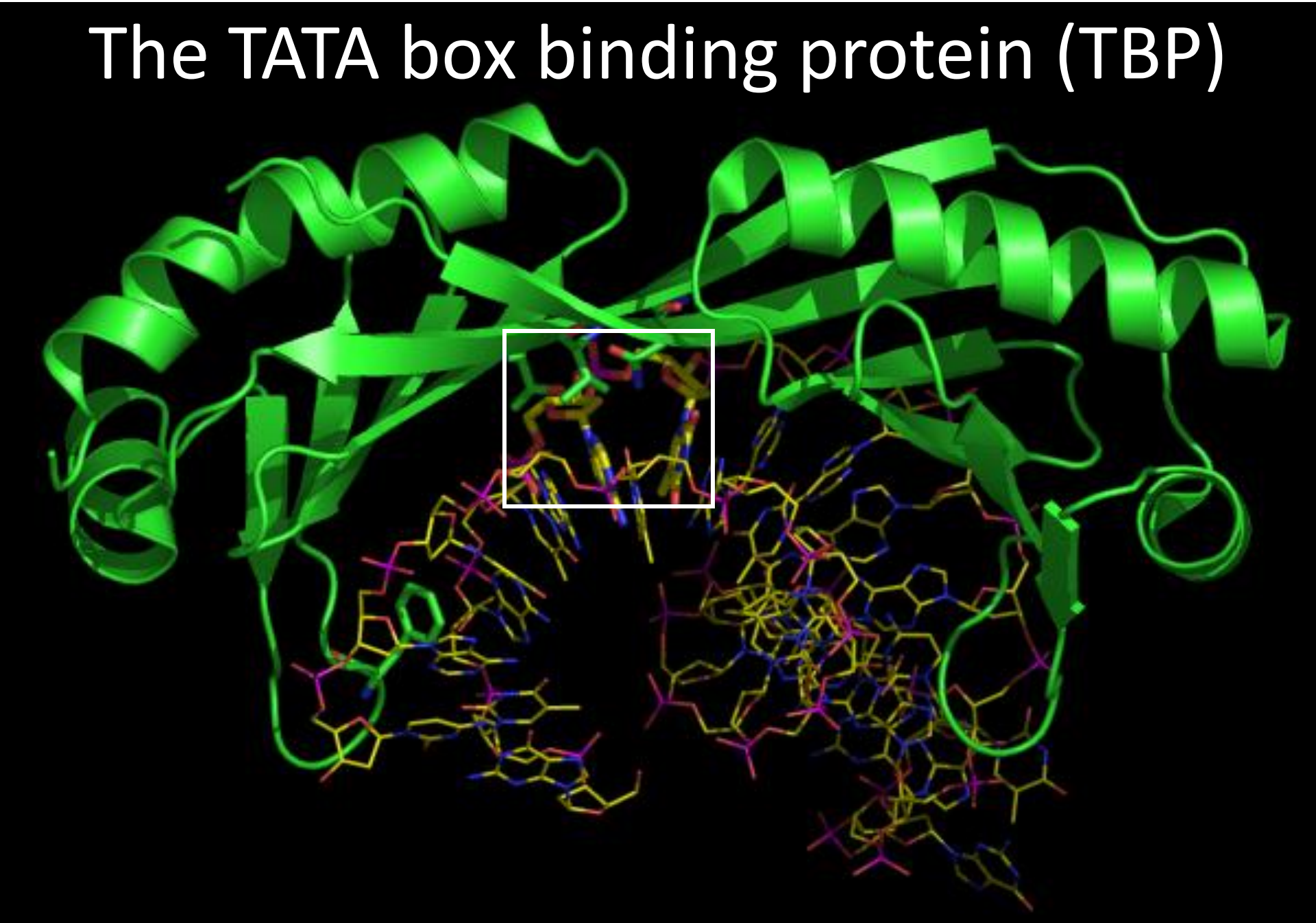


The TATA box binding protein (TBP)

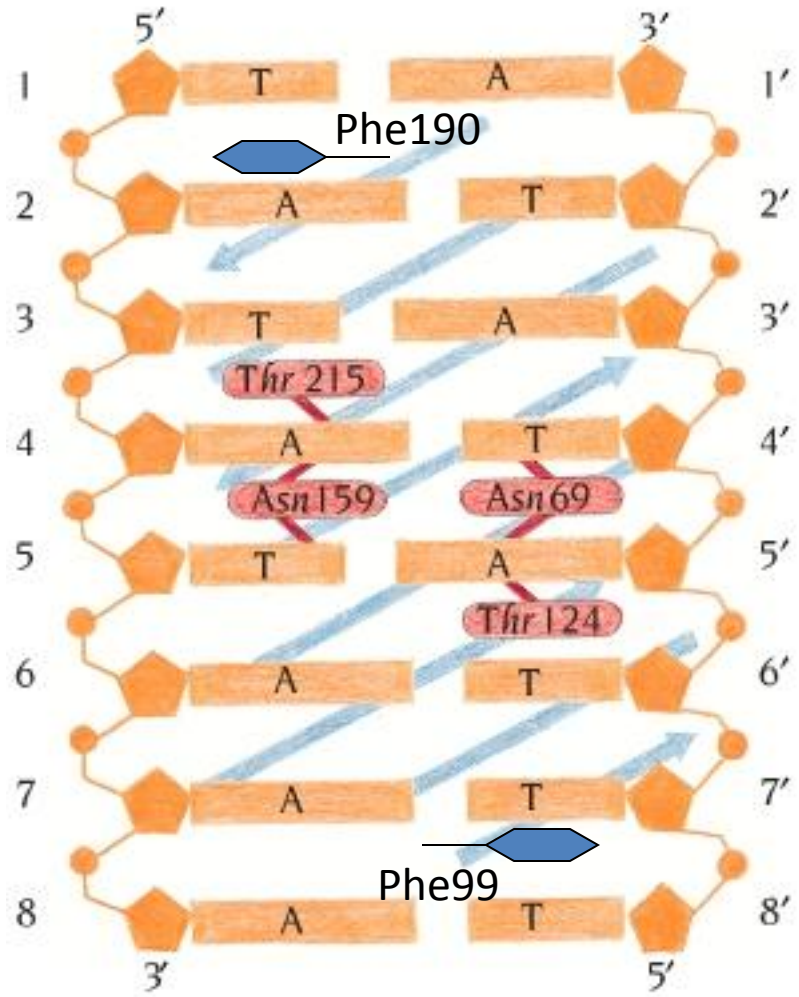



Stacked bases
Stacking between Phe and A base, NOT between 2 bases

The TATA box binding protein (TBP)



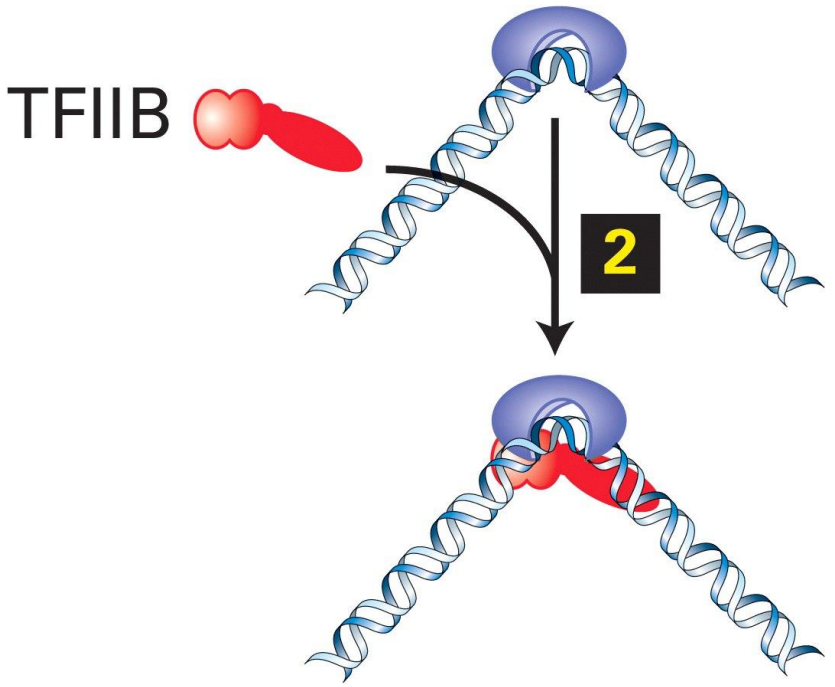
Summary of interactions between TBP and TATA box



TATA box is making a pseudo-twofold sequence-specific interaction with two threonines and two asparagines of TBP

Stacking of phenylalanines against DNA bases is also shown (DNA bending due to stacking is not shown)

TFIIB



- Functions:

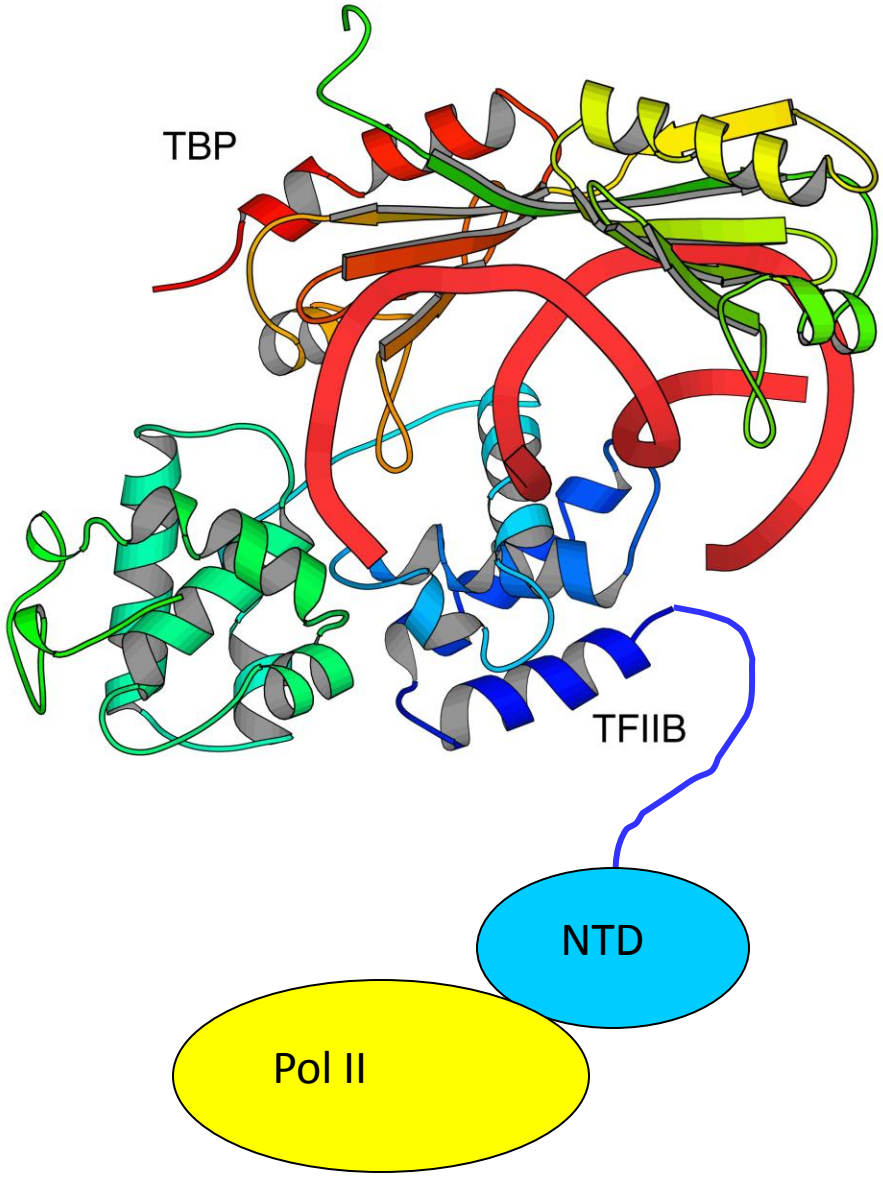
- start site selection for Pol II

- TFIIF-PolII complex recruitment

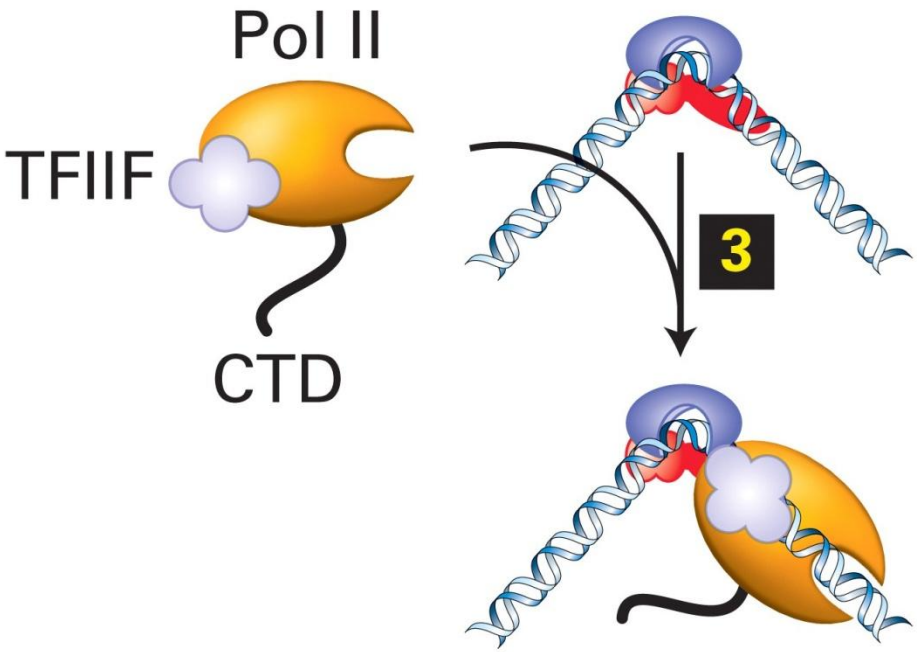
- Some TFIIB mutants result in a shift of transcription start site
- Under certain conditions (BRE element promoter) Pol II together with TFIID and TFIIB can form the minimal initiation complex. At most promoters, however, TFIIE, F and H are necessary

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

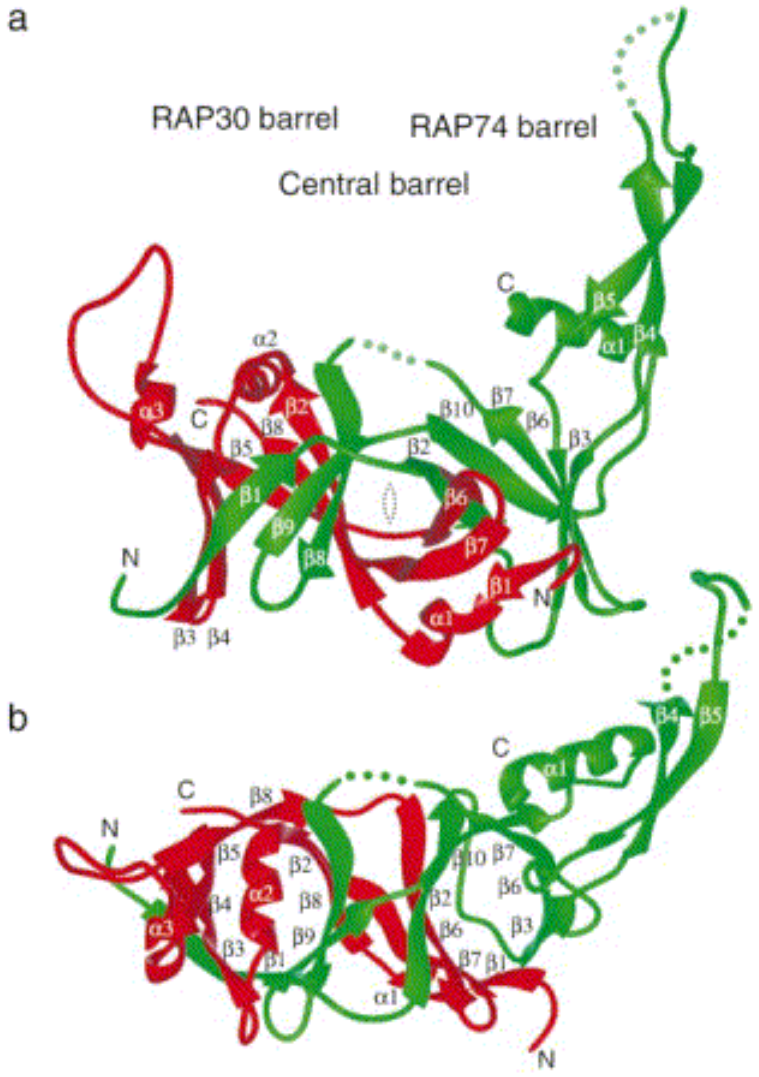


- Functions:

- Recruitment of Pol II to the existing DNA-TFIID-B complex,
- Positioning Pol II over the start site
- Binding to the non-template DNA strand.
- TFIIF also reduces non-specific binding of RNA pol II to DNA.

The 3-D structure of TFIIF

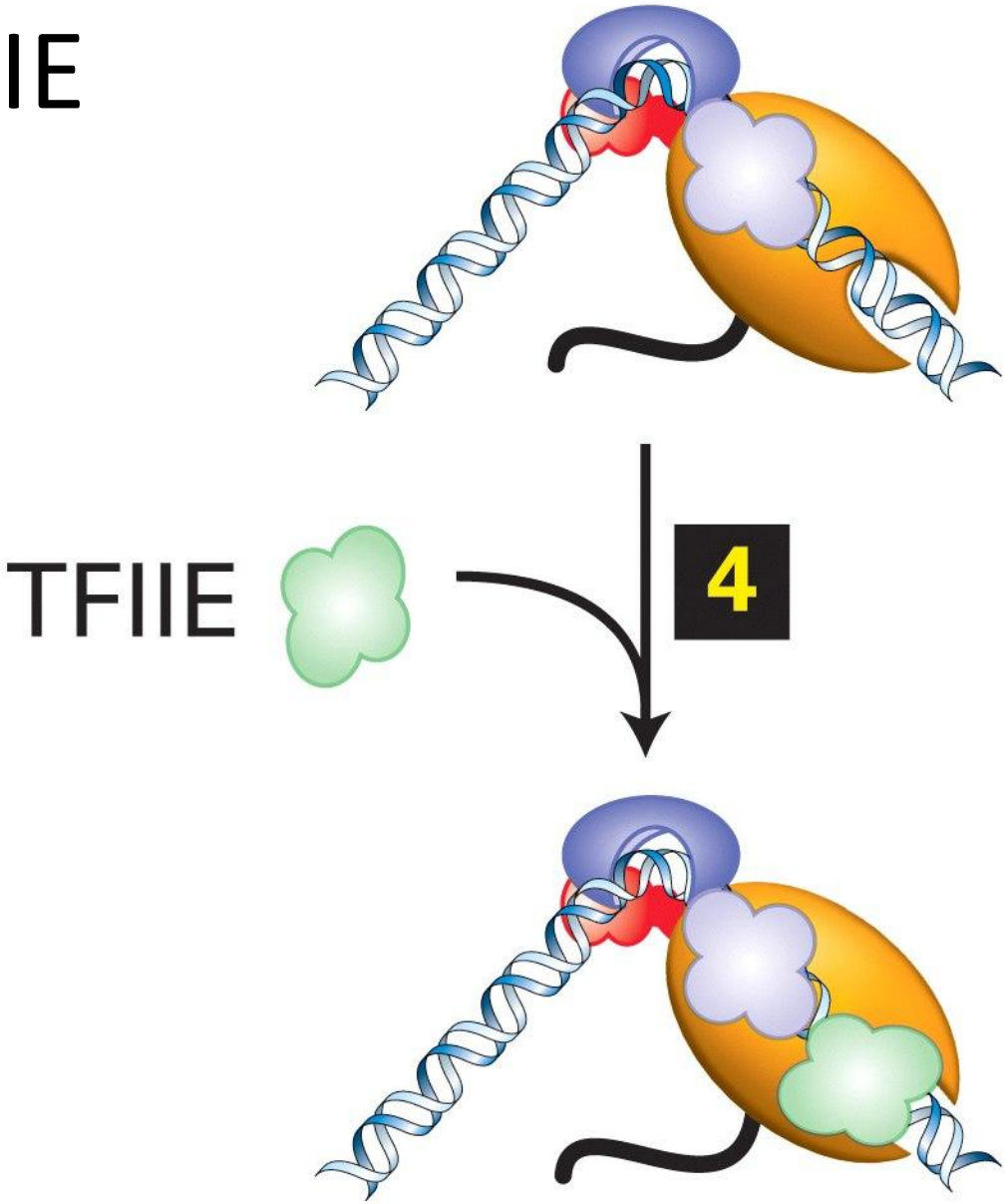
- TFIIF is a heterotetramer, made from two subunits of RAP30 and two subunits of RAP74 proteins
- RAP30-RAP74 dimer within the complex structure has an unusual triple-barrel fold



TFIIE

- TFIIE is a heterotetrameric protein ($\alpha_2\beta_2$)
- Functions:

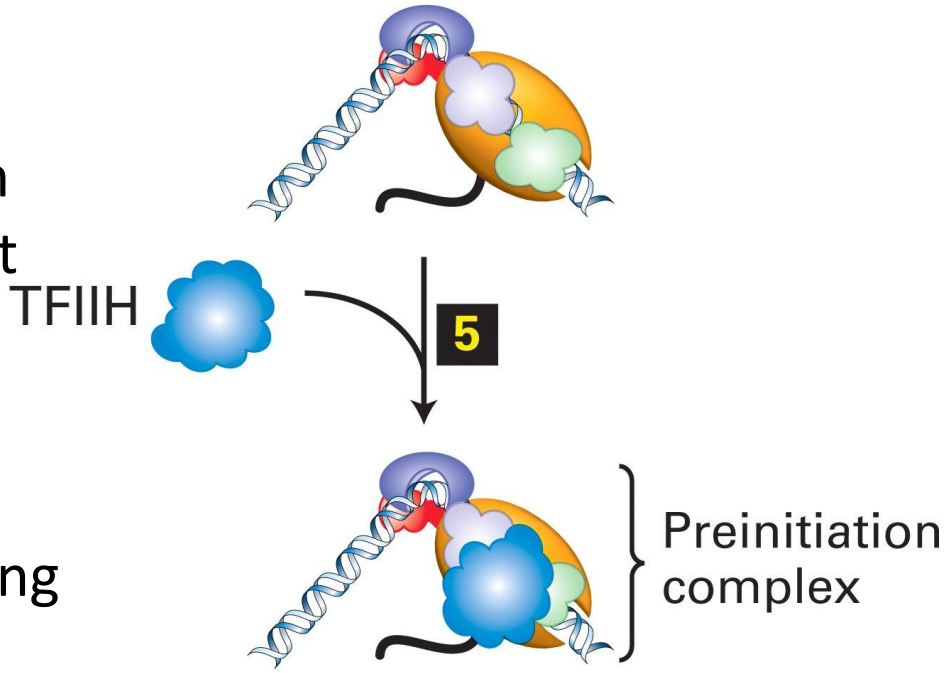
- TFIIE appears to create the docking site for next transcription factor, TFIIH.
- TFIIE also modulates TFIIH enzymatic activities
- In addition TFIIE enhances promoter melting.



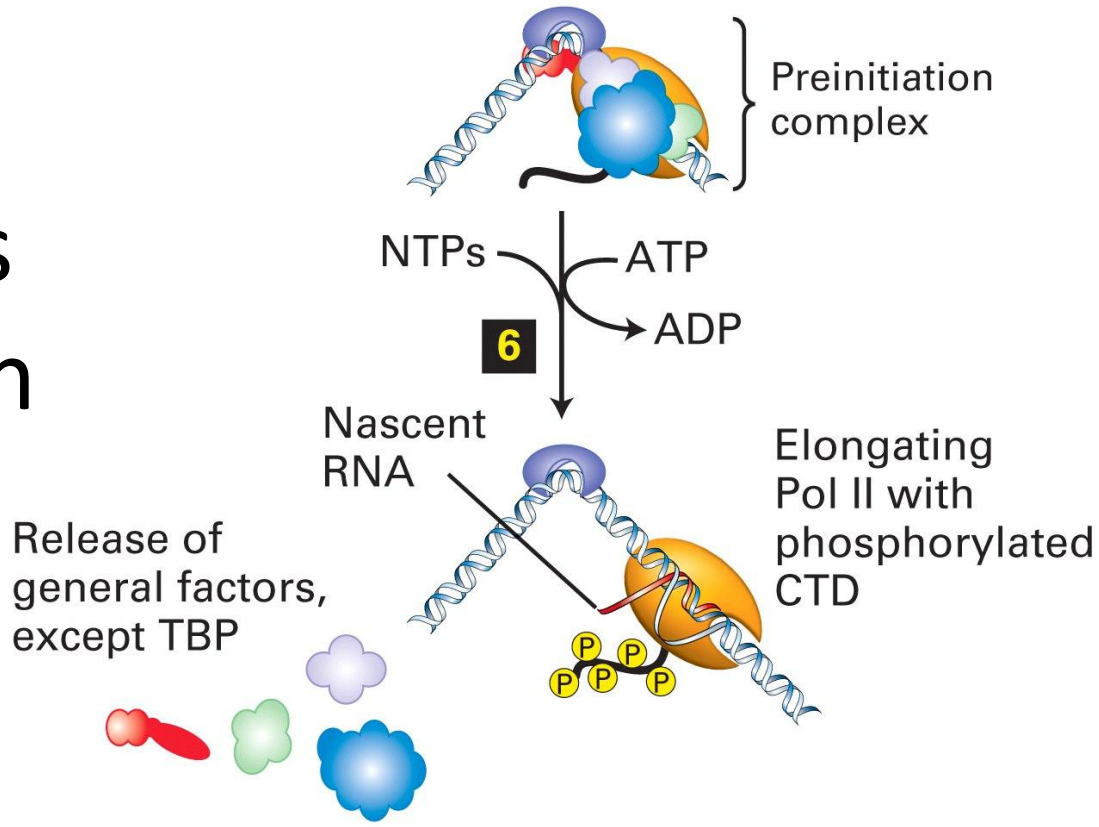
- TFIIH is a multimeric protein, composed of 9 subunits, some of them with distinct enzymatic activities
- Functions:

- TFIIH has a helicase activity, which unwinds the DNA duplex at a start site, allowing Pol II to bind to the template strand.
- TFIIH also has a kinase ativity, it phosphorylates PolII in the begining of elongation
- Other TFIIH subunits have been shown to recruit DNA-repairing enzymes if polymerase reaches damaged region in DNA and gets stalled

TFIIH



Early events in elongation



- As Pol II transcribes away from the start site subunit of TFIIH phosphorylates the Pol II CTD, which results in promoter escape.
- General factors get released

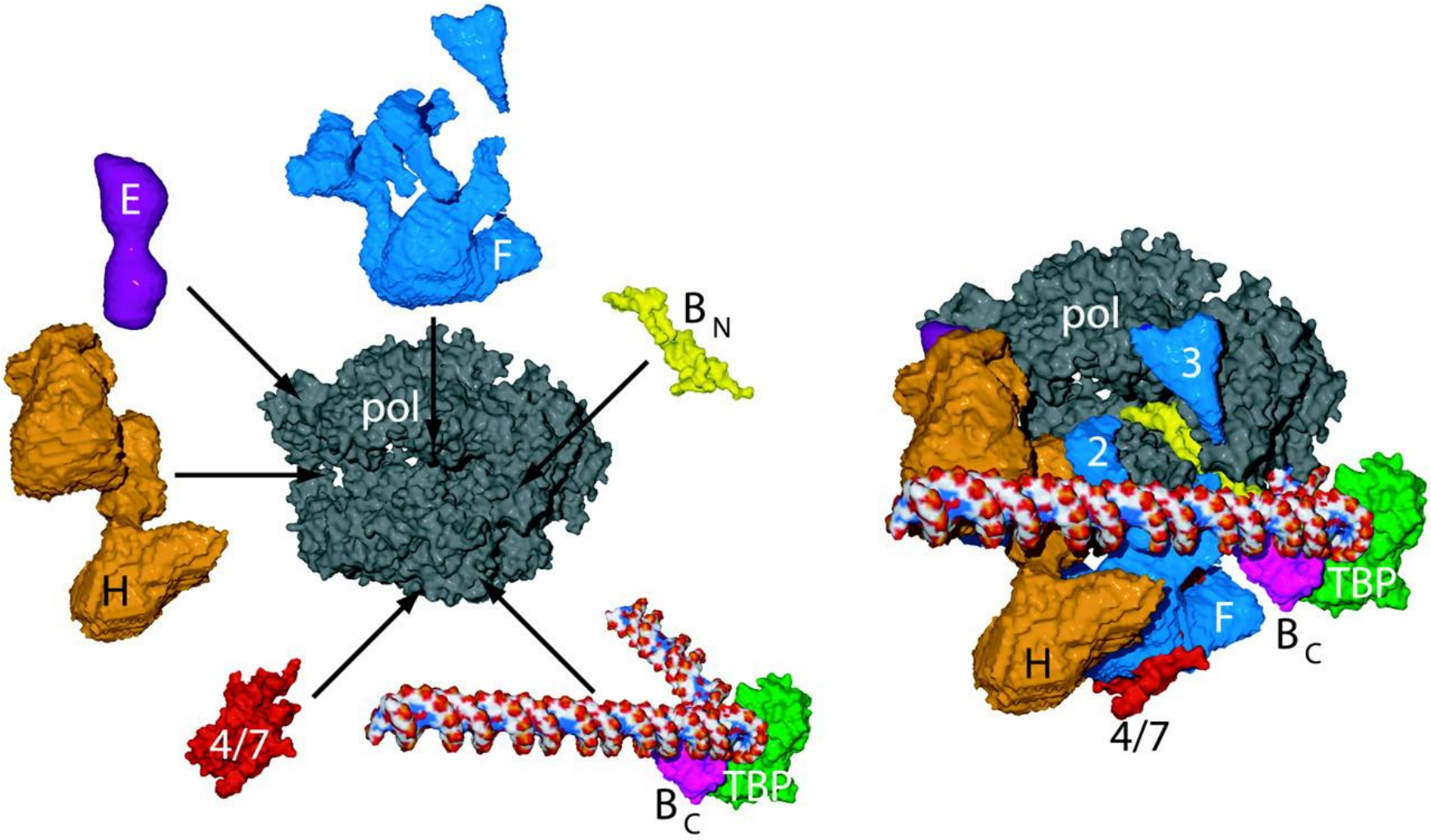
TFIIA

- For transcription *in vivo*, another factor TFIIA is required
- The function of TFIIA is somewhat unclear, but it might help the other factors to bind.
- TFIIA has also shown to have some anti-repressor functions
- TFIIA is not required for transcription *in vitro.*

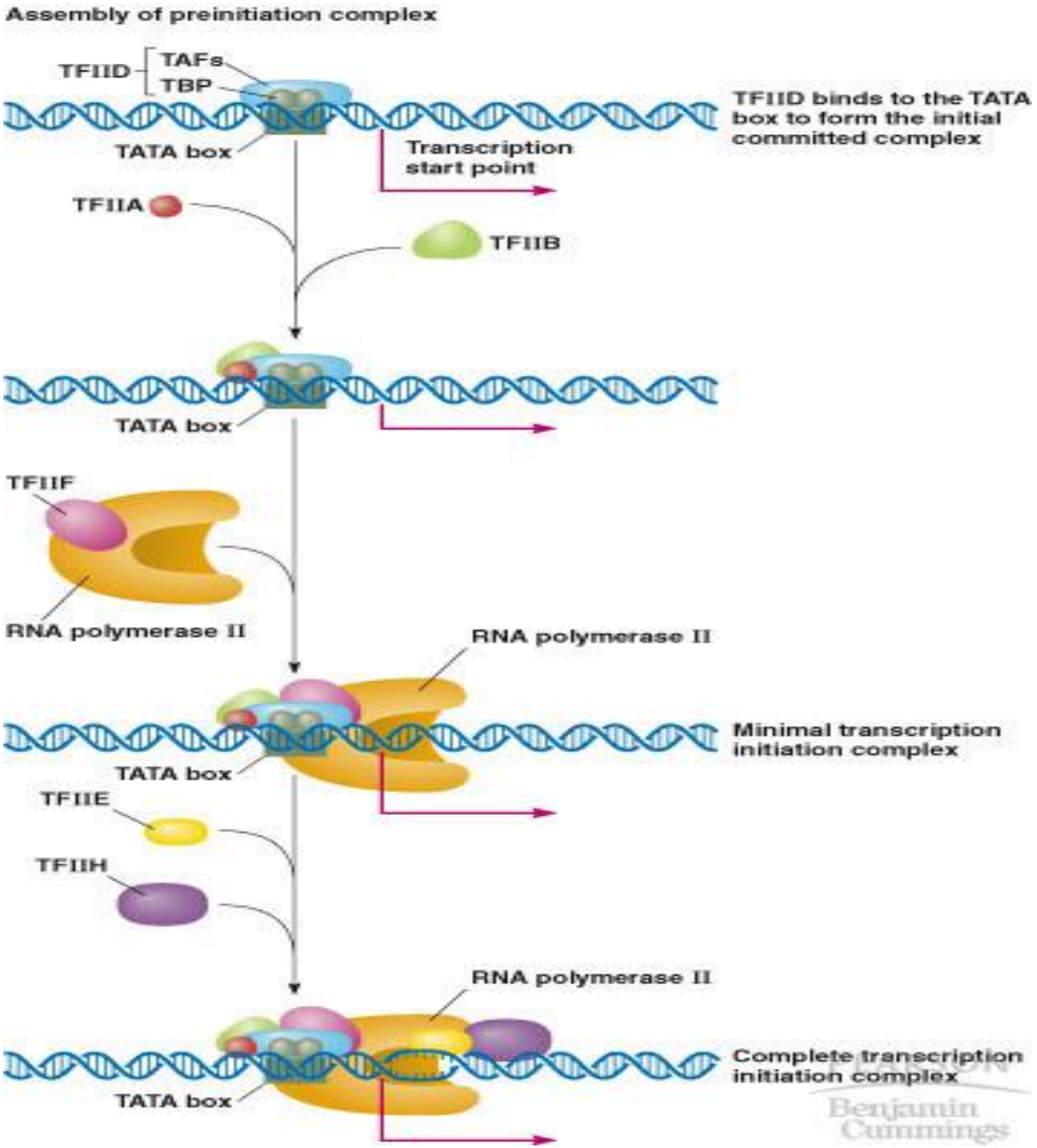
TAFs

- Apart from TBP, TFIID has 13 TAF (**T**BP **a**ssociated **f**actors) subunits
- Some of them seem to be necessary for transcription initiation from promoters, lacking the TATA box
- Other TAFs have been shown to be tissue-specific coactivators
- TAF subunits also interact with other GTFs therefore stabilyzing the complex.

The 3-D structure of Pol II holoenzyme



Transcription Initiation: Assembly Of the initiation machinery



The mRNA molecule has 3 main parts in both prokaryotes and eukaryotes?

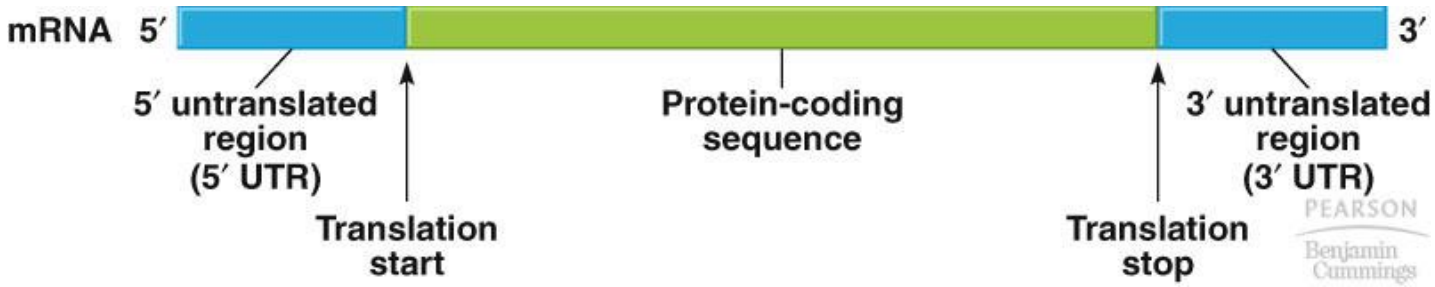
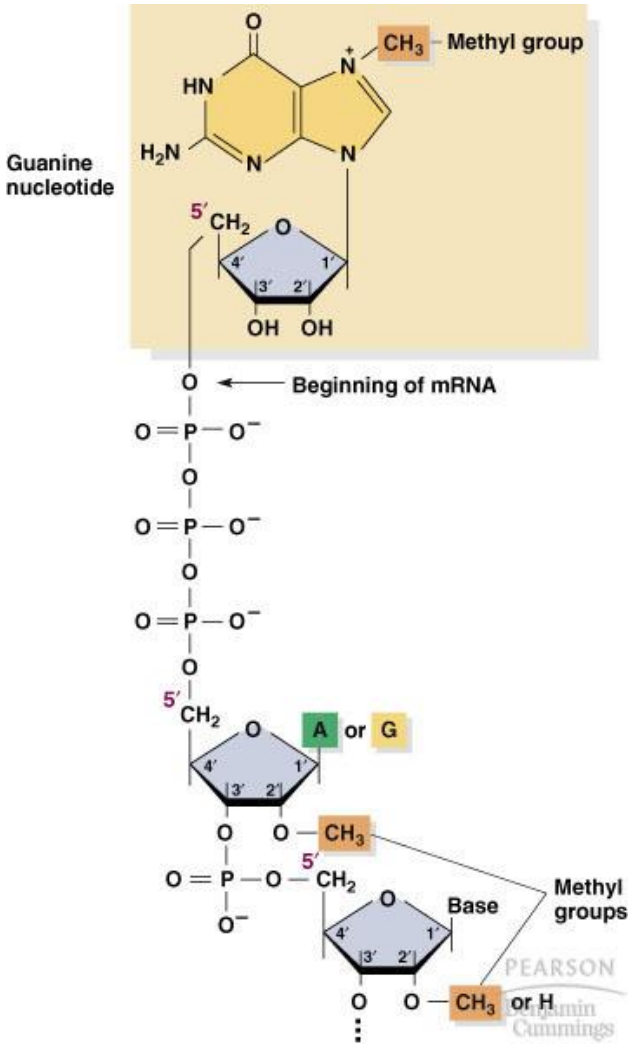


Figure 5.8 shows the general structure of the mature, biologically active mRNA as it exists in both prokaryotic and eukaryotic cells.

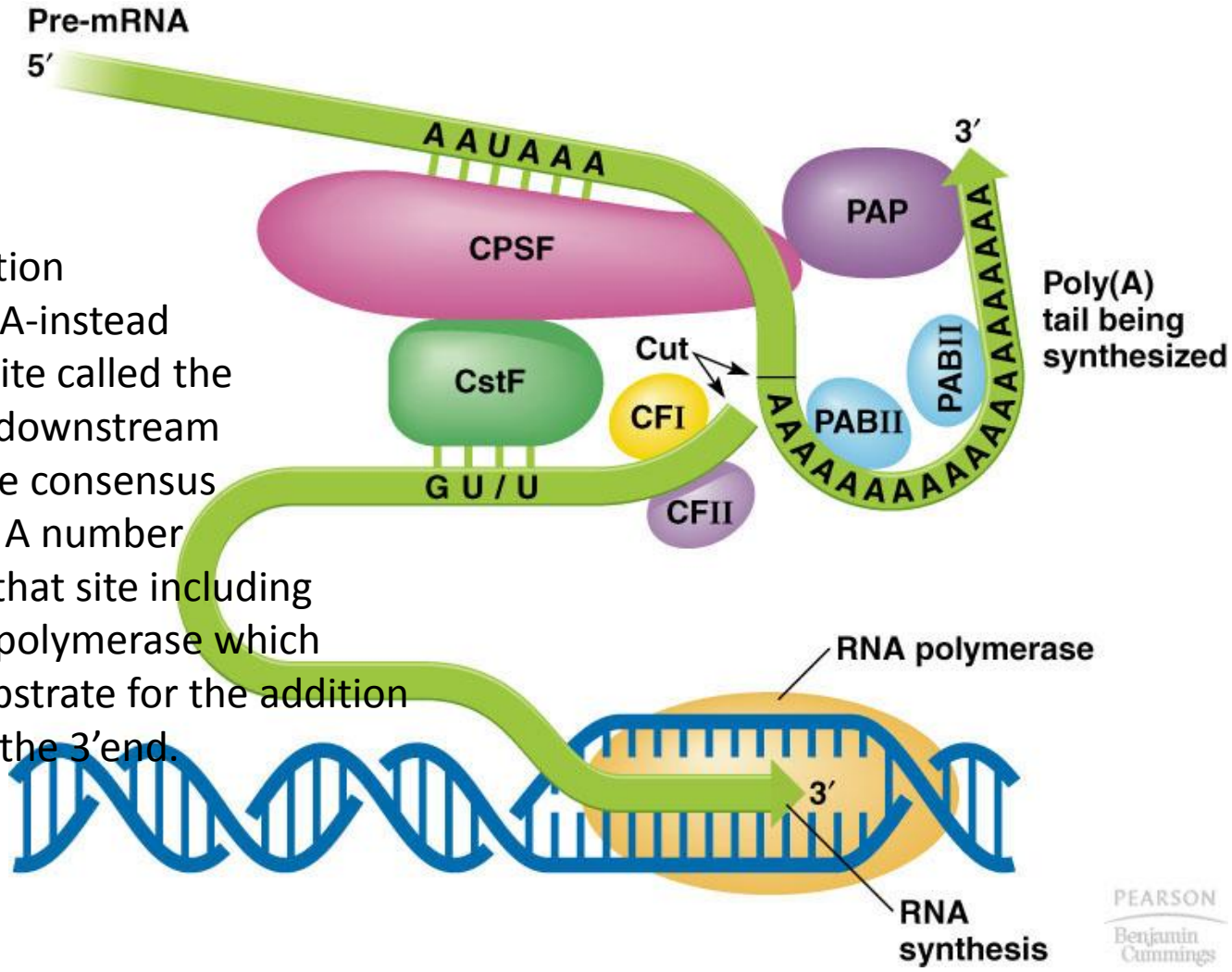
Unlike prokaryotic mRNAs, eukaryotic mRNAs are modified at both the 5' and 3' ends and they have introns which need to be removed to form a mature mRNA



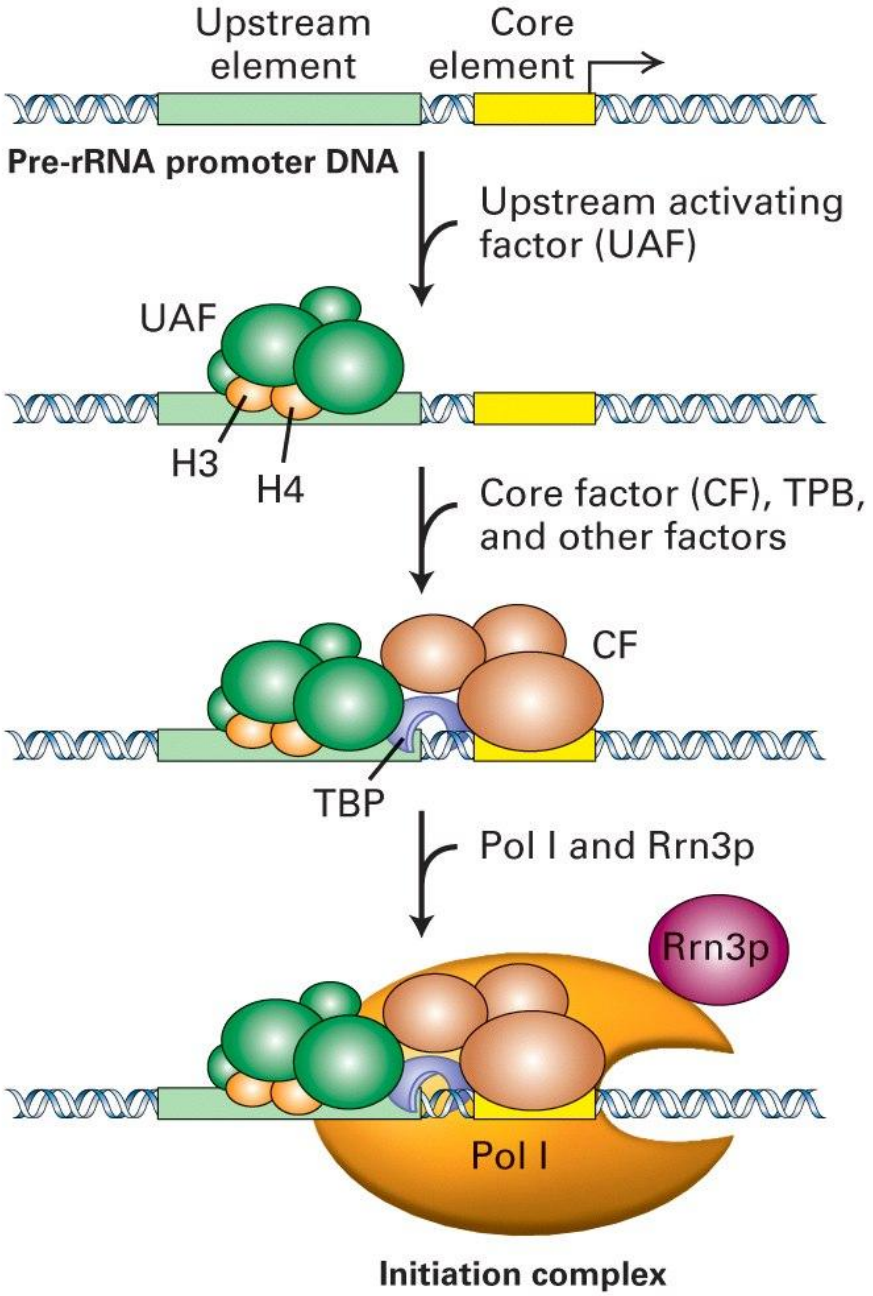
At what end do you find the CAP?

At what end do you find the poly A?

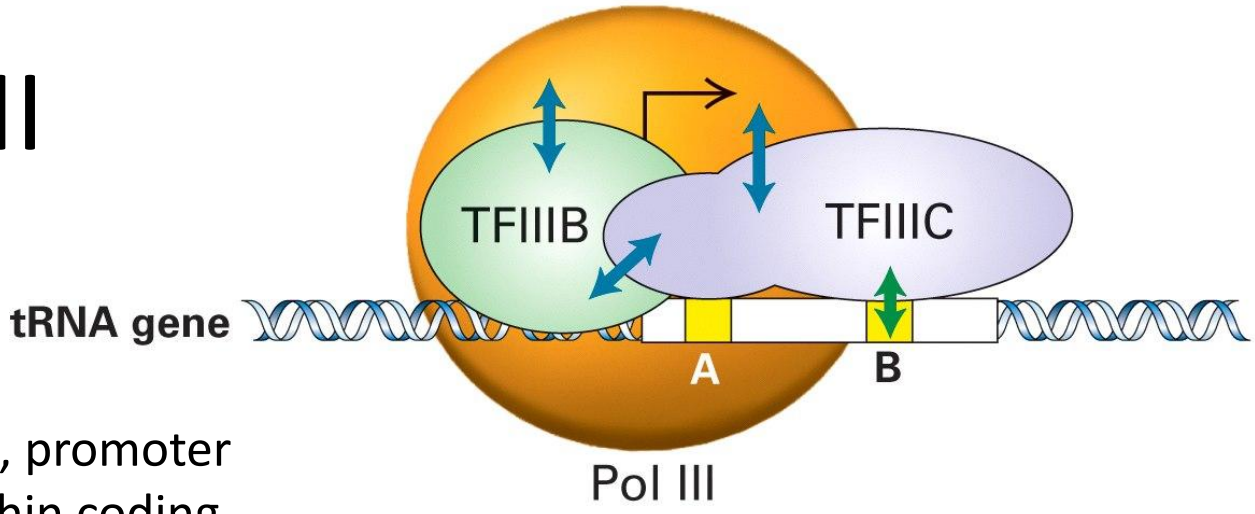
There is no termination
Sequence in the DNA-instead
It continues past a site called the
Poly A site which is downstream
10 to 30 bases of the consensus
Sequence AAUAAA. A number
Of proteins bind at that site including
The enzyme poly A polymerase which
Uses an ATP as a substrate for the addition
Of A nucleotides to the 3'end.



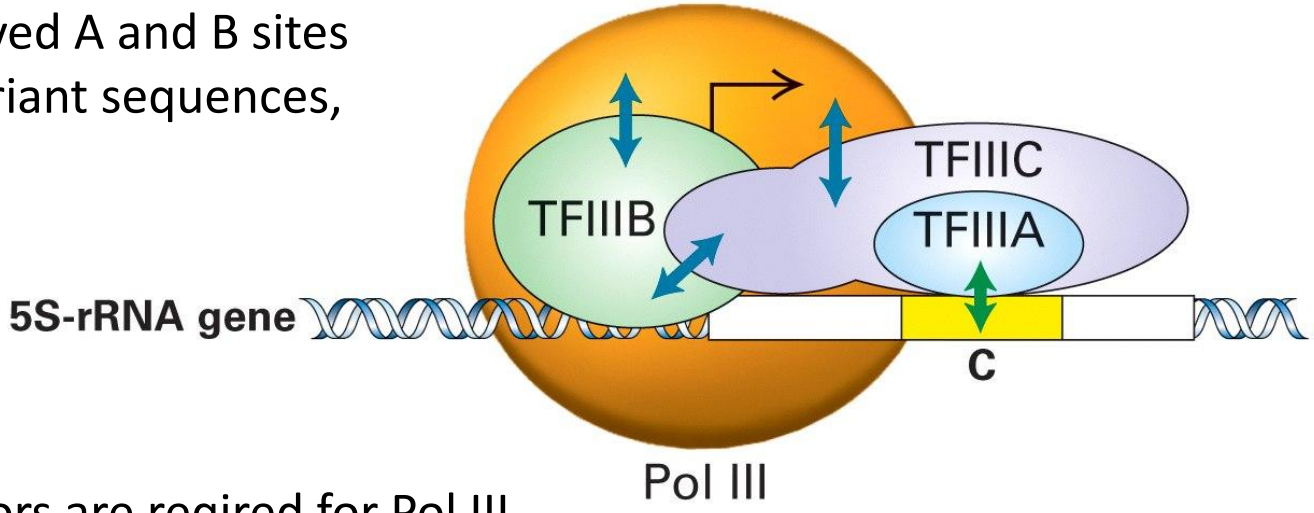
RNA Pol I


Upstream element
Core element
Pre-rRNA promoter DNA
Upstream activating factor (UAF)
UAF
H3
H4
Core factor (CF), TPB, and other factors
CF
TBP
Pol I and Rrn3p
Rrn3p
Pol I
Initiation complex

RNA pol III



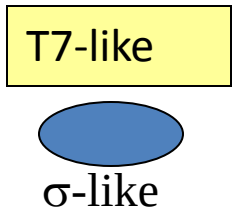
Unlike other polymerases, promoter sequences lie entirely within coding sequence. The conserved A and B sites also code for two invariant sequences, observed in all tRNAs

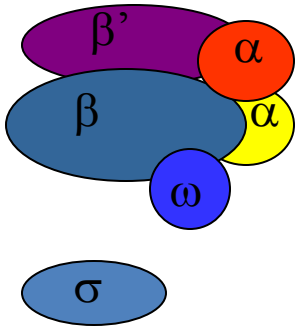


Three initiation factors are reqired for Pol III initiation. TFIIIA is required only for 5S rRNA transcription. TFIIIB has three subunits, one of them similar to TFIIB and another one being TBP (same as for Pol I and II)

Mitochondrial RNA polymerase

- Encoded in nuclear RNA and transported to mitochondrial matrix
- Contains only two subunits
- One subunit similar to bacteriophage T7 RNA polymerase
- Other subunit similar to bacterial σ factor





Chloroplast RNA polymerase

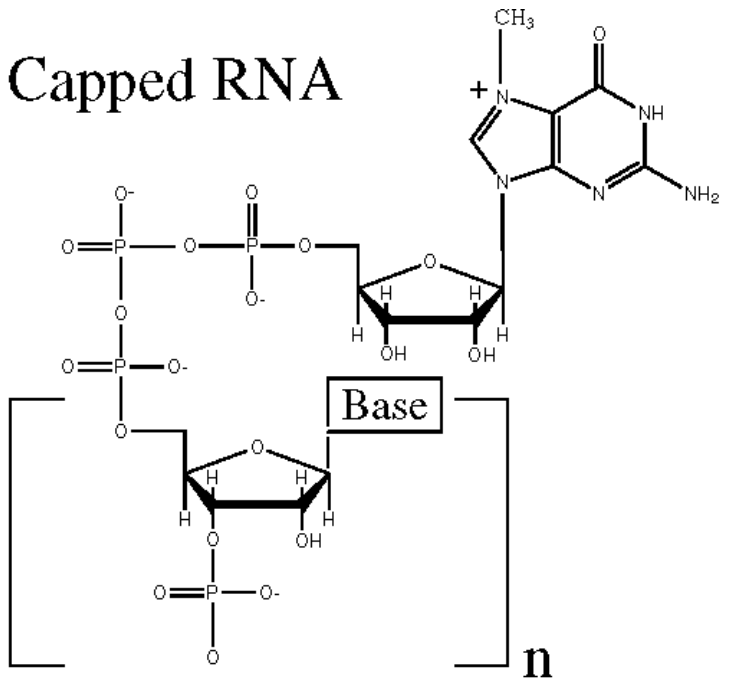
- Encoded by chloroplast genome
- Contains considerable homology to bacterial α, β and β' RNA pol subunits
- No any σ-like factors or general transcription factors

After Transcription

- In prokaryotes, the RNA copy of a gene is messenger RNA, ready to be translated into protein. In fact, translation starts even before transcription is finished.
- In eukaryotes, the primary RNA transcript of a gene needs further processing before it can be translated. This step is called “RNA processing”. Also, it needs to be transported out of the nucleus into the cytoplasm.
- Steps in RNA processing:
  - 1. Add a cap to the 5’ end
  - 2. Add a poly-A tail to the 3’ end
  - 3. splice out introns.

Capping

- RNA is inherently unstable, especially at the ends. The ends are modified to protect it.
- At the 5’ end, a slightly modified guanine (7-methyl G) is attached “backwards”, by a 5’ to 5’ linkage, to the triphosphates of the first transcribed base.
- At the 3’ end, the primary transcript RNA is cut at a specific site and 100-200 adenine nucleotides are attached: the poly-A tail. Note that these A’s are not coded in the DNA of the gene.

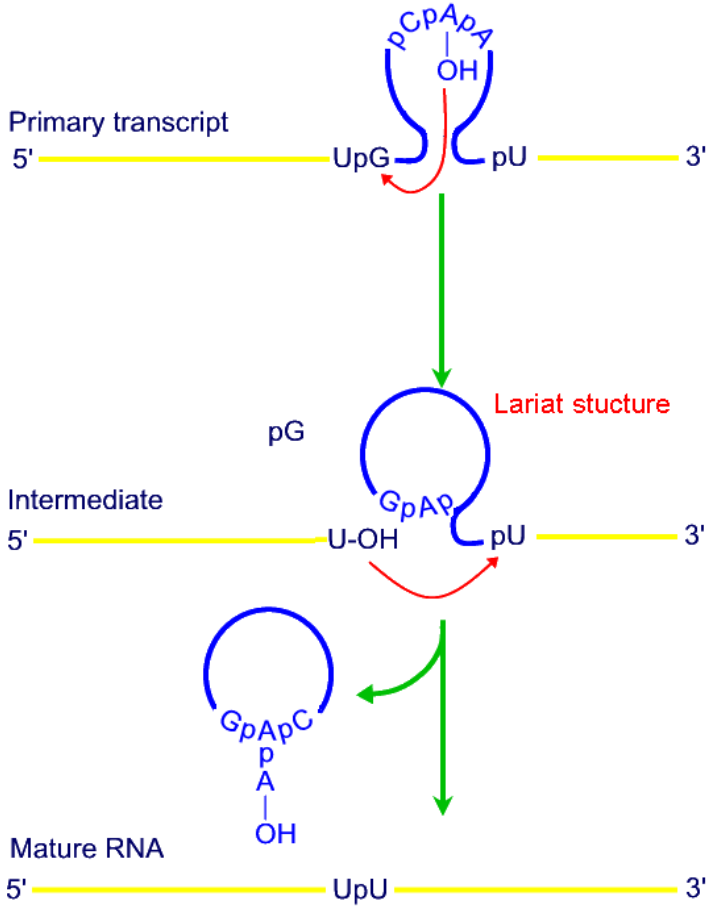


Introns

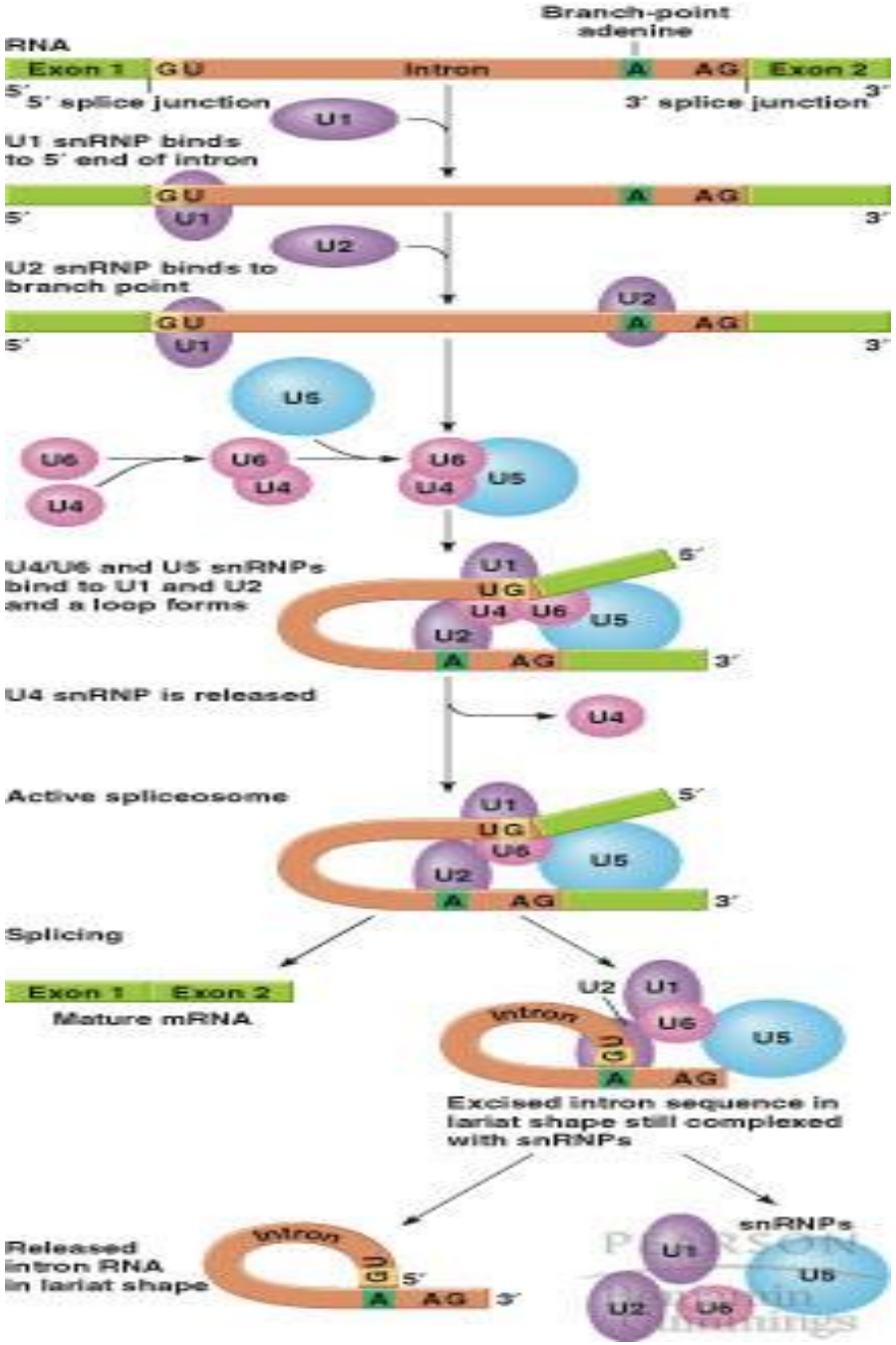
- Introns are regions within a gene that don't code for protein and don't appear in the final mRNA molecule. Protein-coding sections of a gene (called exons) are interrupted by introns.
- The function of introns remains unclear. They may help is RNA transport or in control of gene expression in some cases, and they may make it easier for sections of genes to be shuffled in evolution. But , no generally accepted reason for the existence of introns exists.
- There are a few prokaryotic examples, but most introns are found in eukaryotes.
- Some genes have many long introns: the dystrophin gene (mutants cause muscular dystrophy) has more than 70 introns that make up more than 99% of the gene's sequence. However, not all eukaryotic genes have introns: histone genes, for example, lack introns.

Intron Splicing

- Introns are removed from the primary RNA transcript while it is still in the nucleus.
- Introns are “spliced out” by RNA/protein hybrids called “spliceosomes”. The intron sequences are removed, and the remaining ends are re-attached so the final RNA consists of exons only.

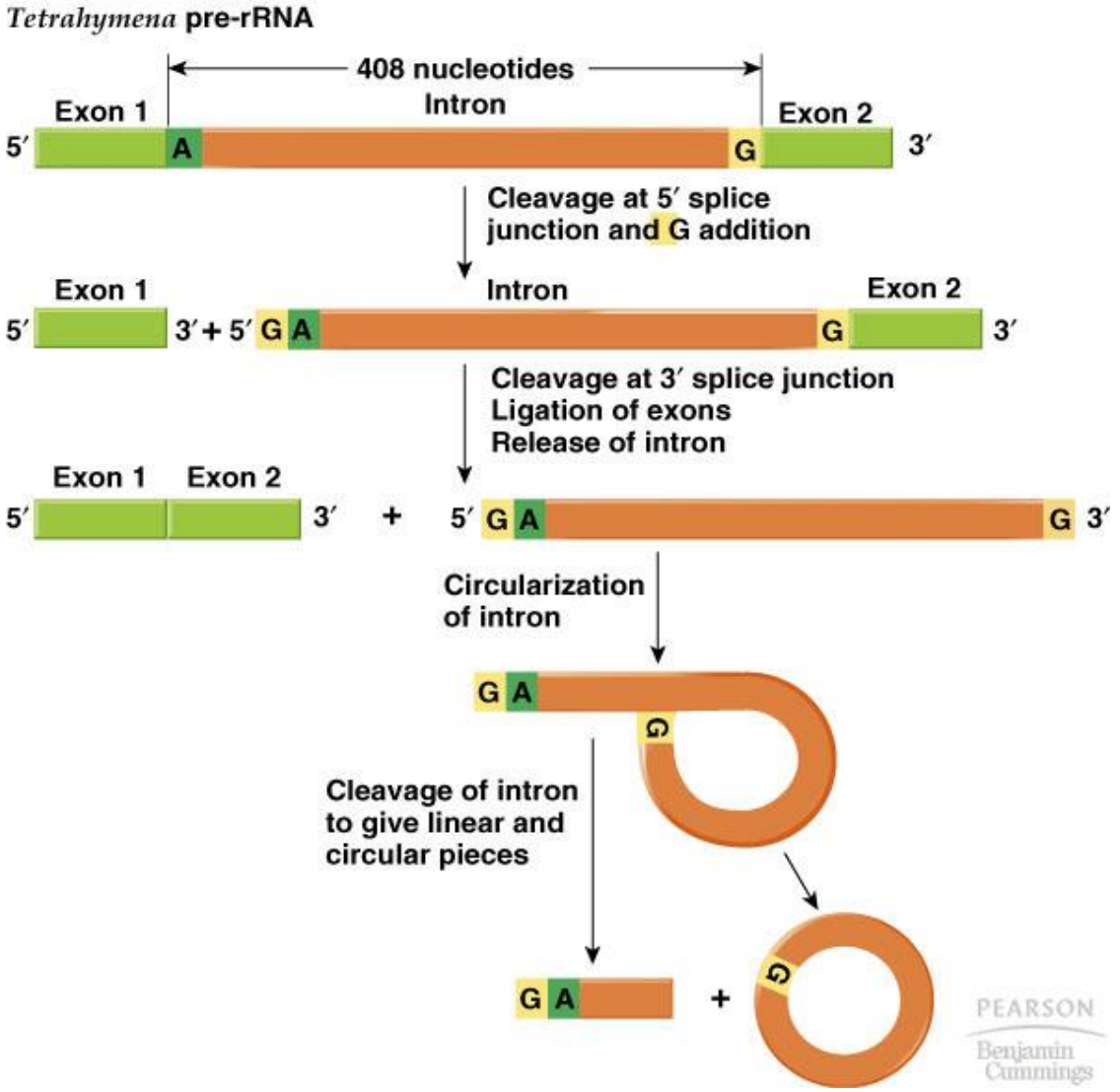


At the end 5’ of an intron is the sequence **GU,** and at the 3’ end is the sequence **AG**. Near the end of the intron is an **A** nucleotide located within the branch-point sequence, which, in mammals, is YNCUR*A*Y, where Y=pyrimidine, N=any base, R=purine, A=adenine, and, in yeast, is UACUA*A*C. (The italic A in each sequence is where the 5’ end of the intron bonds). With the aid of snRNPs, intron removal begins with a cleavage at the first exon–intron junction. The G at the released 5’ of the intron folds back and forms an unusual 2’-5’ bond with the A of the branch-point sequence. This reaction produces a lariat-shaped intermediate. Cleavage at the 3’ intron–exon junction and ligation of the two exons completes the removal of the intron.



Interestingly, some eukaryotic species (a protozoan) have specific RNAs that self-splice; in this case a rRNA

Cleavage at the splice junction takes place first, accompanied by the addition of a G nucleotide to the end of the intron. Cleavage at the splice junction releases the intron; the two exons are spliced. The intron now circularizes by the nucleotide bonding to an internal nucleotide, producing a lariat-shaped molecule. Cleavage of the lariat molecule produces a circular RNA molecule and a linear molecule. This activity is not considered enzymatic



Summary of RNA processing

- In eukaryotes, RNA polymerase produces a “primary transcript”, an exact RNA copy of the gene.
- A cap is put on the 5’ end.
- The RNA is terminated and poly-A is added to the 3’ end.
- All introns are spliced out.
- At this point, the RNA can be called messenger RNA. It is then transported out of the nucleus into the cytoplasm, where it is translated.