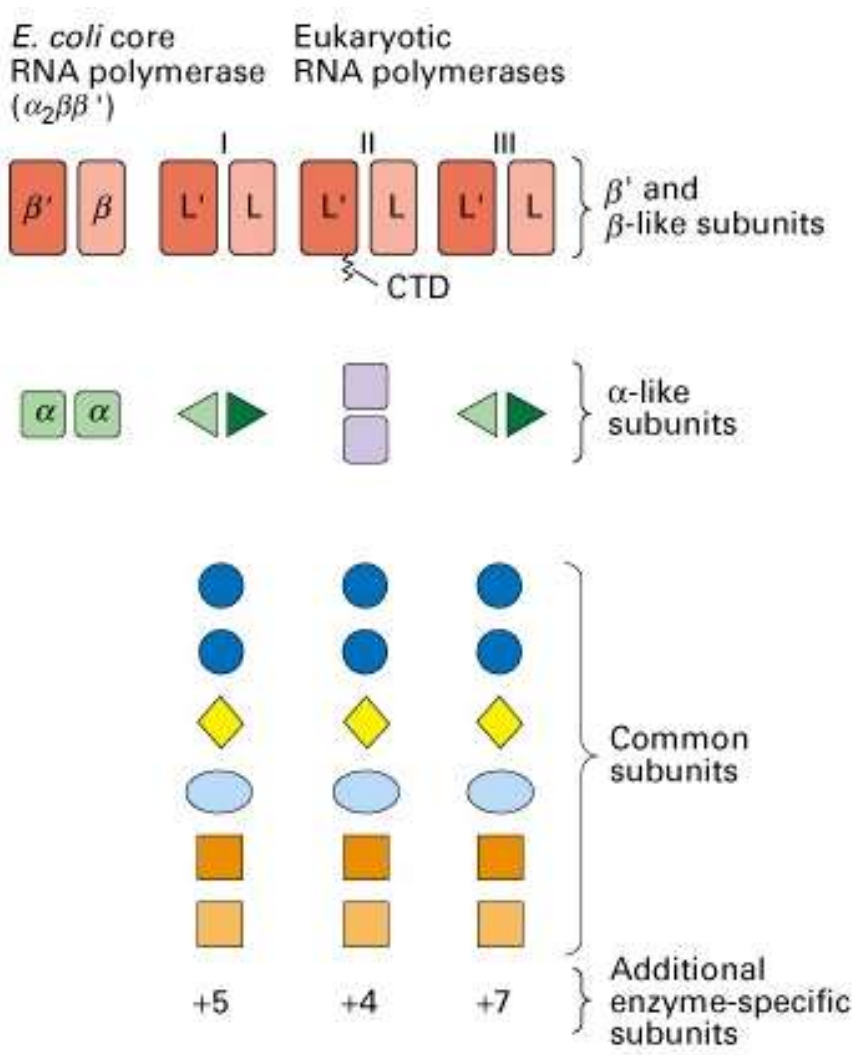
Eukaryotic transcription activation, RNA processing, Ribozymes and the RNA World.

Liam Keegan

31.3.25

Liam.Keegan@ceitec.muni.cz

Eukaryotic RNA polymerases are similar to that of *E. coli* but have 12 subunits.



Which is the original and older type?

- The simpler bacterial RNA polymerase working with sigma factor?
- Or the more complicated eukaryotic polymerase with more subunits?

ARCHAEA

- Three domains of Life- Archaea, Bacteria, Eukarya
- Archaea were first classified as a separate group of prokaryotes in 1977 by Carl Woese and George E. Fox
- *Archaea* share many characteristics with both *Bacteria* and *Eukarya*
- *Archaea* are split into two major phyllums:-

  Crenarchaeota

  Euryarchaeota

Grand Prismatic Spring of Yellowstone National Park

- Best known for the "extremophiles" – Archaea which thrive in extremely harsh environments

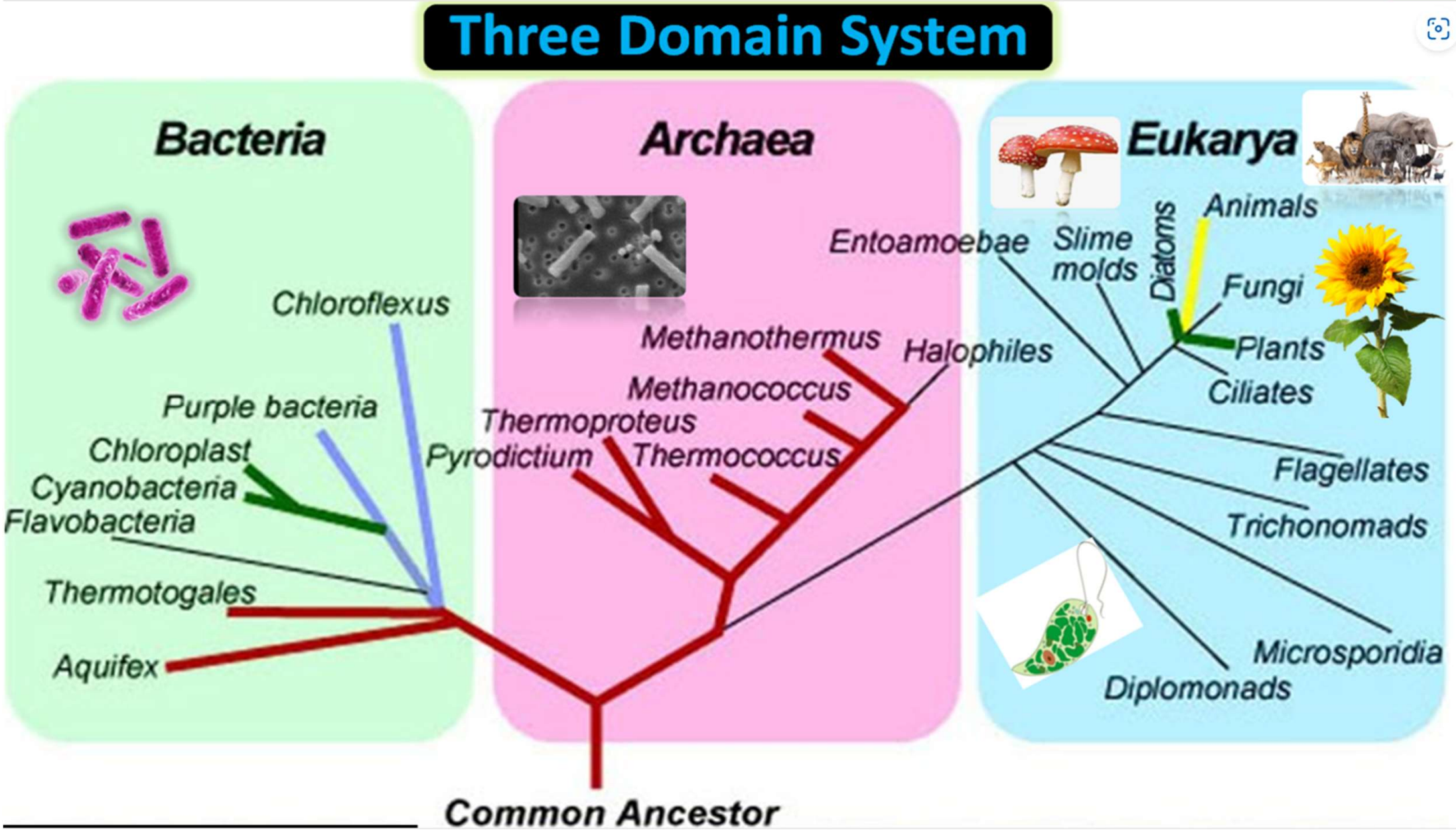
- Thermophiles – thrive at 60-80°C (>176°F!)

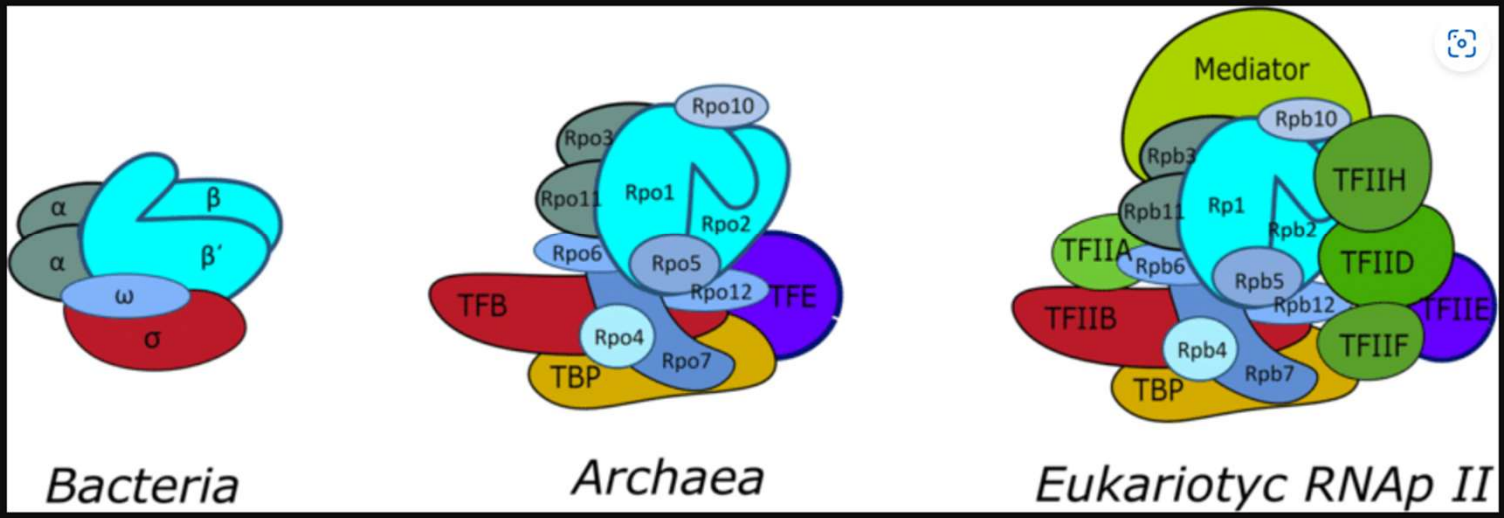
Acidophiles – thrive at pH at or below pH 3

Xerophiles – grow in extremely dry conditions

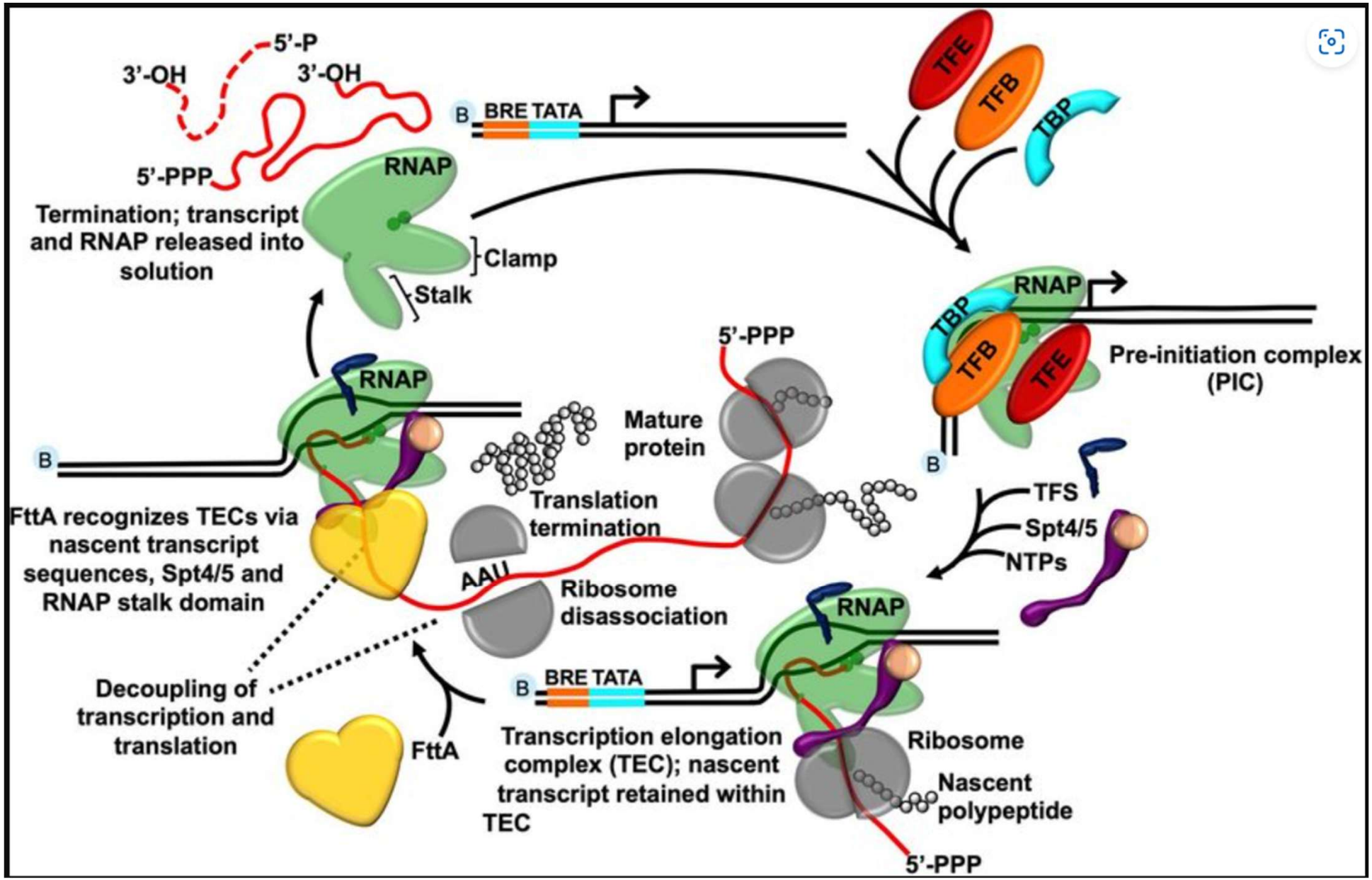
Halophiles – require extremely high concentrations of salt.

Three Domains of Life



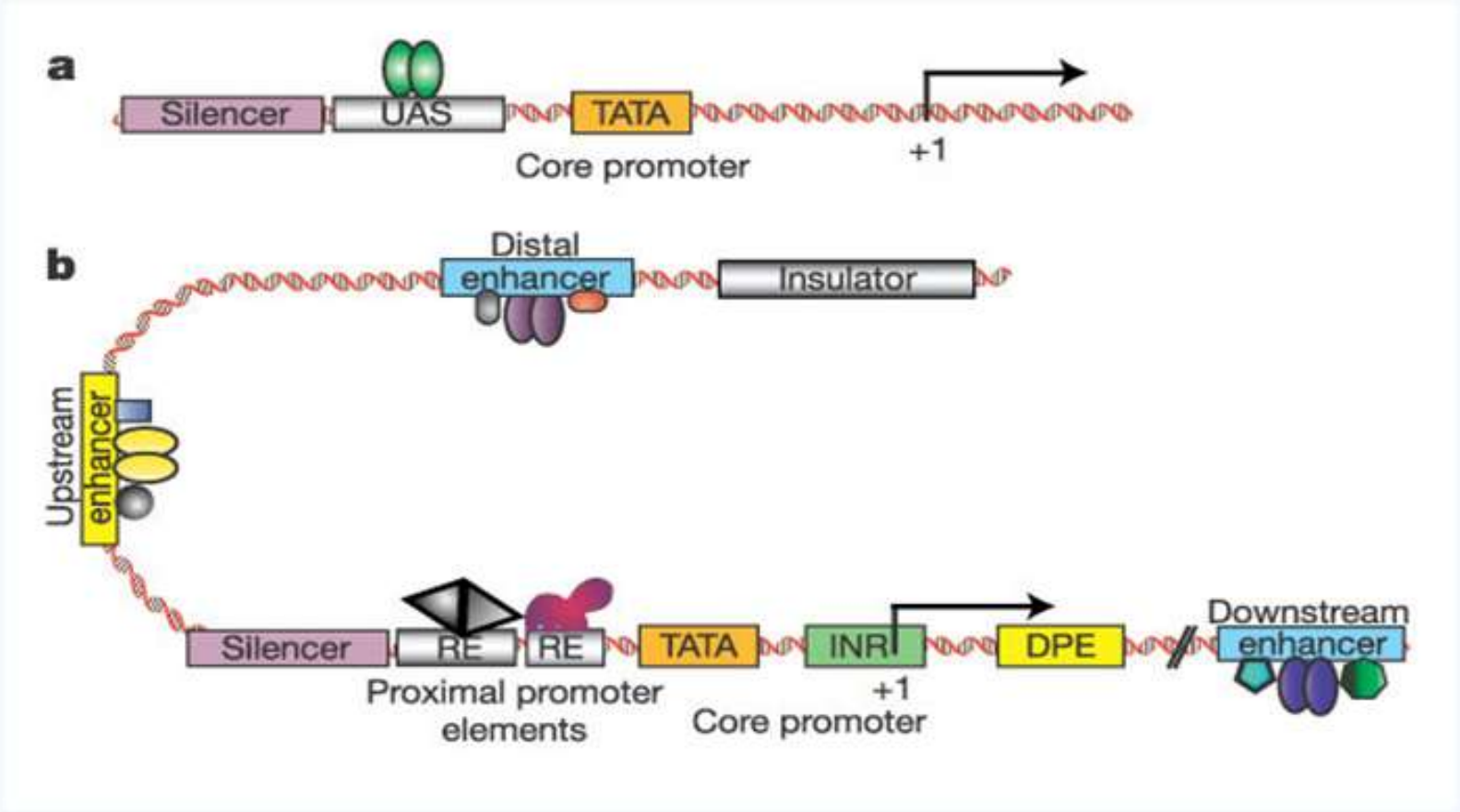


Transcription in Archeae is Eukaryote-like



How have eukaryotic sequence-specific DNA-binding proteins been identified and characterized?

Regulatory elements controlling RNA Pol. II transcription in yeast and higher eukaryotes.



- Promoters
- Proximal regulatory elements
- Enhancers

25.7 Response Elements Are Recognized by Activators - promoter deletion analyses

- Response elements may be located in promoters or enhancers. Proximal regulatory elements were targets of the earliest studies.

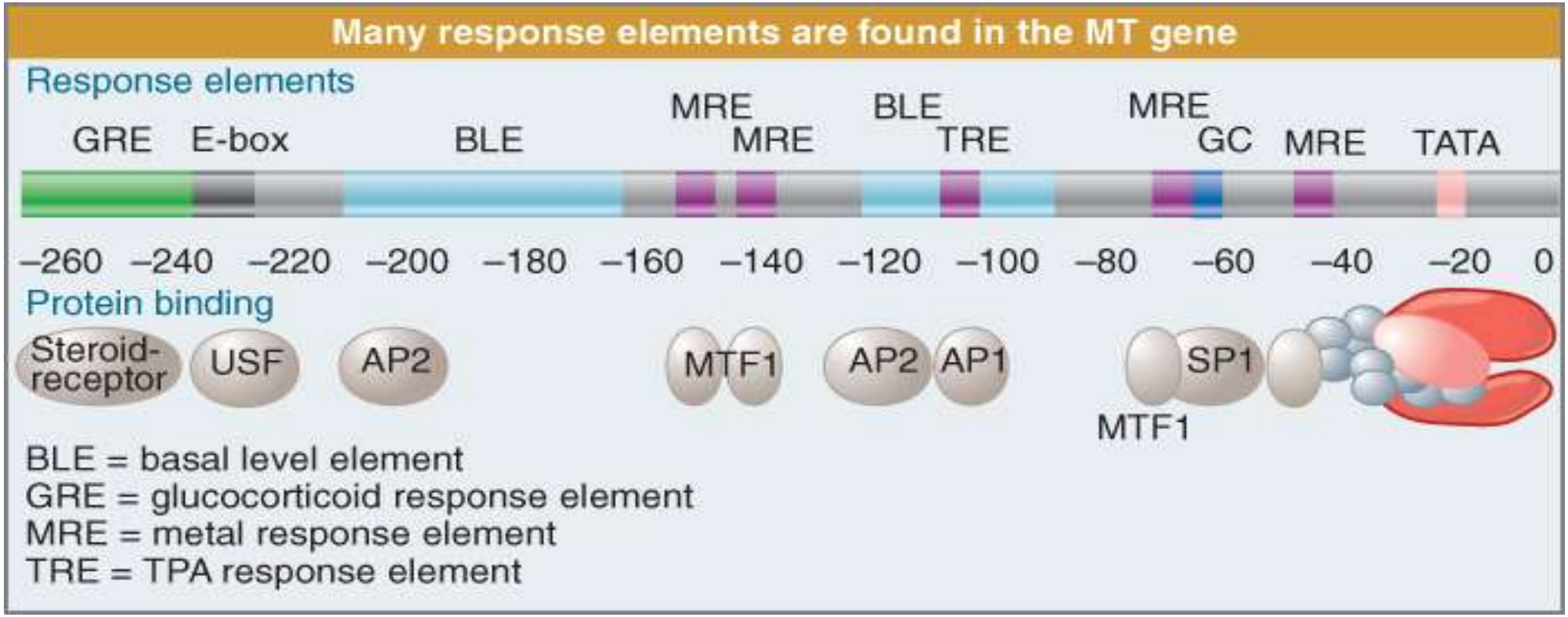
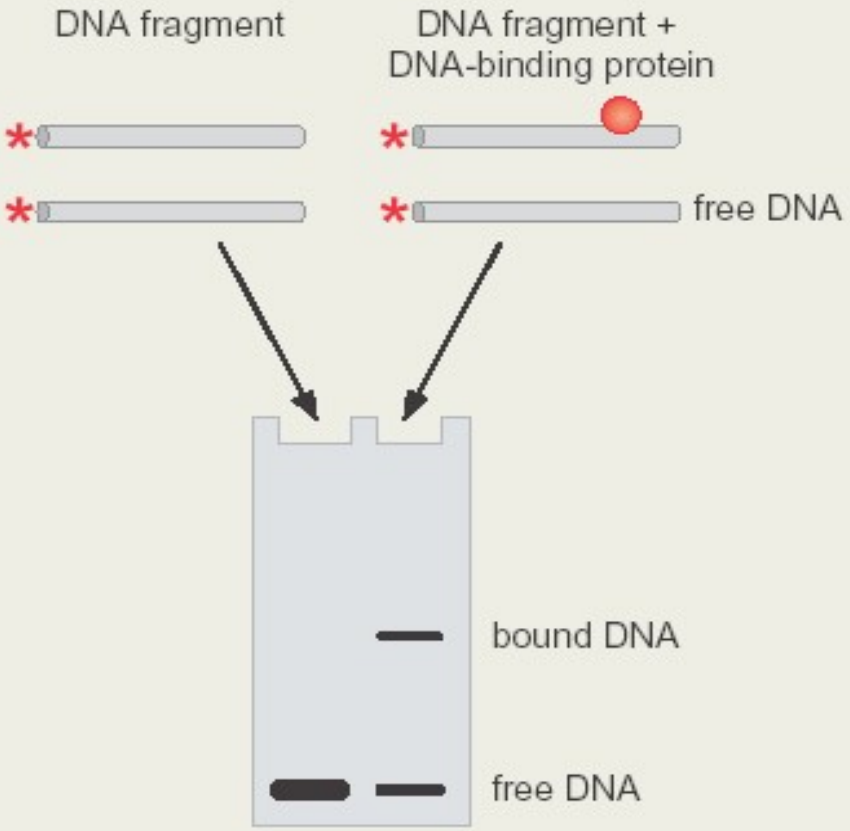


Figure 25.11

Bandshift assay or Gel mobility shift assay.

Many gene regulators from mammalian cells were identified and purified using this assay.

(also Gel Retardation assay, Electrophoretic Mobility Shift Assay (EMSA)).



BOX 16-1 FIGURE 2 Gel mobility shift assay. The principle of the mobility shift assay is shown schematically. A protein is mixed with radiolabeled probe DNA containing a binding site for that protein. The mixture is resolved by acrylamide gel electrophoresis and visualized using autoradiography. DNA not mixed with protein runs as a single band corresponding to the size of the DNA fragment (left lane). In the mixture with the protein, a proportion of the DNA molecules (but not all of them at the concentrations used) binds the DNA molecule. Thus, in the right-hand lane, there is a band corresponding to free DNA, and another corresponding to the DNA fragment in complex with the protein.

Gel is non-denaturing and buffer is mild to avoid disrupting complex.
Gel matrix ‘traps’ protein DNA complex keeping components together to prevent disassociation.

Bandshift assay identifies DNA-binding proteins in crude extracts.

- DNase1 footprinting required pure proteins but bandshift assays could detect different scarce DNA-binding proteins even in crude nuclear extracts.

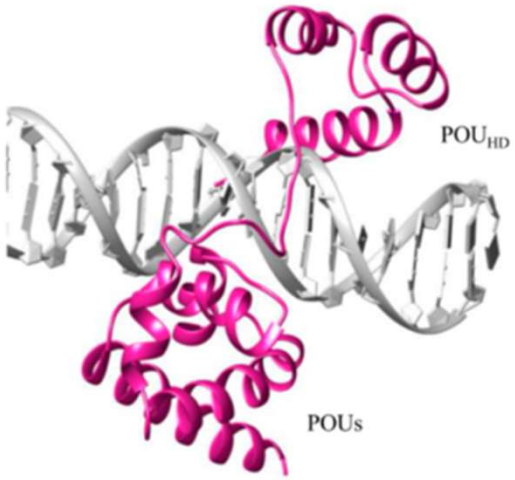
- If the probe DNA fragment comes from an enhancer then a point mutation that inactivates the enhancer should also prevent DNA binding by the key regulatory protein in the extract. (Some proteins bind DNA ends or other sites in the probe DNA)

- The specific bandshift provides a convenient assay to purify the regulatory protein.

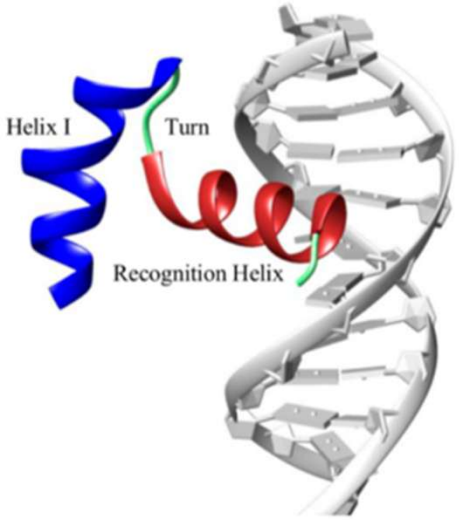
Examples of important transcription regulators identified after purification from nuclear extracts.

- **Fos/jun, AP1**. oncoproteins studied by Tom Curran and others. Steve MacKnight showed these proteins contain a leucine zipper.
- **Nf-kappaB** (Baltimore lab, MIT) Nuclear factor binding the immunoglobulin kappa B enhancer. Important protein in innate immunity and response to viral infection.
- **SRF** (Maniatis lab and Richard Treisman in London) Serum response factor activates genes involved in growth of cells in tissue culture.
- **SP1** (Tjian lab)

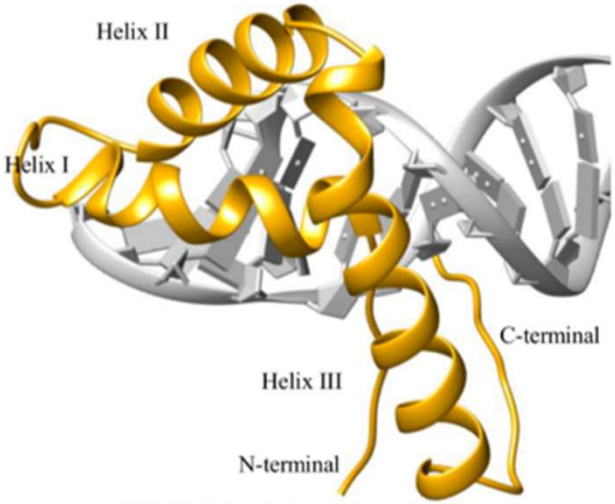
Sequence-specific DNA-binding domains in transcriptional regulators



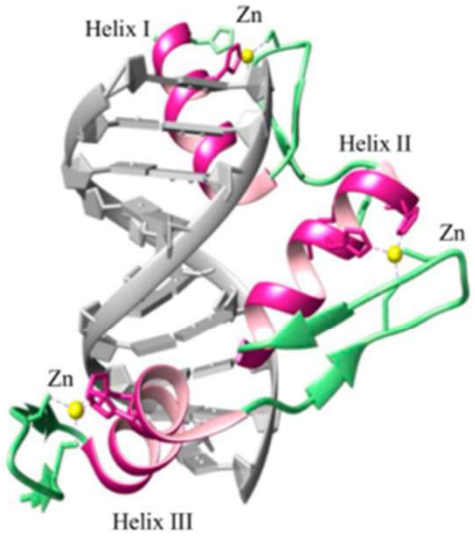
HTH-HD domain (3l1p)



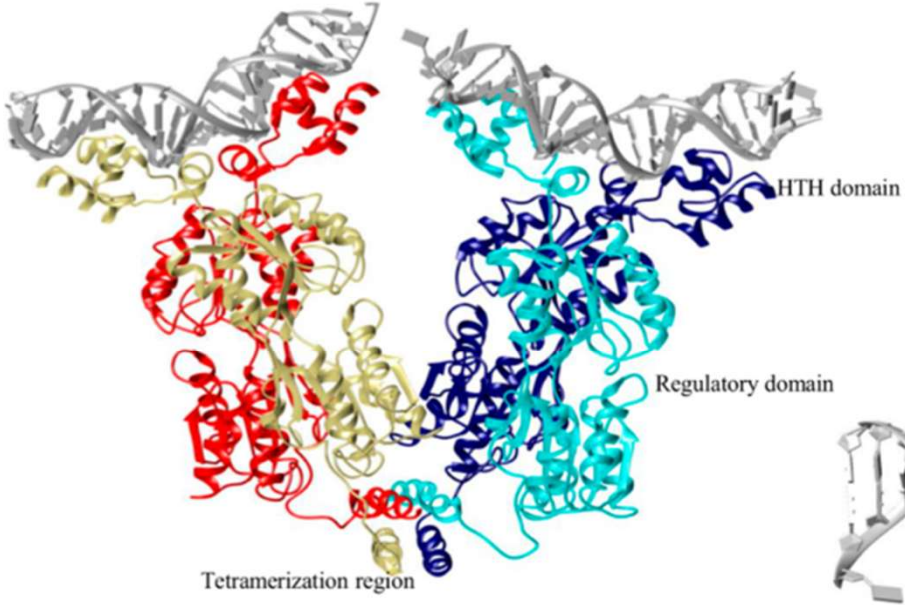
HTH domain (5d5v)



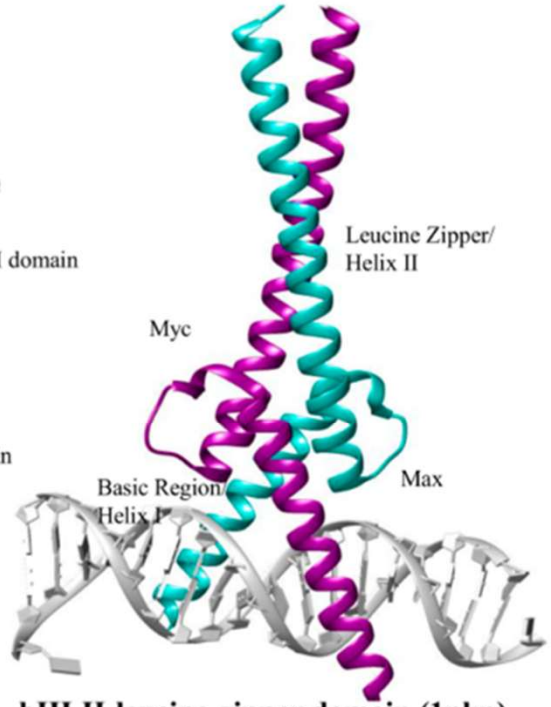
HTH-HMG domain (1gt0)



Zn finger domain (4m9e)



Lac repressor-DNA complex (1lbg)



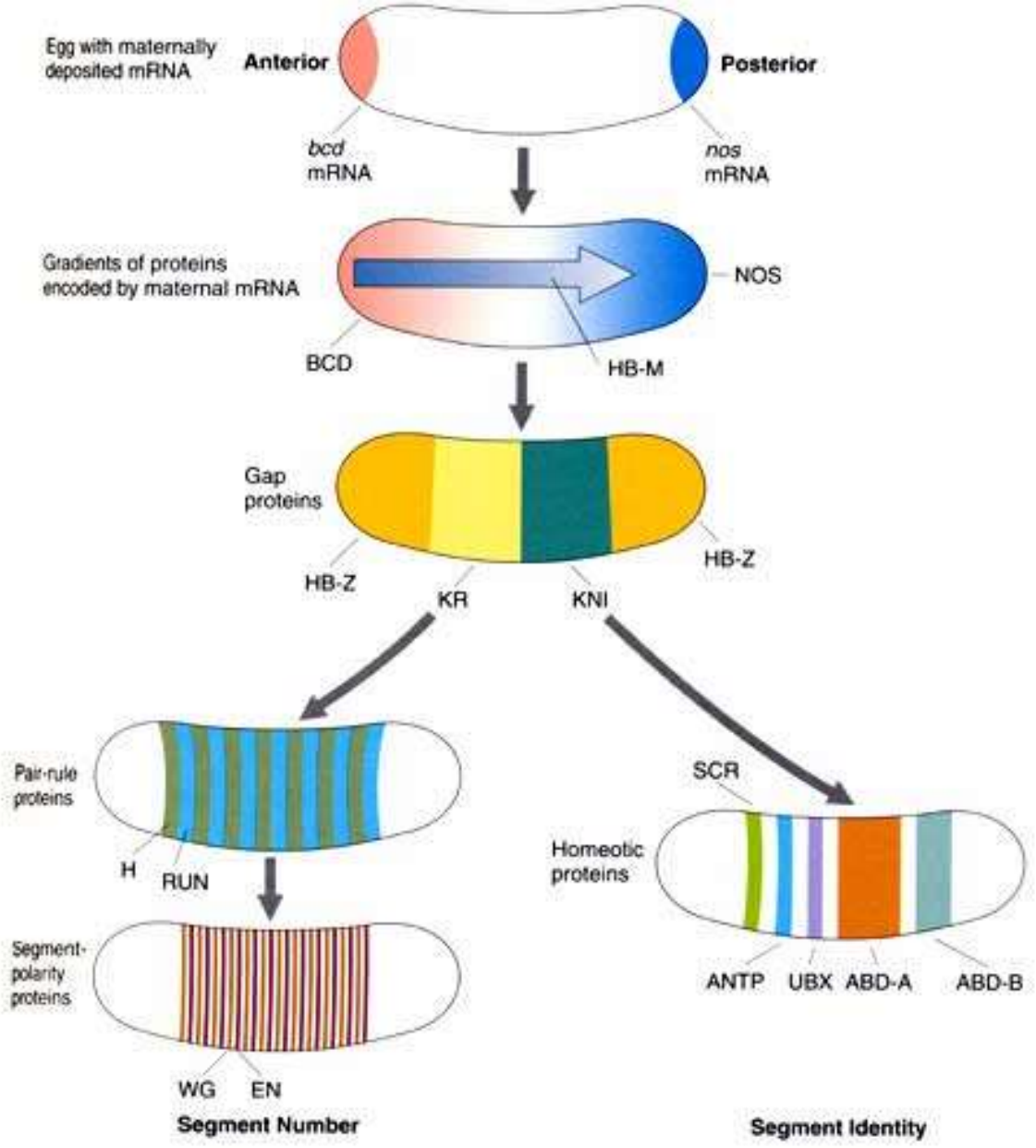
bHLH-leucine-zipper domain (1nkp)

1445 putative sequence-specific TFs in mouse genome divided in families by DNA-binding domain types

Domain	No. of genes
Homeobox	227
bHLH	116
HMG	58
bZIP	57
Nuclear Rec	50
Forkhead	40
ETS	28
ZF C_2H_2	490
ZF PHD	60
ZF C_2CH	39
ZF btb/poz	28
Other	252
TF Total	1445
Cofactors*	133
Non-TFs*	336
Total genes	1914

Gray, P.A. et. Al…, Science (2004), 306, 2255 **Mouse brain organization revealed through direct genome-scale TF expression analysis.**

A hierarchy of sequence-specific DNA-binding proteins control *Drosophila* embryonic segmentation.

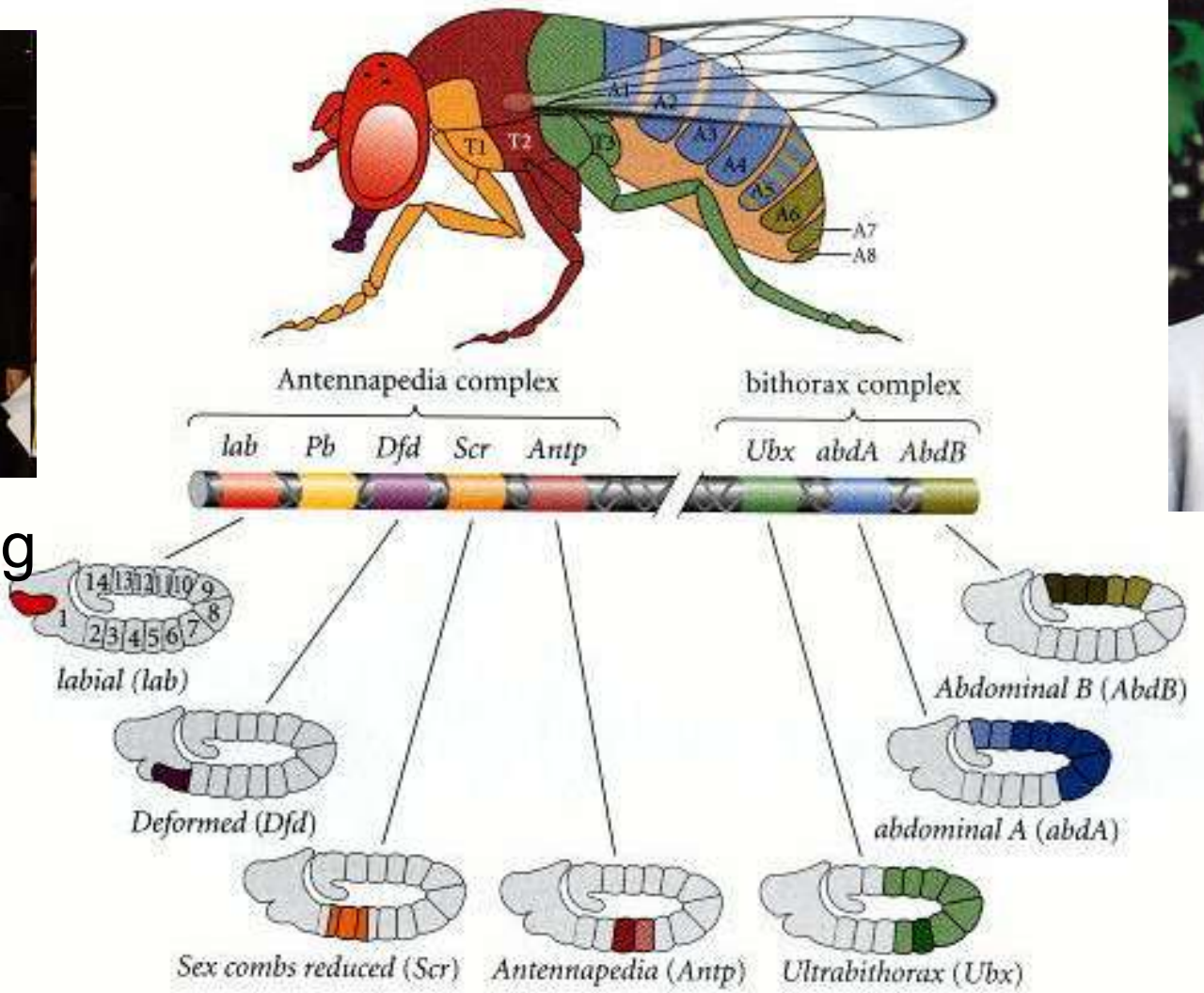


Christiane Nüsslein-Volhard (1942 -)

Eric Wieschaus (1947 -)

After the segment specification stage the segment identities are set by homeotic genes encoding homeodomain transcription factors.

Walter Gehring



Ed Lewis

The mechanism of transcription activation by gene-specific regulators.

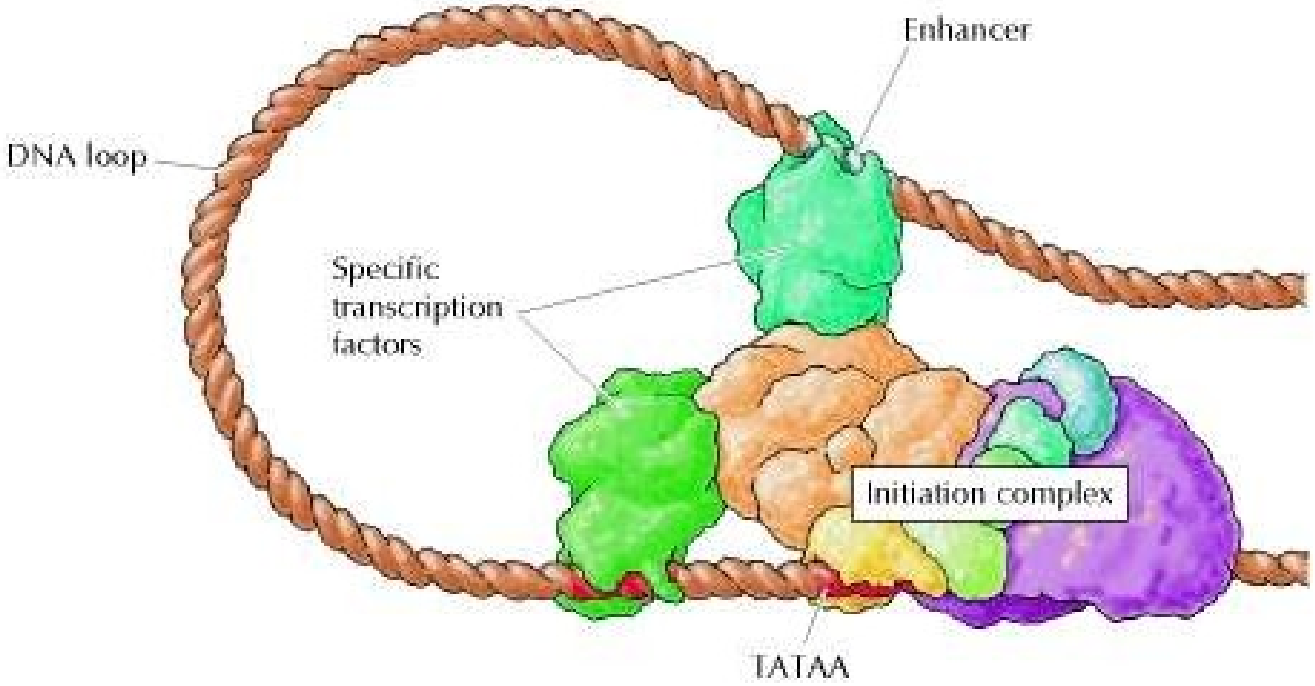
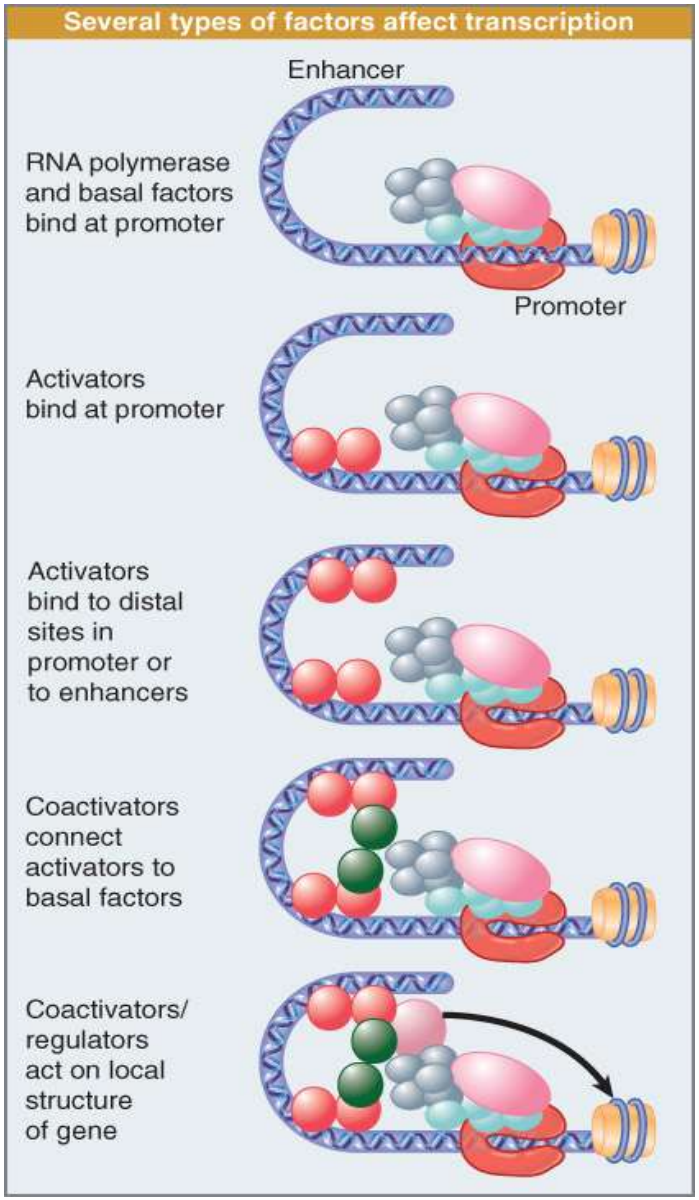


Figure 6.22 DNA looping

Transcription factors bound at distant enhancers are able to interact with general transcription factors at the promoter because the intervening DNA can form loops. There is therefore no fundamental difference between the action of transcription factors bound to DNA just upstream of the promoter and to distant enhancers.

25.2 Gene specific transcription factors (Activators/Coactivators) are distinct from General Transcription Factors (GTFs) for initiation.



- **The basal apparatus determines the startpoint for transcription.**
- **Activators determine the frequency of transcription.**
- **Activators work by making protein–protein contacts with the basal factors.**
- **Activators may work via coactivators.**
- **Some components of the transcriptional apparatus work by changing chromatin structure. Coactivator p300/CBP (CREB-binding protein) is a histone acetylase targeting H4.**

Figure 25.2

- **Several factors in the basal apparatus are targets with which activators or coactivators interact.**
- Many activators contact different TAFs. Acidic activators contact TFIIB.

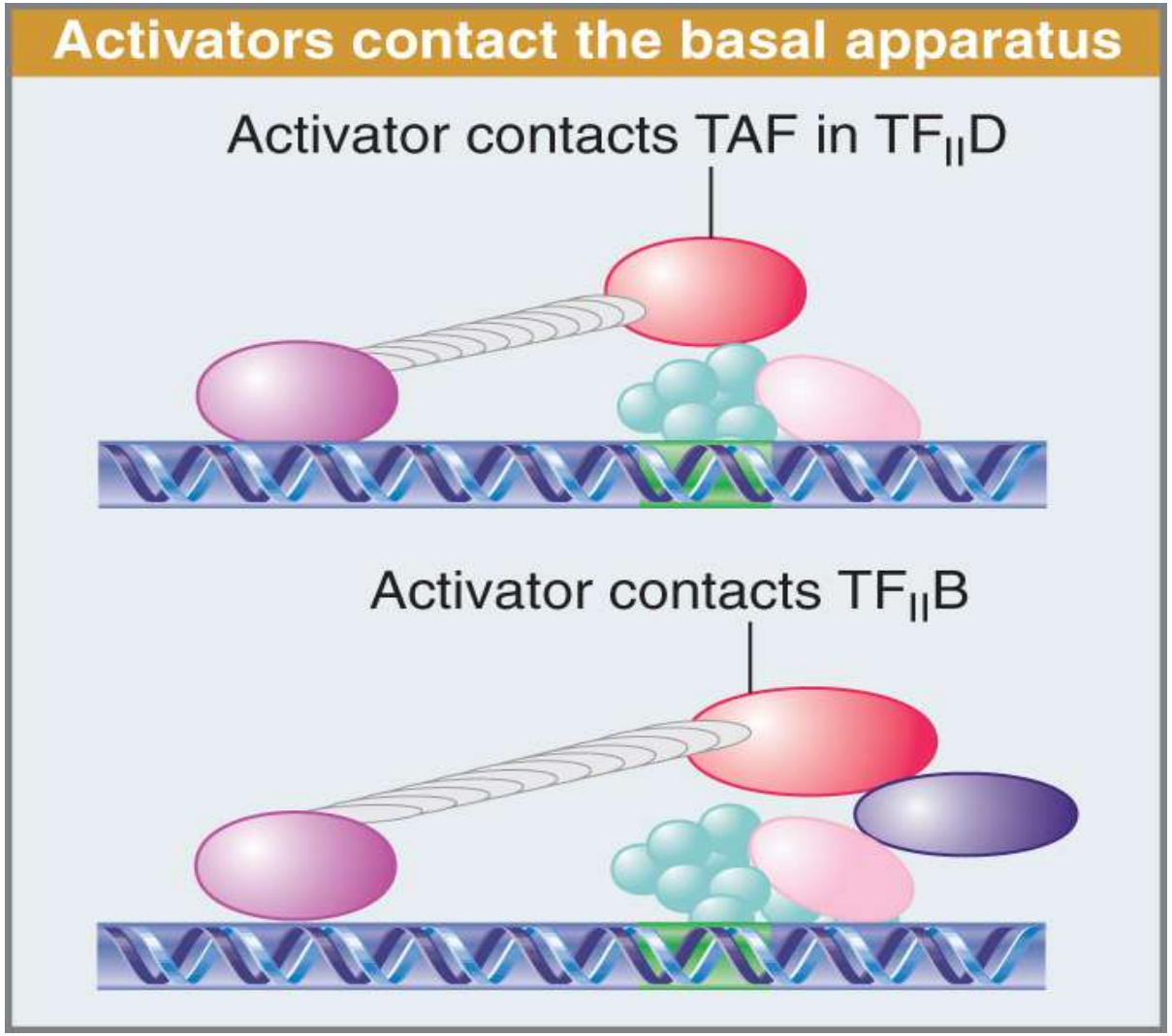


Figure 25.8

- **An activator that does not have an activating domain may work by binding a coactivator that has an activating domain**.
- Strong viral activators like Herpesvirus VP16 and Adenovirus E1A are recruited by binding other proteins.

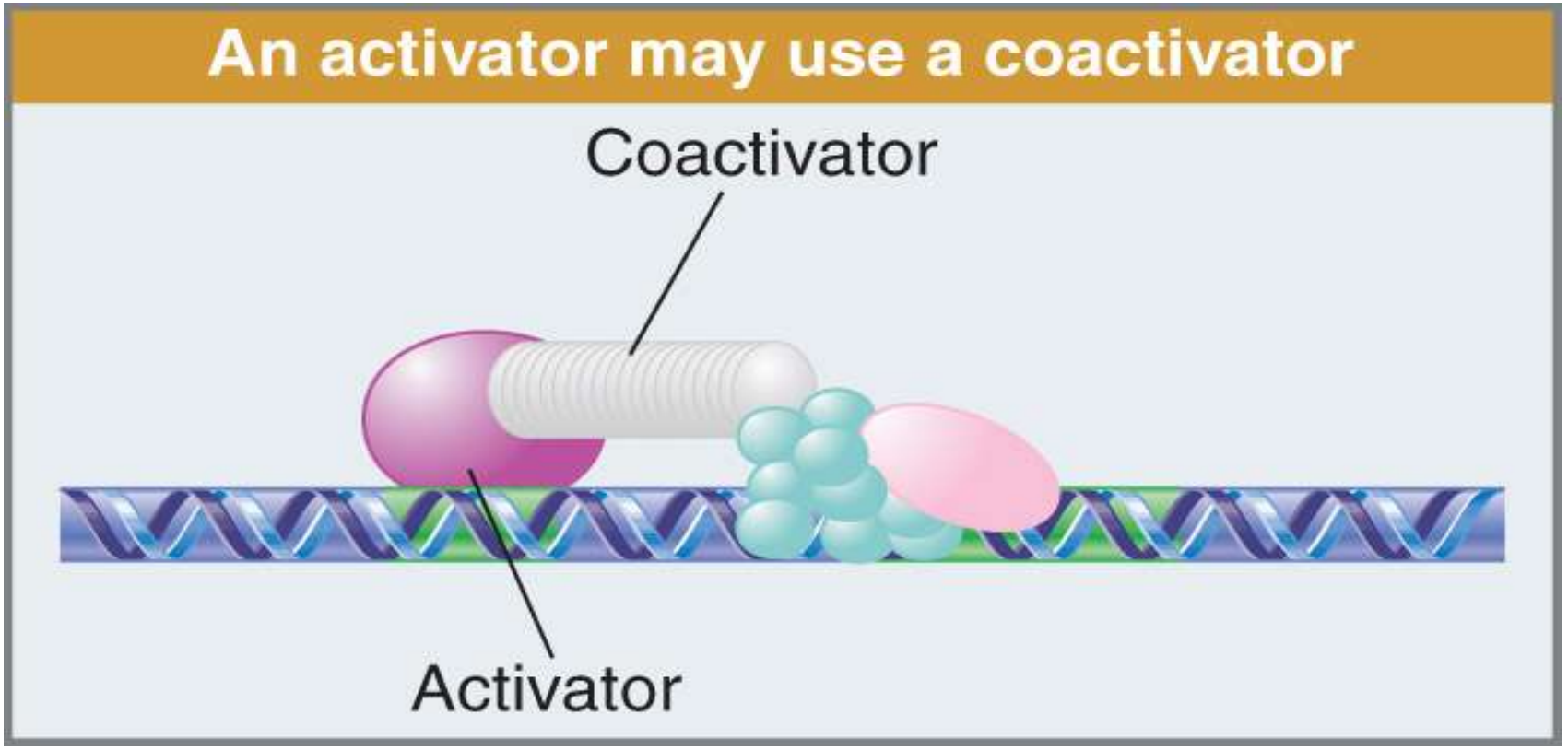
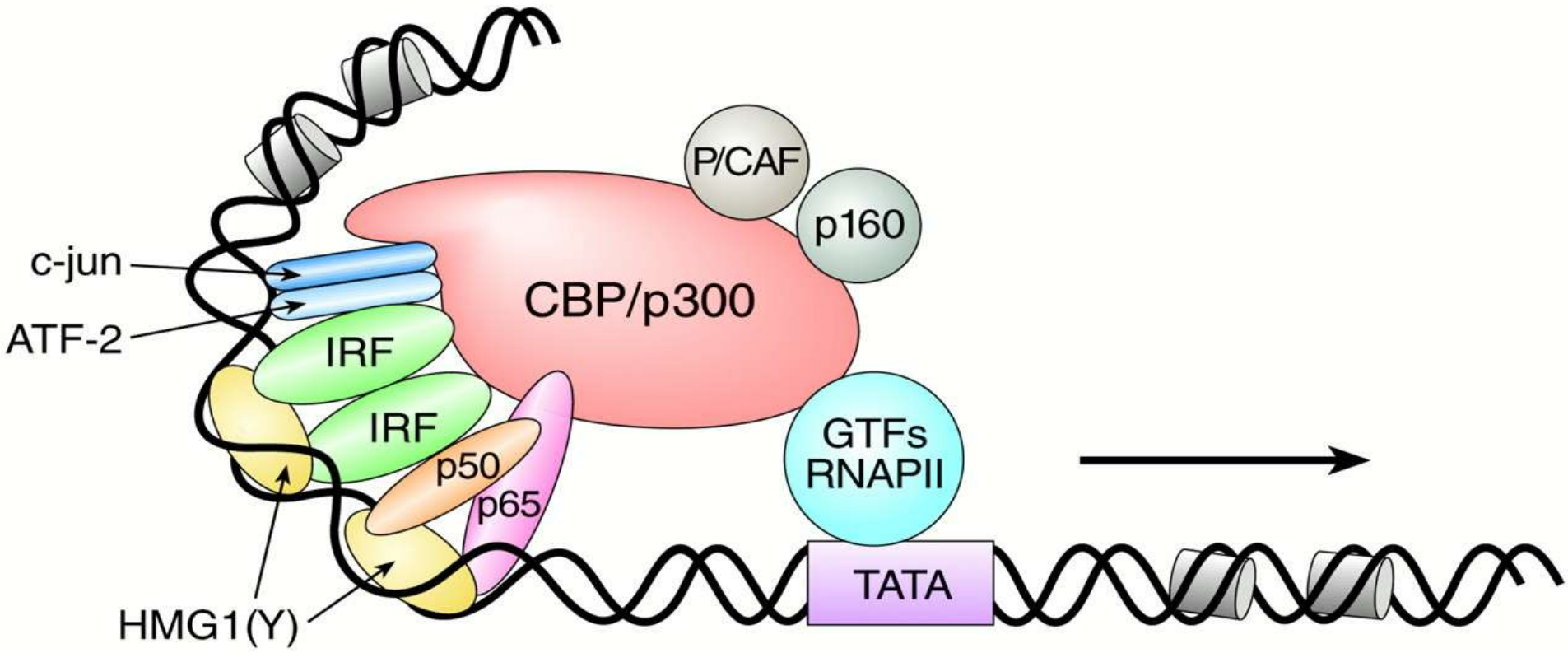


Figure 25.7

The IFN-β enhanceosome complex.



Vo N , Goodman R H J. Biol. Chem. 2001;276:13505-13508

Acidic activation domains in GAL4, NF-κB, p53 and other specific TFs.

Many transcription activators recruit the coactivator p300 through a 9 amino acid transactivation domain (**9aaTADs**).

Trans-activating domain [edit]

Trans-activating domains (TADs) are named after their amino acid composition. These amino acids are either essential for the activity or simply the most abundant in the TAD. Transactivation by the Gal4 transcription factor is mediated by acidic amino acids, whereas hydrophobic residues in Gcn4 play a similar role. Hence, the TADs in Gal4 and Gcn4 are referred to as acidic or hydrophobic activation domains, respectively.[40]

Nine-amino-acid transactivation domain (9aaTAD) defines a novel domain common to a large superfamily of eukaryotic transcription factors represented by Gal4, Oaf1, Leu3, Rtg3, Pho4, Gln3, Gcn4 in yeast and by p53, NFAT, NF-κB and VP16 in mammals.[41] Prediction for 9aa TADs (for both acidic and hydrophilic transactivation domains) is available online from ExPASy [42] and EMBnet Spain [43]

9aaTAD transcription factors p53, VP16, MLL, E2A, HSF1, NF-IL6, NFAT1 and NF-κB interact directly with the general coactivators TAF9 and CBP/p300.[44] p53 9aaTADs interact with TAF9, GCN5 and with multiple domains of CBP/p300 (KIX, TAZ1,TAZ2 and IBiD).[45]

KIX domain of general coactivators Med15(Gal11) interacts with 9aaTAD transcription factors Gal4, Pdr1, Oaf1, Gcn4, VP16, Pho4, Msn2, Ino2 and P201.[46] Interactions of Gal4, Pdr1 and Gcn4 with Taf9 were reported.[47] 9aaTAD is a common transactivation domain recruits multiple general coactivators TAF9, MED15, CBP/p300 and GCN5.[48]

	Annotated 9aaTAD	Peptide - KIX interaction (NMR data)
p53 TAD1	E TFSD LWKL	LSPEETFSDLWKLPE
p53 TAD2	D DIEQ WFTE	QAMDDLMLSPDDIEQWFTEDPGPD
MLL	S DIMD FVLK	DCGNILPSDIMDFVLKNTP
E2A	D LLDF SMMF	PVGTDKELSDLLDFSMMFPLPVT
Rtg3	E TLDF SLVT	*E2A homolog*
CREB	R KILN DLSS	RREILSRRPSYRKILNDLSSDAP
CREBαB6	E AILA ELKK	*CREB-mutant binding to KIX*
Gli3	D DVVQ YLNS	*TAD homology to CREB/KIX*
Gal4	D DVYN YLFD	*Pdr1 and Oaf1 homolog*
Oaf1	D LFDY DFLV	DLFDYDFLV
Pip2	D FFDY DLLF	*Oaf1 homolog*
Pdr1	E DLYS ILWS	EDLYSILWSDWY
Pdr3	T DLYH TLWN	*Pdr1 homolog*

Martin Piscacek,
MUNI

Mediator.

- *In vitro* transcription extracts were developed in which Pol II and the General Transcription Factors responded with increased transcription when GAL4 or another activator was added.
- This allowed further purifications to isolate additional components needed to mediate the activation. Depending on what activator was used this purified coactivators or mediator complex.
- Some argue that the complex of Pol II, GTFs and Mediator can exist as a 'holoenzyme' that recruits to promoters as in *E. coli*.

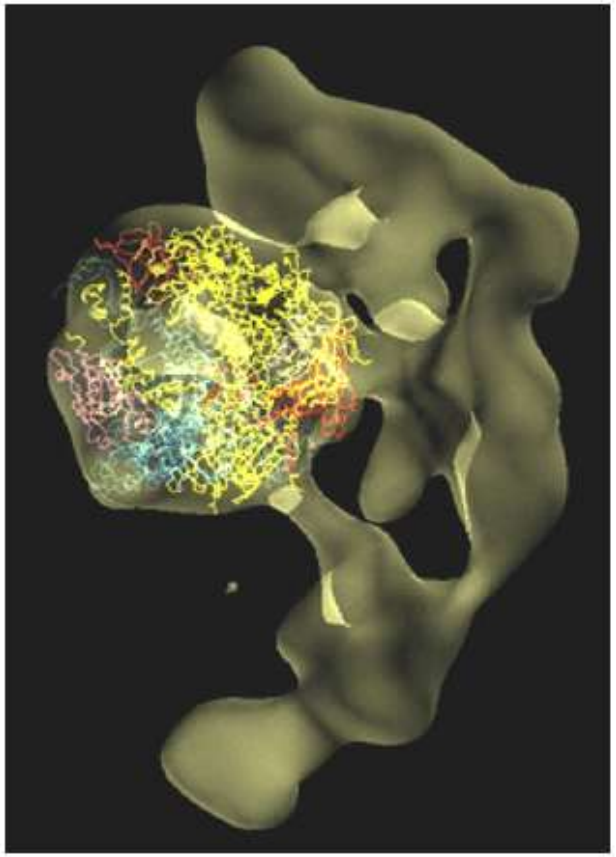
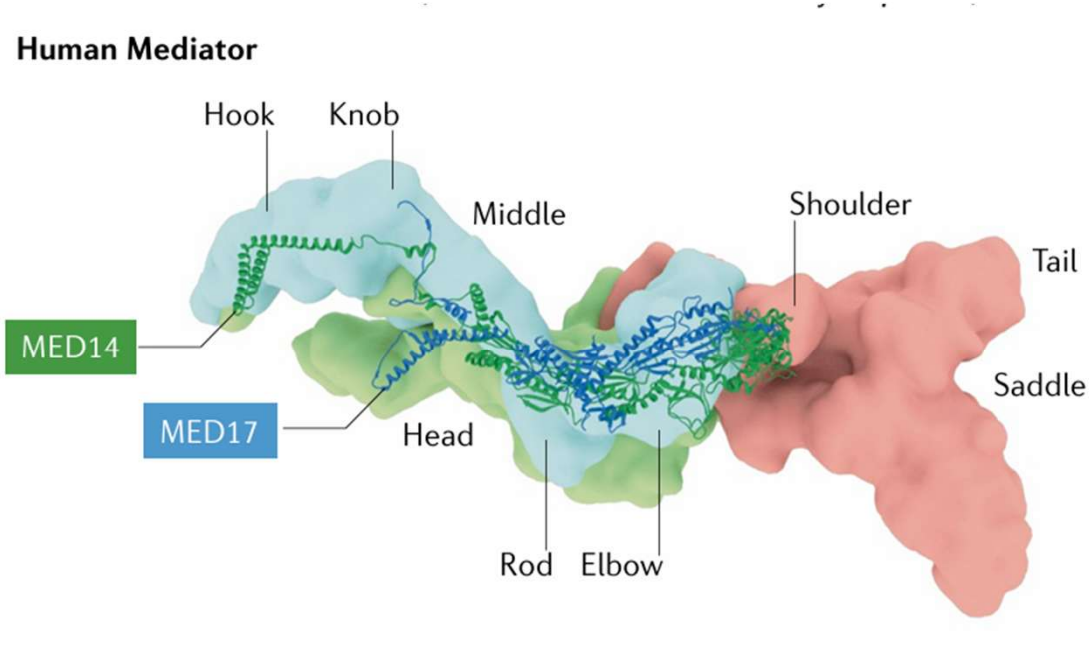


Fig. 18.

Cryo-EM structure of an RNA polymerase II-Mediator complex. The pol II structure was docked in the central density, and is shown in a similar direction of view and color scheme as Fig. 5.



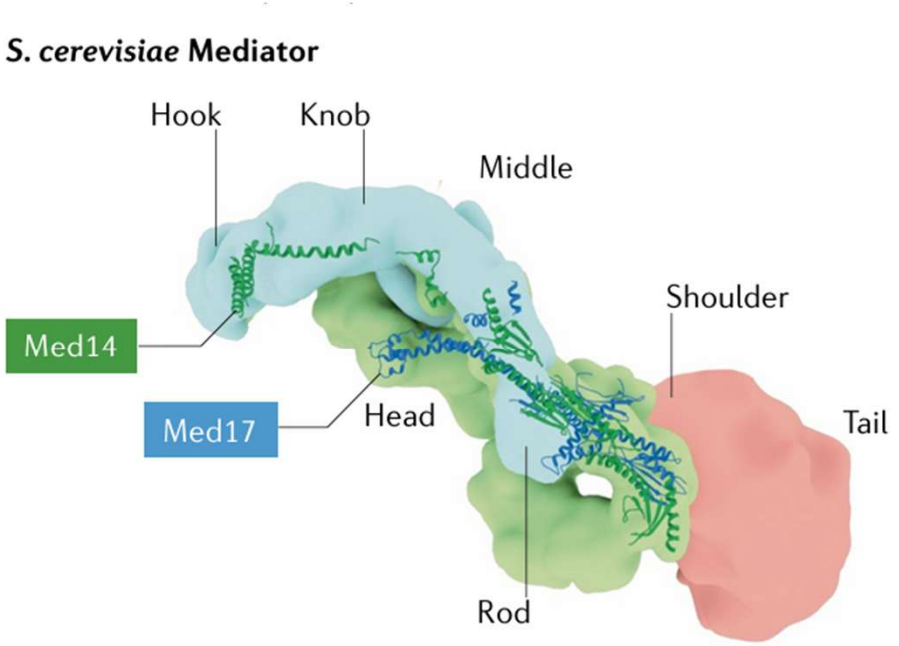


Table 2 | **A representative set of mammalian Mediator subunits and transcription factors that bind them**

Subunit	Transcription factors	Refs
MED1	TRα, TRβ, RARα, RXRα, PPARγ, VDR, PPARα, ER, AR, GR, HNF4, PGC1α, POU2AF1	95,220–226
MED14	HNF4, PPARγ	224,227
MED15	SMAD2–SMAD4, SMAD3–SMAD4, SREBP1A	90,228
MED17	VP16, p53	229
MED19	REST	230
MED23	RUNX2, E1A, ELK1	83,231,232
MED24	TR	221
MED25	ETS factors, ATF6α	233,234
CDK8[a]	MYC	235
MED12[a]	β-Catenin, REST, GLI3	236–238

The published evidence of transcription factor–Mediator interactions is too extensive to catalogue here; this is a curated list. AR, androgen receptor; ATF6α, activating transcription factor 6α; CDK8, cyclin-dependent kinase 8; E1A, adenovirus early region 1A; ER, oestrogen receptor; ETS, E26 transformation-specific; GLI3, GLI family zinc-finger 3; GR, glucocorticoid receptor; HNF4, hepatocyte nuclear factor 4; PGC1α, proliferator-activated receptor-γ coactivator 1α; POU2AF1, POU class 2 homeobox-associating factor 1; PPAR, peroxisome proliferator-activated receptor; RARα, retinoic acid receptor-α; REST, RE1 silencing transcription factor; RUNX2, RUNX family transcription factor 2; RXRα, retinoid X receptor-α; SREBP1A, sterol regulatory element-binding protein 1A; TR, thyroid hormone receptor; VDR, vitamin D receptor. [a]Subunit of the Mediator kinase module.

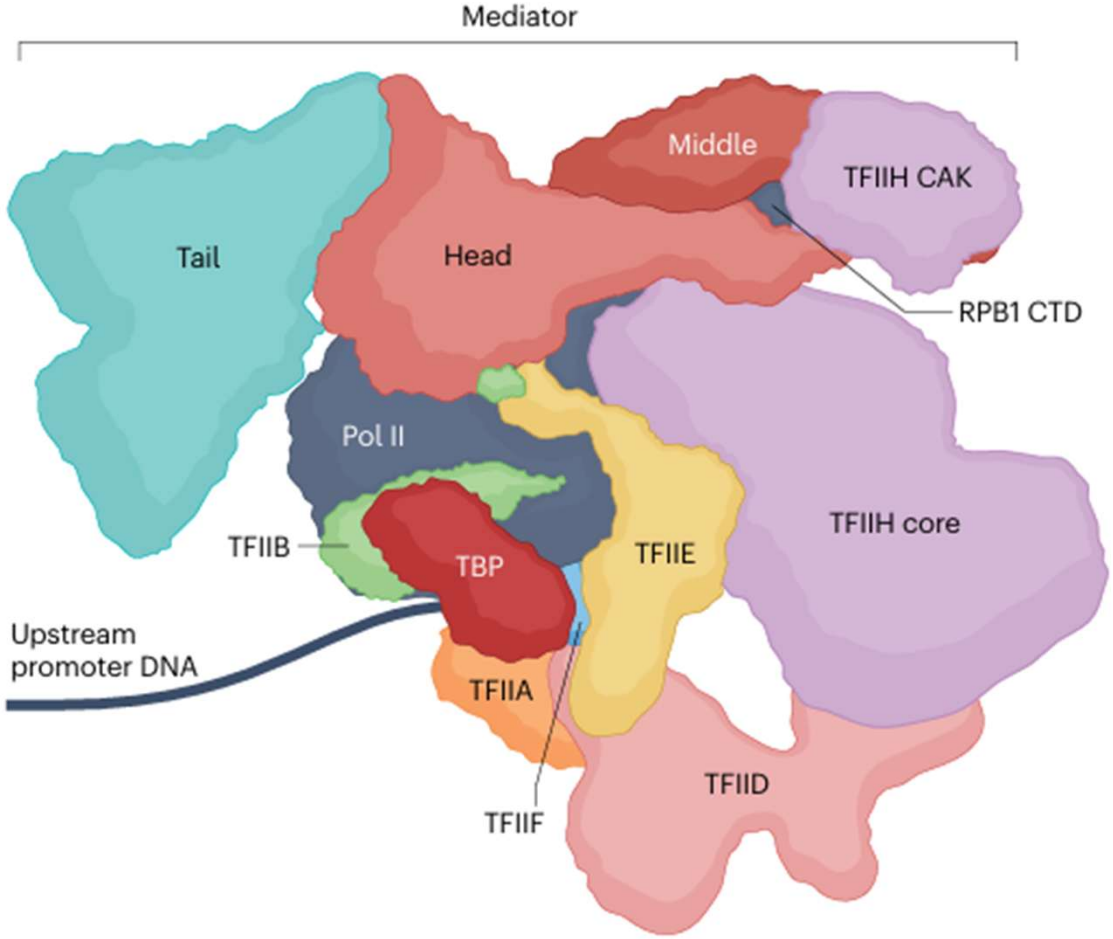


Fig. 7 | A TFIID-containing and Mediator-containing PIC. A model for a TFIID-containing and Mediator-containing pre-initiation complex (PIC) adapted from ref. 43. Note the additional stabilizing interactions relative to a TATA binding protein (TBP)-nucleated minimal PIC that does not contain Mediator (Fig. 1). Highlighted here is the stabilization of the RPB1 carboxy-terminal domain (CTD), but numerous other subtle, and perhaps dynamic, interactions between the coactivators, RNA polymerase II (Pol II) and general transcription factors (GTFs) also take place, as discussed and referenced in the main text. Additional stabilizing interactions of the PIC might come from promoter proximally bound activators and the downstream +1 nucleosome via TFIID (Fig. 5) and Mediator (not shown; see text). Other general cofactors that may be fulfilling architectural and other roles may also contribute but are not shown (Box 2). CAK, CDK-activating kinase.

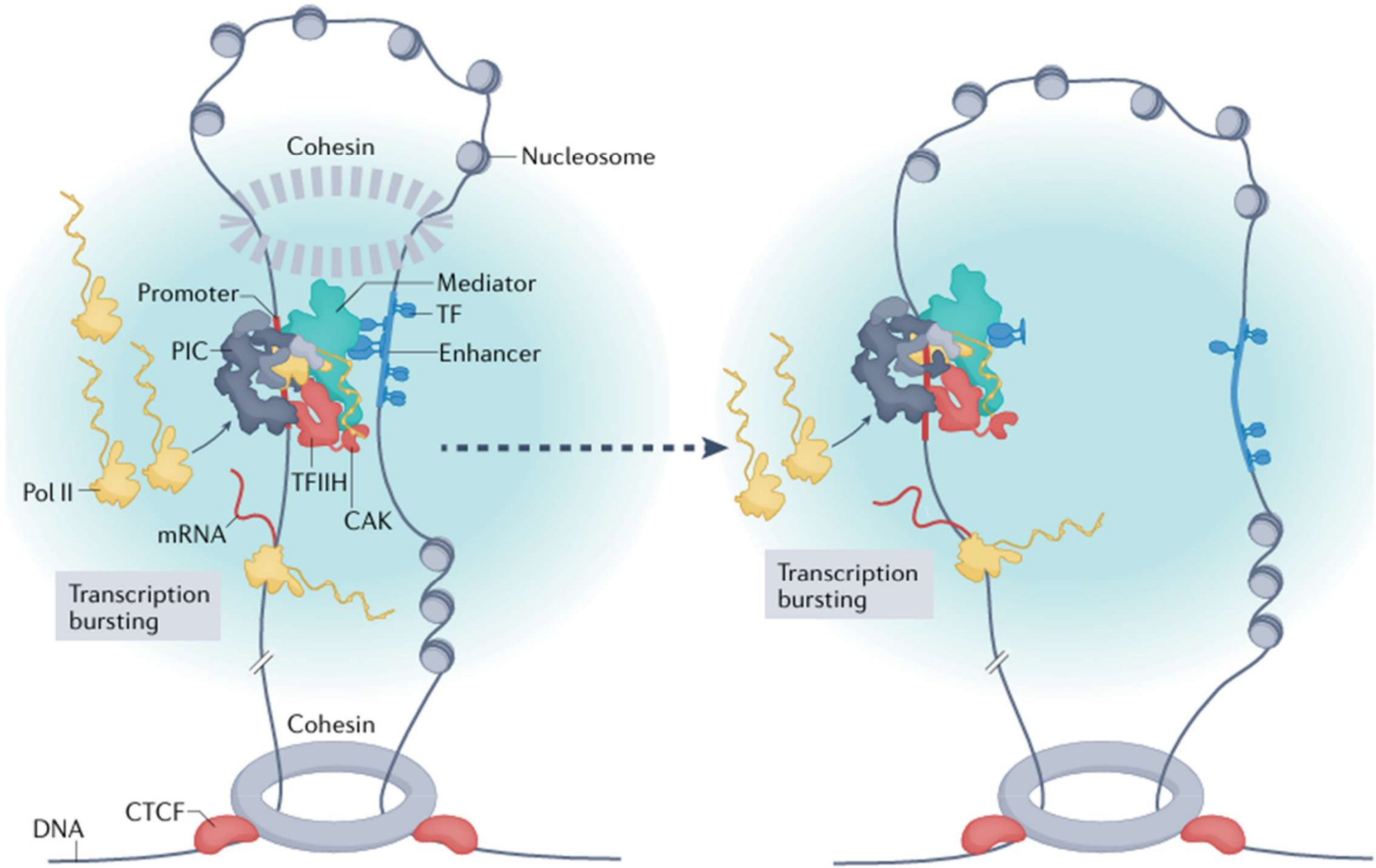


Fig. 3 | **A working model for Mediator function.** An enhancer–promoter interaction (loop) is shown on the left, within a larger topologically associating domain formed by CTCF and cohesin. Mediator is bound to one or more transcription factors (TFs) that occupy the enhancer, and the preinitiation complex (PIC) at the promoter is fully assembled and active.

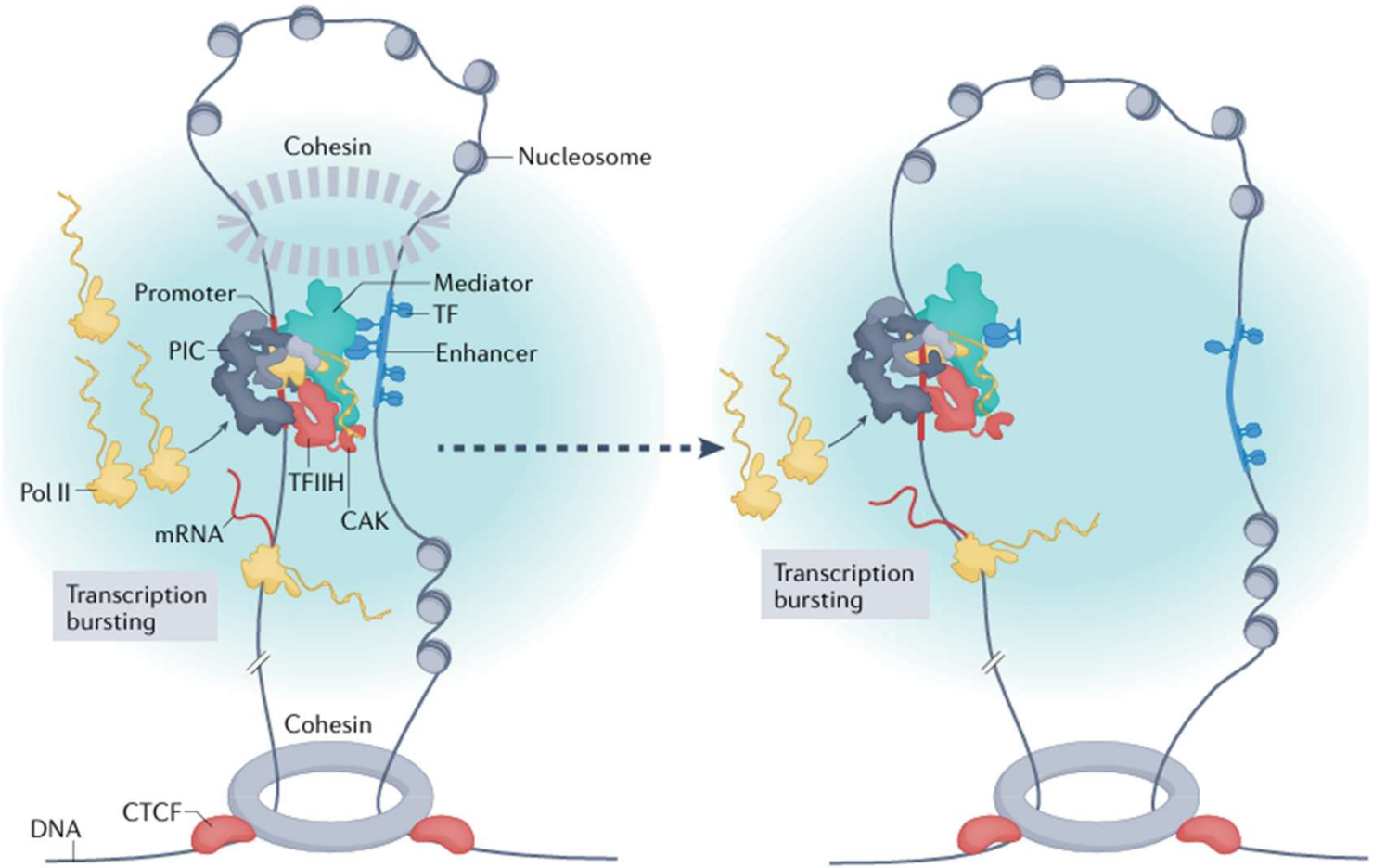


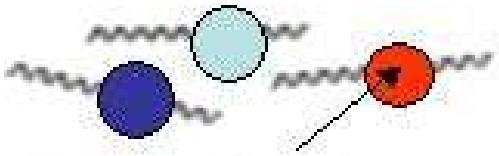
Fig. 3 | **A working model for Mediator function.** An enhancer–promoter interaction (loop) is shown on the left, within a larger topologically associating domain formed by CTCF and cohesin. Mediator is bound to one or more transcription factors (TFs) that occupy the enhancer, and the preinitiation complex (PIC) at the promoter is fully assembled and active. Such local architecture of enhancer–promoter chromatin looping could be further stabilized by Mediator-associated cohesin[167], but this association would be transient (dashed circle) relative to topologically associating domain boundaries (solid circle). Following a brief, direct enhancer–promoter interaction, the enhancer detaches from the promoter (for example, through dissociation of TFs from enhancer DNA); however, if one or more TFs remain bound to Mediator, the complex could remain in an active conformational state. This state could allow continued transcription reinitiation (bursting) from the PIC scaffold complex, provided RNA polymerase II (Pol II) and other PIC factors continue to associate for reinitiation (right). Ultimately, reinitiation may stop (not shown), because of TF–Mediator dissociation, binding of the kinase module to Mediator (which would block Mediator–Pol II interaction) or PIC disassembly. The light blue shading represents a hub or condensate that establishes a high local concentration of PIC components that promotes transcription initiation and bursting. TFIIH, transcription factor IIH.

Transcription activation in chromatin.

- Getting transcription initiation and later regulated transcription on chromatin templates required extra factors that operate on nucleosomes.
- Constructs with enhancers reintroduced in mammalian chromosomes were often silenced. This turned out later to be due to lack of regulatory elements like the globin gene Locus Control Region (LCR).
- Only certain sequence-specific DNA binding proteins can still bind their sites if histones are added first (pioneer transcription factors).

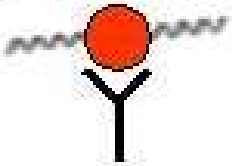
Chromatin immunoprecipitation for genome-wide location analysis of chromosomal proteins in living cells

1. Formaldehyde cross link DNA to transcription factors.
2. Lyse cells and fragment chromatin by sonication

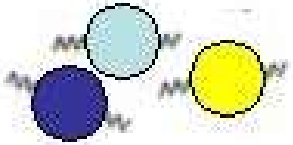


Factor X bound to target gene

3. Immunoprecipitation



Fragments bound to anti-X antibodies
Purify on beads



Unbound fragments are washed away

4. Reverse cross-links, extract DNA

5. Analyse which DNA fragments are enriched in the immunoprecipitated fraction.

6. Best done now by Next Generation sequencing (ChIP-Seq)

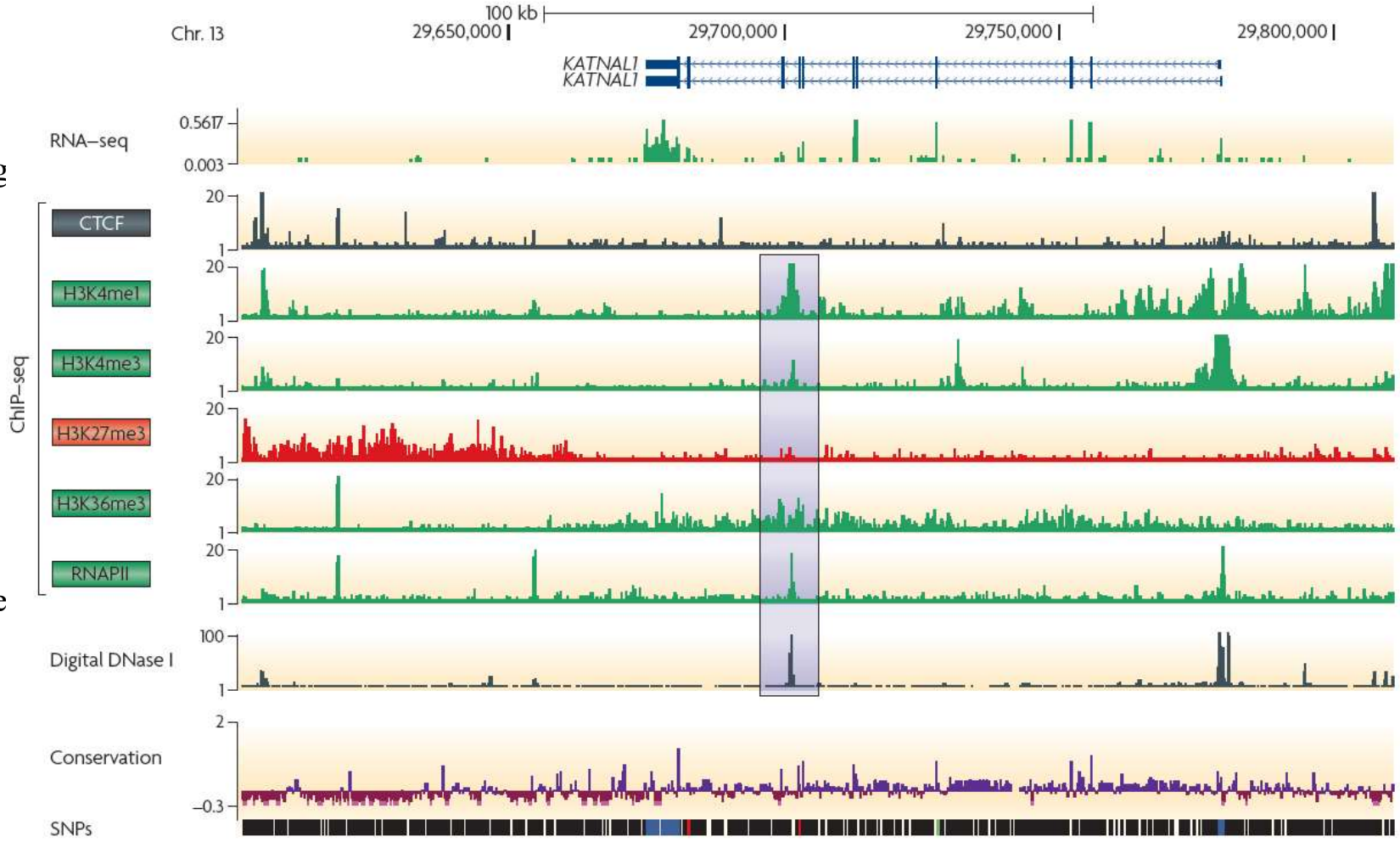
7. Compare reads of immunoprecipitated sequences back to the genome sequence to locate w the protein was bound.

Mapping Polycomb-repressed domains in the bithorax complex using in vivo formaldehyde cross-linked **chromatin**.
Orlando V, **Paro** R.
Cell. 1993 Dec 17;75(6):1187-98.

Histone modifications associated with ACTIVE promoters, polII and DNAseI sensitivity. (H3K4 Me1,Me3)

Histone modifications associated with a silent region (H3K27 Me3).

- RNA-seq = reads of cDNA sequence showing where transcripts are coming from.
- CTCF = insulator binding protein, blocks effects of distant enhancers, thought to separate regions of gene control
- SNPs = single nucleotide polymorphisms between human individuals ($3X10^6$ per person!). Some of these cause disease – how do we identify which ones?



Most yeast promoters are relatively "nucleosome free" …many mammalian promoters also….

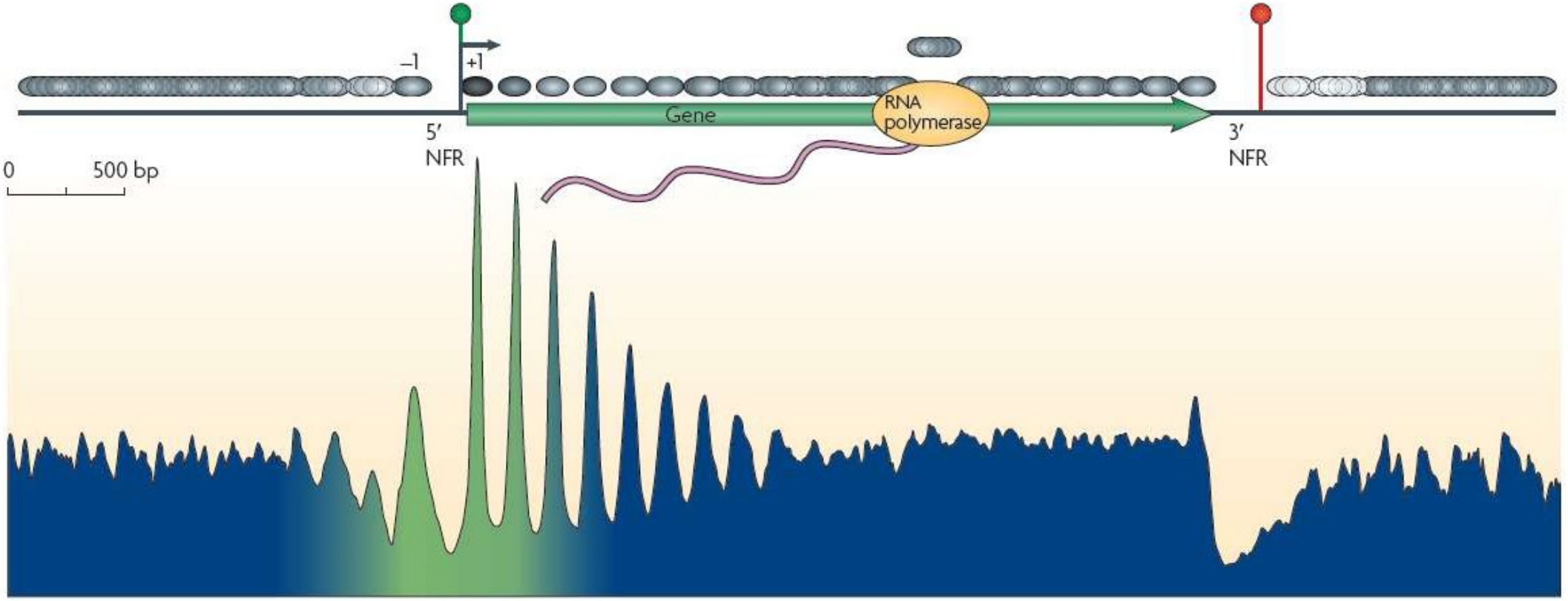


Figure 2 | **Nucleosomal landscape of yeast genes.** The consensus distribution of nucleosomes (grey ovals) around all yeast genes is shown, aligned by the beginning and end of every gene. The resulting two plots were fused in the genic region. The peaks and valleys represent similar positioning relative to the transcription start site (TSS). The arrow under the green circle near the 5′ nucleosome-free region (NFR) represents the TSS. The green –blue shading in the plot represents the transitions observed in nucleosome composition and phasing (green represents high H2A.Z levels, acetylation, H3K4 methylation and phasing, whereas blue represents low levels of these modifications). The red circle indicates transcriptional termination within the 3′ NFR. Figure is reproduced, with permission, from REF. 20 © (2008) Cold Spring Harbor Laboratory Press.

Relatively nucleosome-free regions, DNAse hypersensitive sites and specific histone modifications help identify enhancers in chromatin.

- Chromatin immunoprecipitation methods are easier to use with abundant nucleosome modifications than with scarce sequence-specific DNA-binding proteins.

- Modified histones associated with active (H3K4 Me1,Me3) and inactive (H3K27 Me3) genes have been mapped extensively. This is a proxy for mapping the regulatory proteins themselves and allows likely enhancers and promoters to be identified.

- Promoters and enhancers also tend to be nucleosome-free. **ATAC-Seq, (Assay for Transposase-Accessible Chromatin)**, mutagenises isolated chromatin *in vitro* with a hyperactive Tn5 transposase inserting sequencing tags, subsequent sequencing identifies insertions in nucleosome-free regions.

Summary

- Eukaryotic promoters
- General Transcription Factors
- Transcription initiation
- Specific gene regulators, Specific transcription factors, DNA-binding assays
- Gene activation mechanisms, cofactors and Mediator
- Transcription in chromatin

Pre-mRNA processing in eukaryotes.
- Ribozymes and the RNA world-

Eukaryotic total RNA.

Ribosomal RNAs and tRNAs are major bands, mRNA is a smear on denaturing gel stained with Ethidium Bromide.

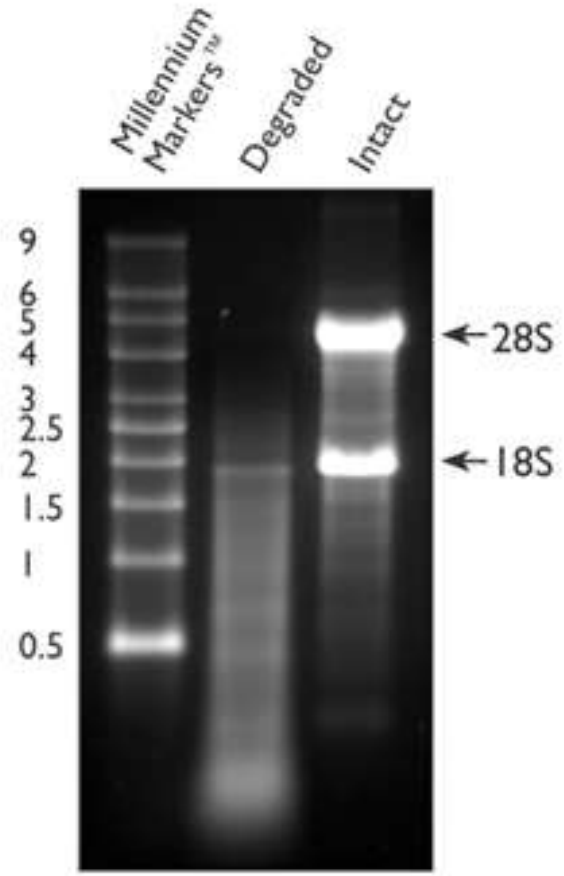
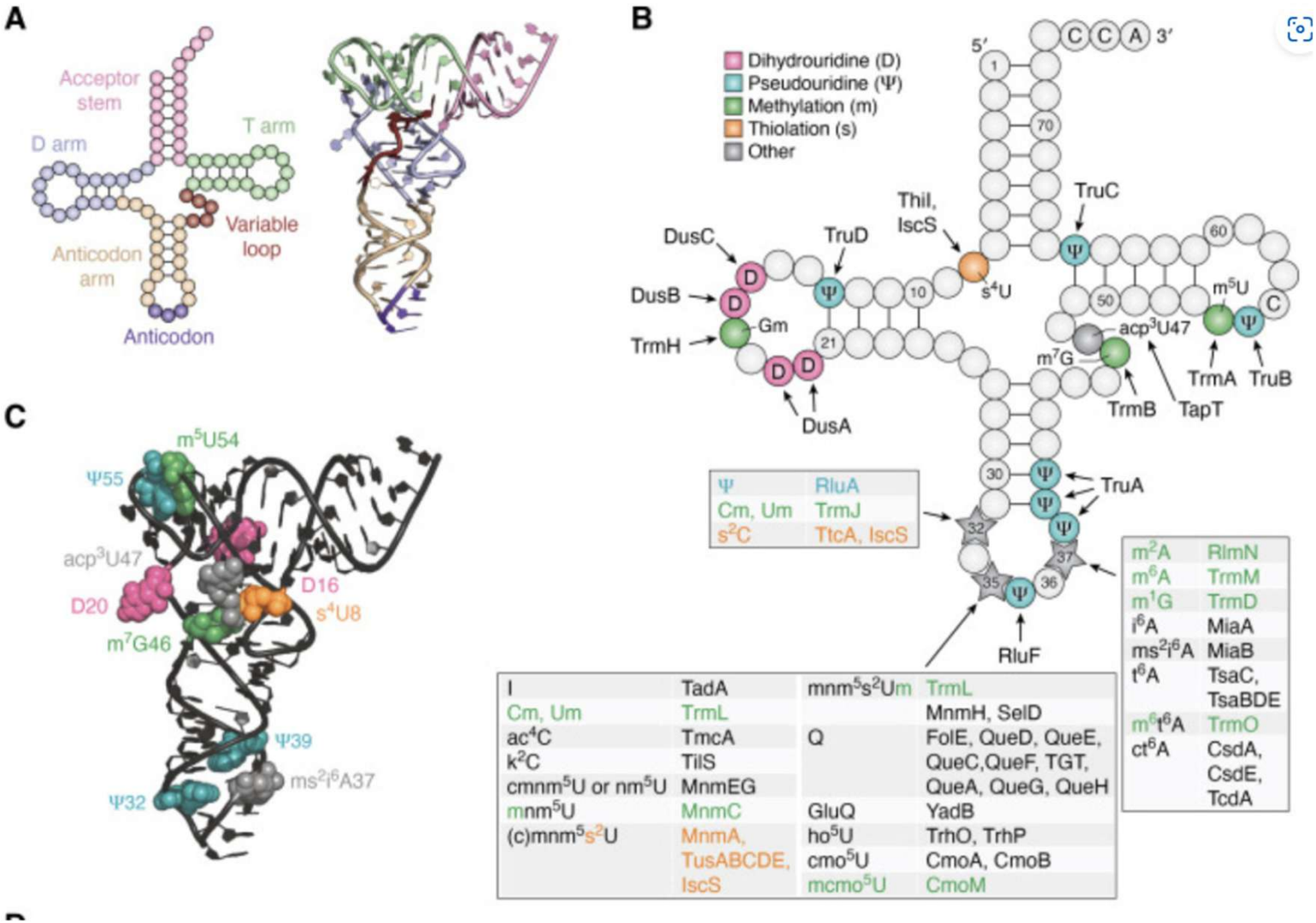


Figure 1. Intact vs. Degraded RNA. Two µg of degraded total RNA and intact total RNA were run beside Ambion's RNA Millennium Markers™ on a 1.5% denaturing agarose gel. The 18S and 28S ribosomal RNA bands are clearly visible in the intact RNA sample. The degraded RNA appears as a lower molecular weight smear.

Enzymatic base modifications help stabilize tRNAs, rRNAs and other stable RNAs (snRNAs, snoRNAs etc) to fold correctly and function correctly

Enzymatically modified bases fold the L-shaped tRNA and help make translation more accurate

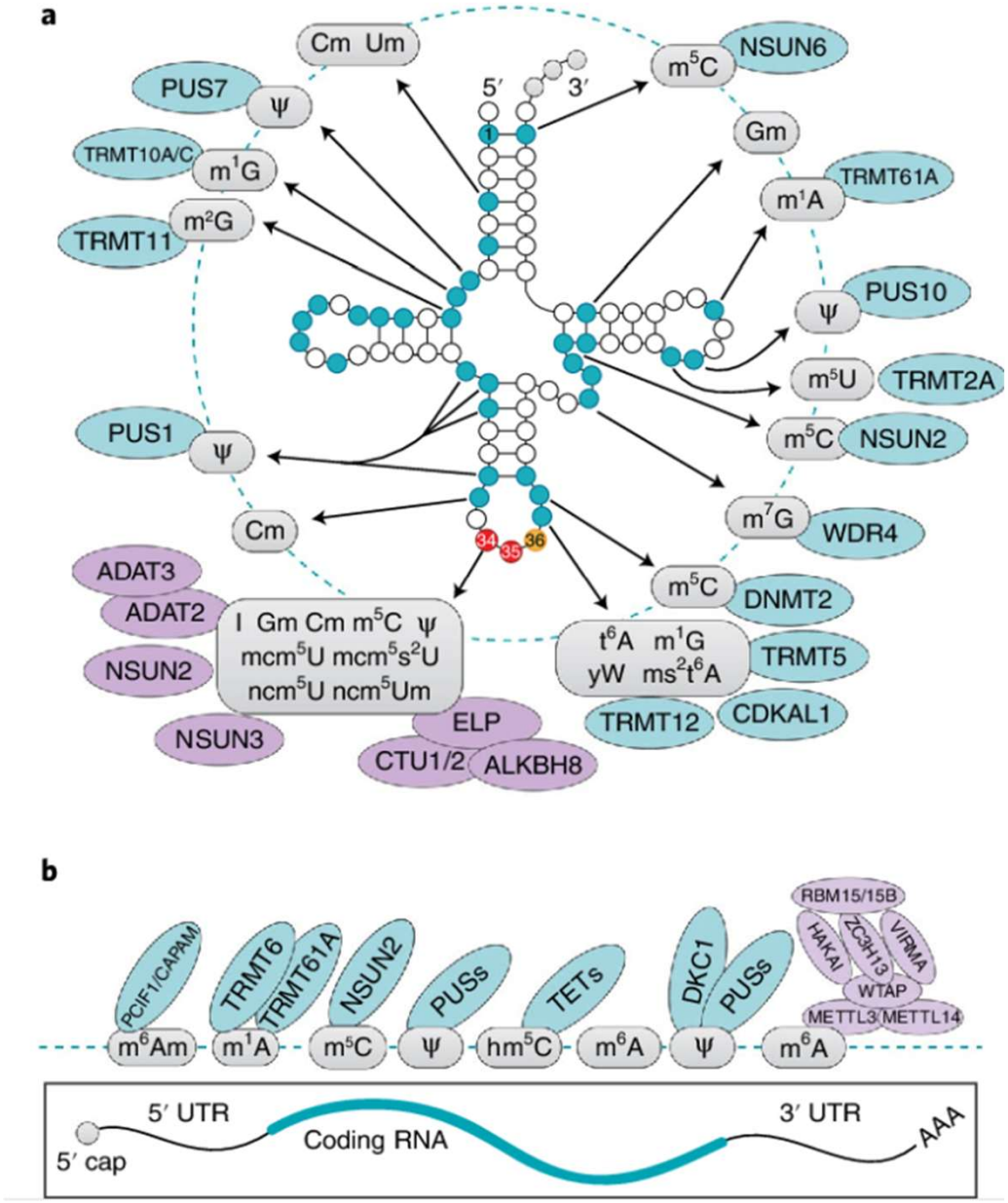


Ψ	RluA
Cm, Um	TrmJ
s^2C	TtcA, IscS

I	TadA	mnm^5s^2Um	TrmL
Cm, Um	TrmL		MnmH, SelD
ac^4C	TmcA	Q	FolE, QueD, QueE, QueC,QueF, TGT, QueA, QueG, QueH
k^2C	TilS		
$cmnm^5U$ or nm^5U	MnmEG		
mnm^5U	MnmC	GluQ	YadB
$(c)mnm^5s^2U$	MnmA, TusABCDE, IscS	ho^5U	TrhO, TrhP
		cmo^5U	CmoA, CmoB
		$mcmo^5U$	CmoM

m^2A	RlmN
m^6A	TrmM
m^1G	TrmD
i^6A	MiaA
ms^2i^6A	MiaB
t^6A	TsaC, TsaBDE
m^6t^6A	TrmO
ct^6A	CsdA, CsdE, TcdA

tRNA modified bases and the enzymes that modify them

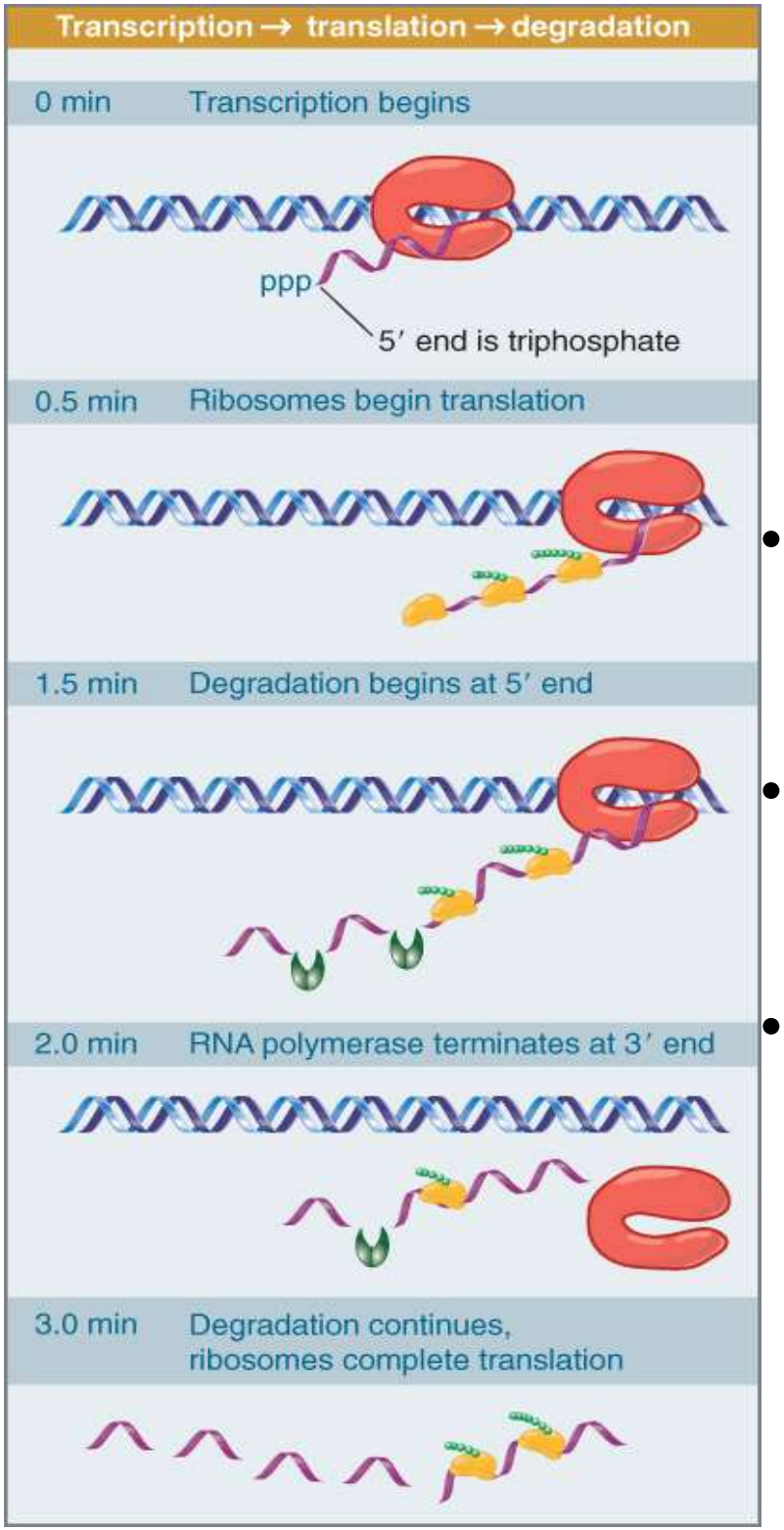


- rRNAs also have pseudouridines and ribose 2'O-methyl groups added by many different protein RNA complexes with snoRNA guides

Capping and polyA tailing of pre-mRNAs.

Lecture outline

- Eukaryotic mRNA stability, 5’ cap and polyA addition and splicing.
- Mechanism of pre-mRNA splicing and alternative splicing. RNA editing.
- Self-splicing Group II introns, RNA catalysis, Ribozymes and the The RNA World.



The life cycle of a bacterial mRNA.

- Bacterial mRNA is unstable and has a half-life of only a few minutes.
- Transcription in both bacteria and eukaryotes is not very fast at 40 nucleotides/sec, 2.4 kb/min.
- Most bacterial mRNAs can be transcribed in a few minutes whereas a human transcript of 100 kb would take 40 minutes.

Figure 7.14

• pre-mRNA processing

- A typical eukaryotic mRNA:
-
- 5'cap....._____.....AAAAAAAAAn
- nontranslated coding long 3' trailer n =100-200
- leader <300nt often >1000nt

Pre-mRNA 5’ cap addition and 3’ polyadenylation and splicing are completed before export from the nucleus.

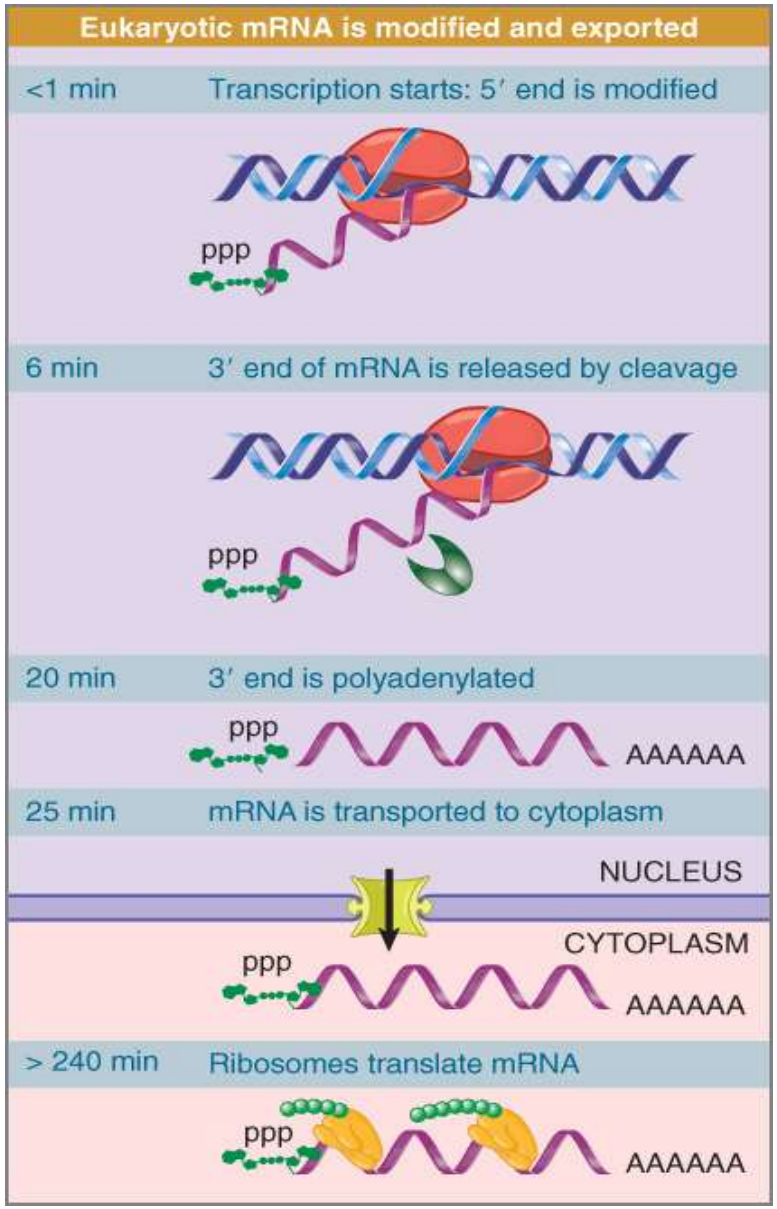


Figure 7.17

The 5′ End of Eukaryotic mRNA Is Capped to protect it against 5’ exonucleases.

- A 5′ cap is formed by adding a G to the terminal base of the transcript via a 5′–5′ link.
- The capping enzyme is **guanylyl transferase.**
  - 1-3 methyl groups are added to the base of the new terminal guanosine
  - 2 methyl-riboses near cap.
  - Base modification m6A in 5’ cap and also inside mRNAs at RRACH consensus, especially around stop codon.

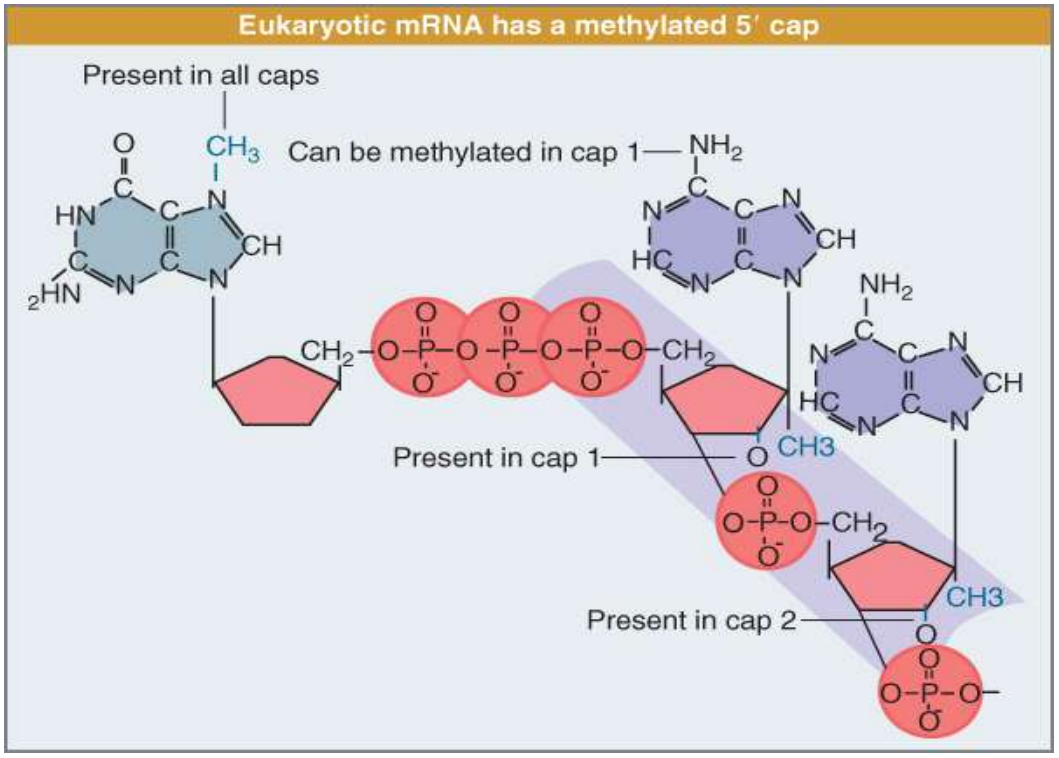
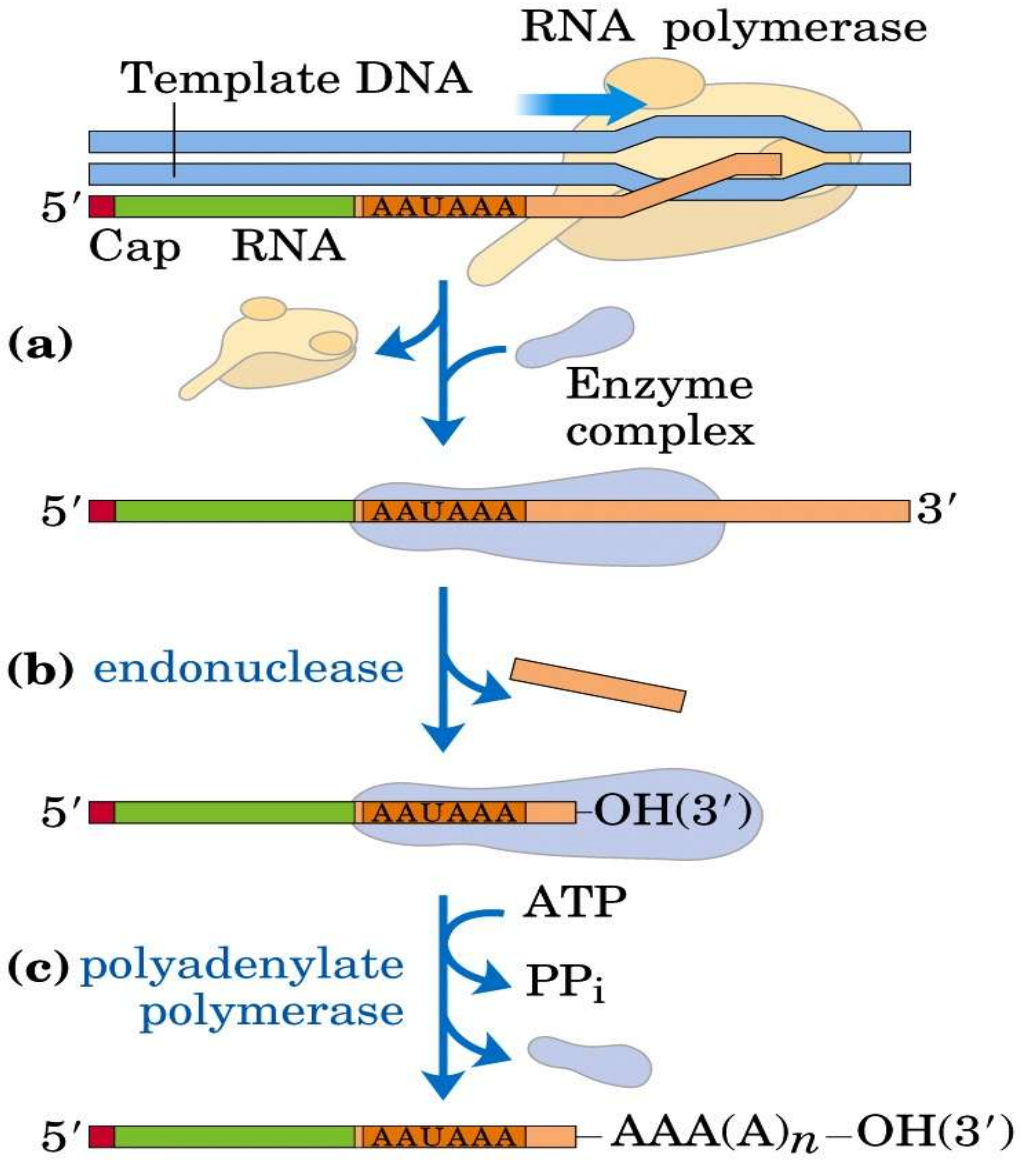


Figure 7.18

3’ end cleavage and polyadenylation.



The cleavage and polyadenylation specificity factor (CPSF) recognizes **AAUAAA (AtwoU, Athree)**.

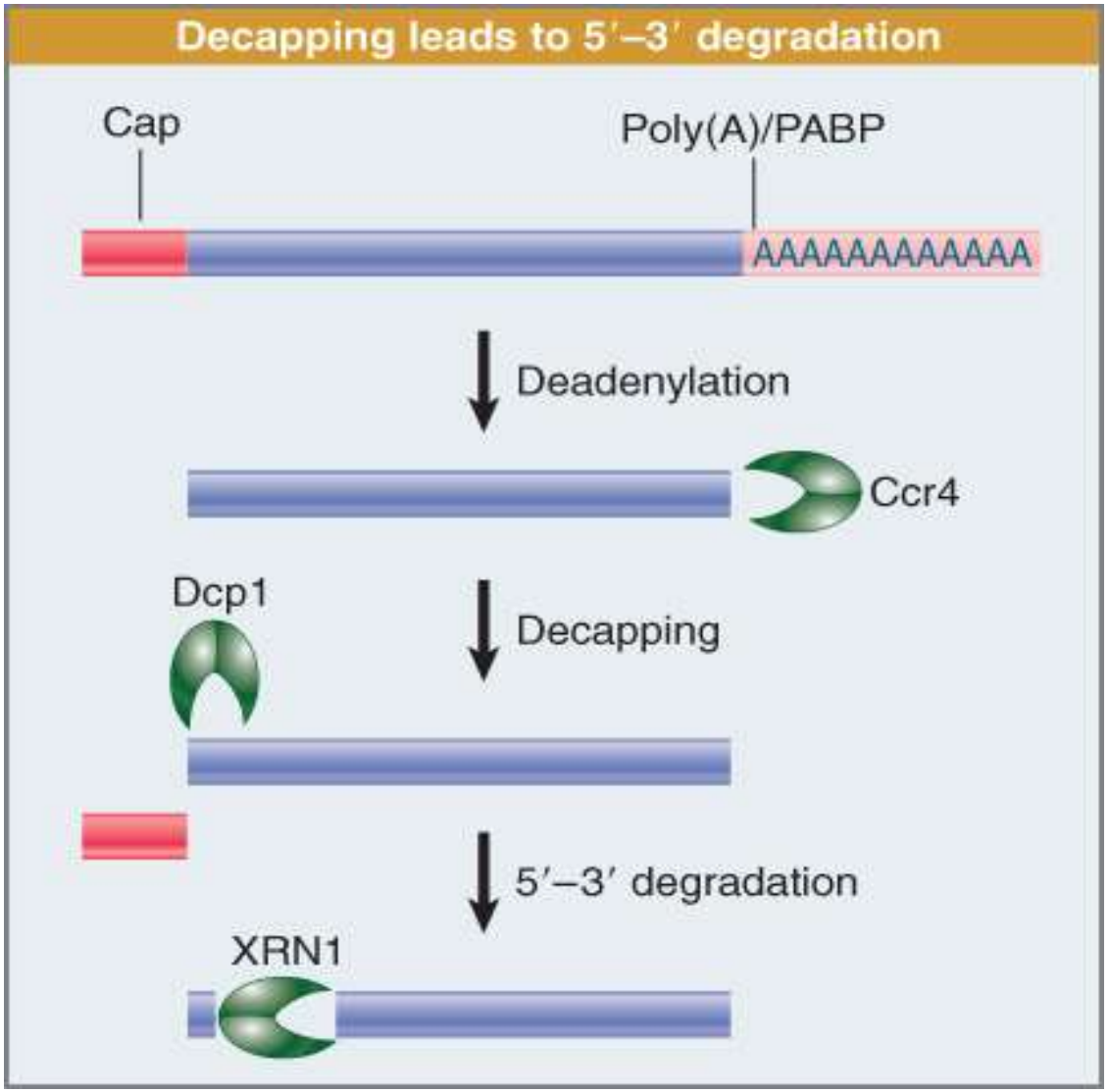
Cleavage factors (CFI and CFII), endonuclease cleaves RNA 30 bases downstream.

Poly(A) polymerase add ~200 A residues processively to the 3′ end.

The polyA tail stabilizes mRNA, facilitates translation.

- Poly A tail is >200 nucleotides long and binds many copies of Poly A-binding protein (PABP).
- PABP stabilizes the mRNA against degradation.
- PABP binds eIF4G (eukaryotic initiation factor 4G) and facilitates translation of message by the ribosome.
- 5’ and 3’ ends are thought to contact one another by forming a loop of the mRNA.

Removal of cap and polyA tails are required for regulated degradation of mRNA.



- Degradation of yeast mRNA requires removal of the 5′ cap and the 3′ poly(A).
- One yeast pathway involves exonucleolytic degradation from 5′→3′.

Figure 7.24

- Another yeast mRNA degradation pathway uses a complex of several exonucleases that work in the 3′→5′ direction The **exosome**.

- The deadenylase of animal cells may bind directly to the 5′ cap.

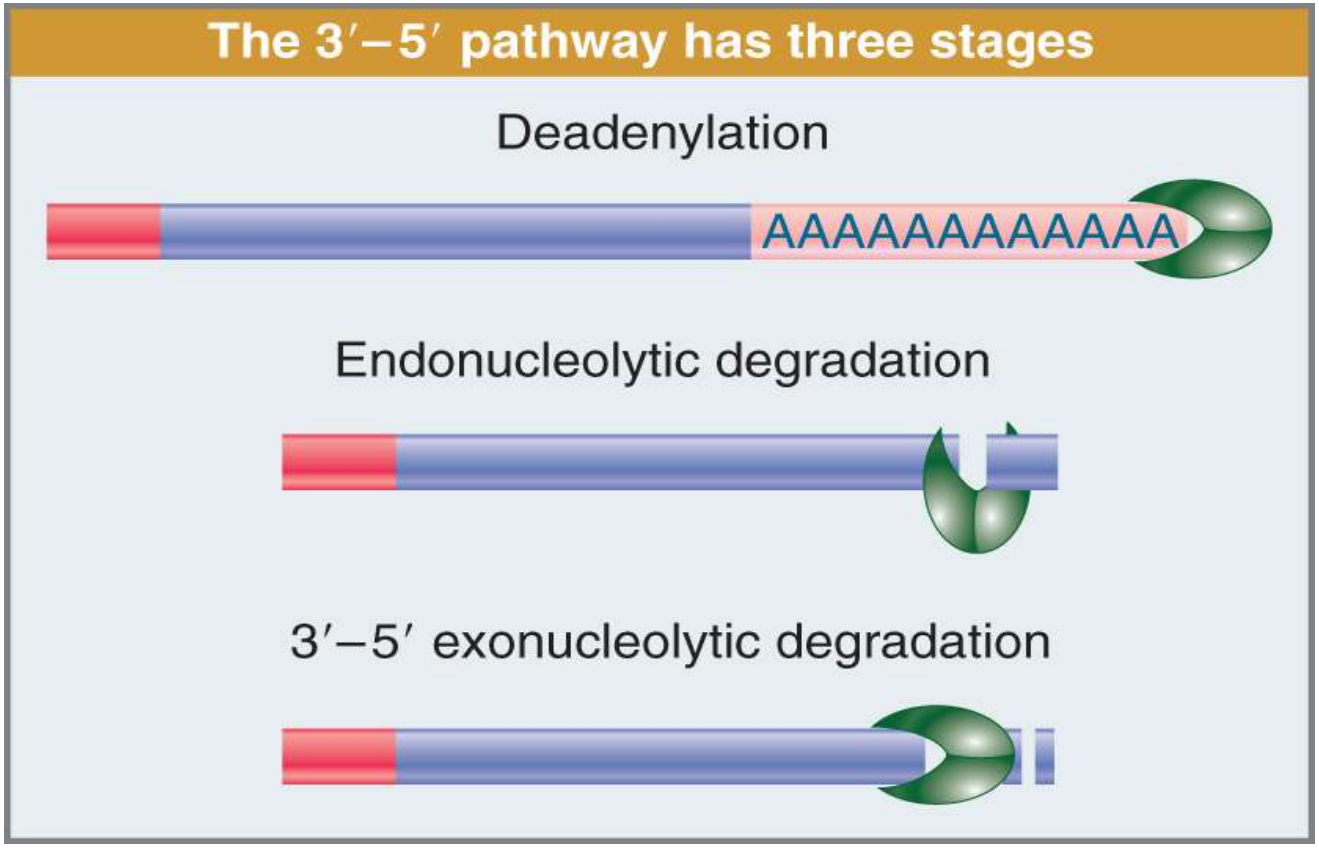


Figure 7.25

The mechanism of eukaryotic pre-mRNA splicing.

Pre-mRNA splicing.

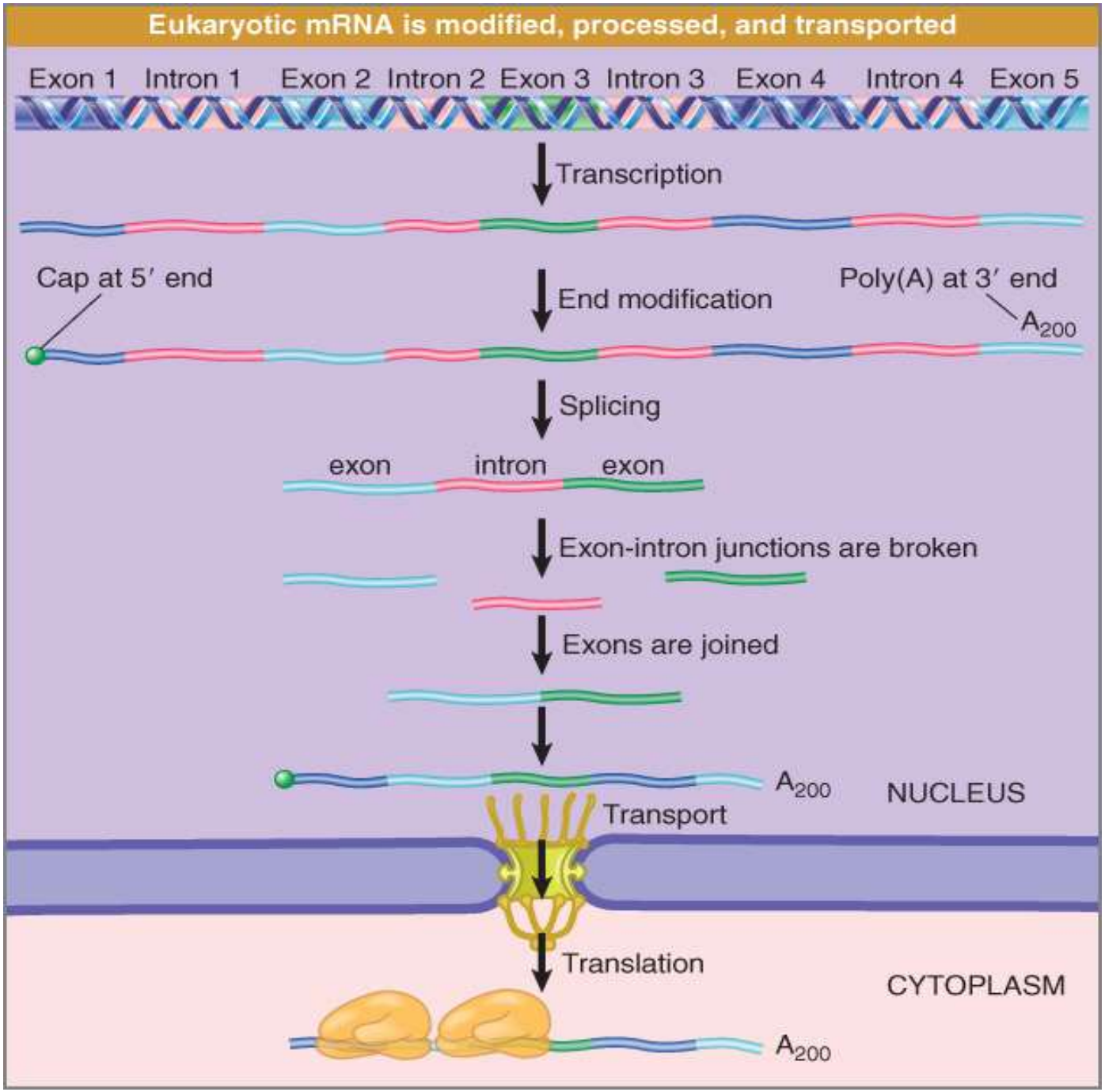
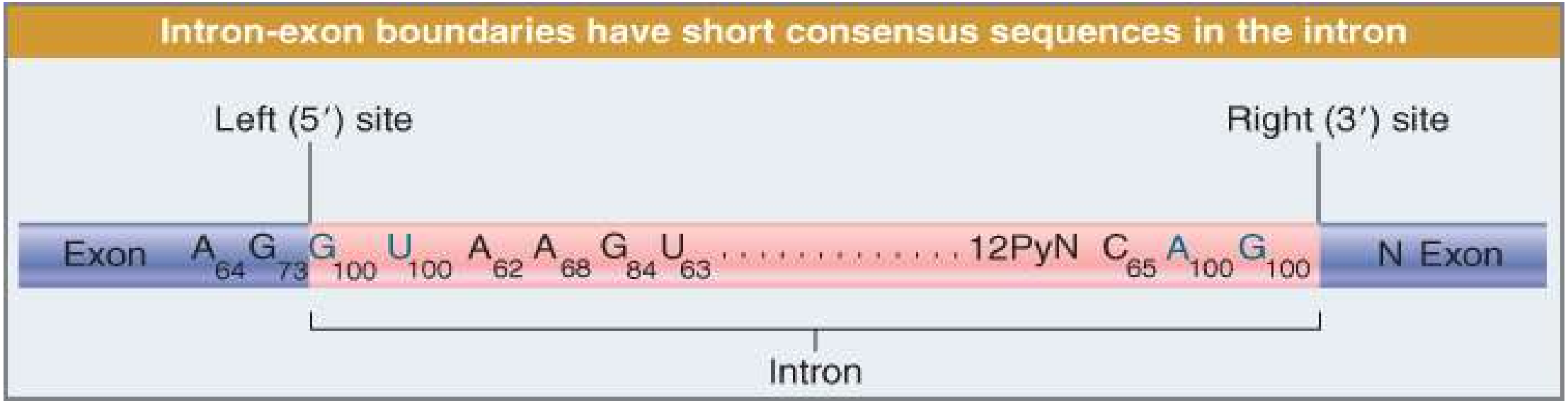


Figure 26.2

Consensus 5’ and 3’ splice sites.
The GU-AG rule



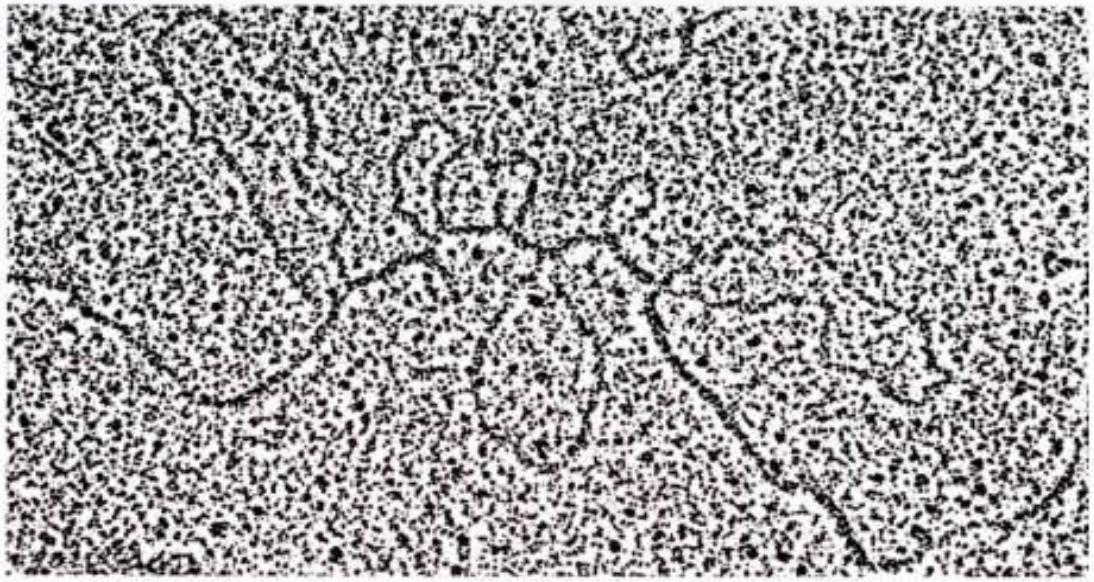
Protein coding sequence must be ‘in frame’ across the splice junction.

Remember GU…AG are splice junction sequences. Mnemonic; GULAG maybe?

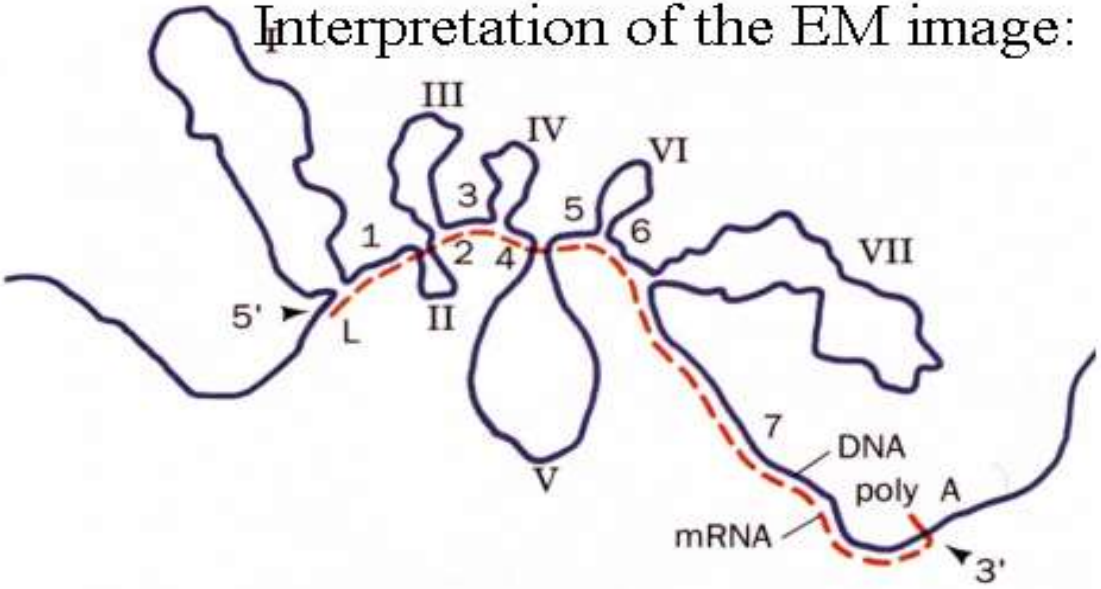
Intron sequences are generally not conserved even between closely related genes.

Figure 26.3

EM heteroduplex analysis of Adenovirus RNA transcripts annealed to Adenovirus DNA showed formation of D (displacement) loops.



Interpretation of the EM image:



Annealing RNA from virus-infected cells with viral DNA revealed the existence of seven introns-transcribed regions of the DNA removed from the mature mRNA.

Splicing must preserve the open reading frame.

- 5’ and 3’ splice site have to be in the same reading frame.
- Reading frames are 1, after first base of codon, 2 after second base and 3 after third base, ie. between codons
- GU..AG consensus are the ends of the intron to be removed, not in the actual coding sequence.

Splice Junctions Are Read in Pairs

- All 5′ splice sites are functionally equivalent, and all 3′ splice sites are functionally equivalent.
- Usually draw exons as boxes to emphasise them with splices joining box corners and introns as lines between the boxes .
- Gene rearrangements within big introns have a good 1/3 chance to produce fusion proteins. (Exon shuffling in evolution).

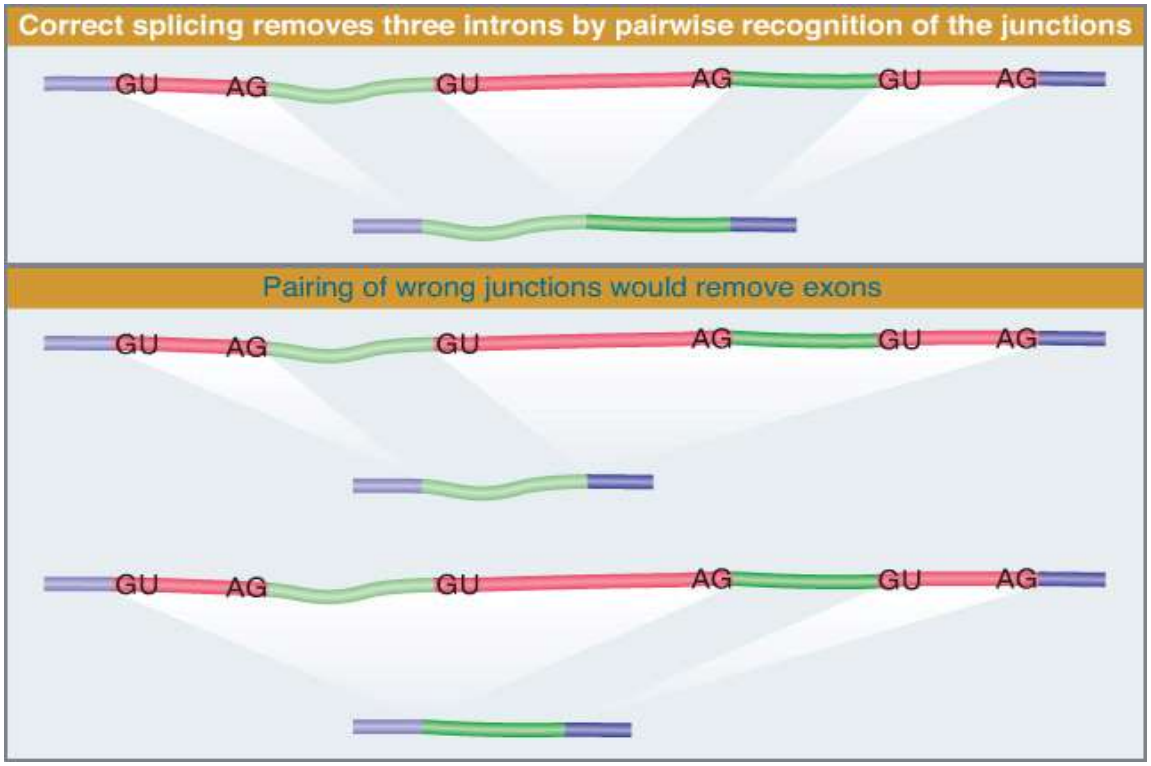
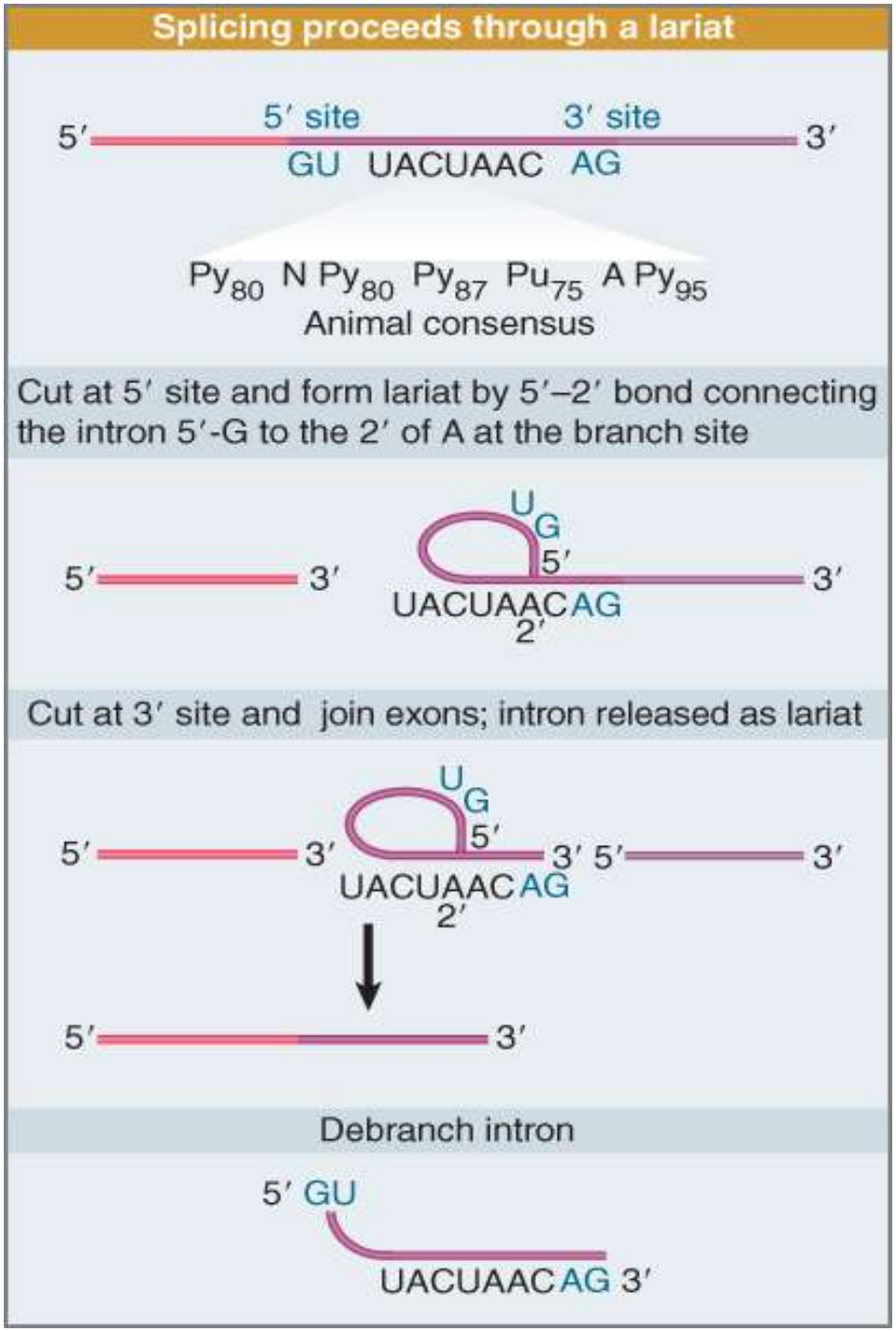


Figure 26.4


Splicing proceeds through a lariat
5′ site
3′ site
5′
3′
GU UACUAAC AG
Py80 N Py80 Py87 Pu75 A Py95
Animal consensus
Cut at 5′ site and form lariat by 5′–2′ bond connecting the intron 5′-G to the 2′ of A at the branch site
U
G
5′
5′
3′
3′
UACUAACAG
2′
Cut at 3′ site and join exons; intron released as lariat
U
G
5′
5′
3′
3′ 5′
3′
UACUAACAG
2′
5′
3′
Debranch intron
5′ GU
UACUAACAG 3′

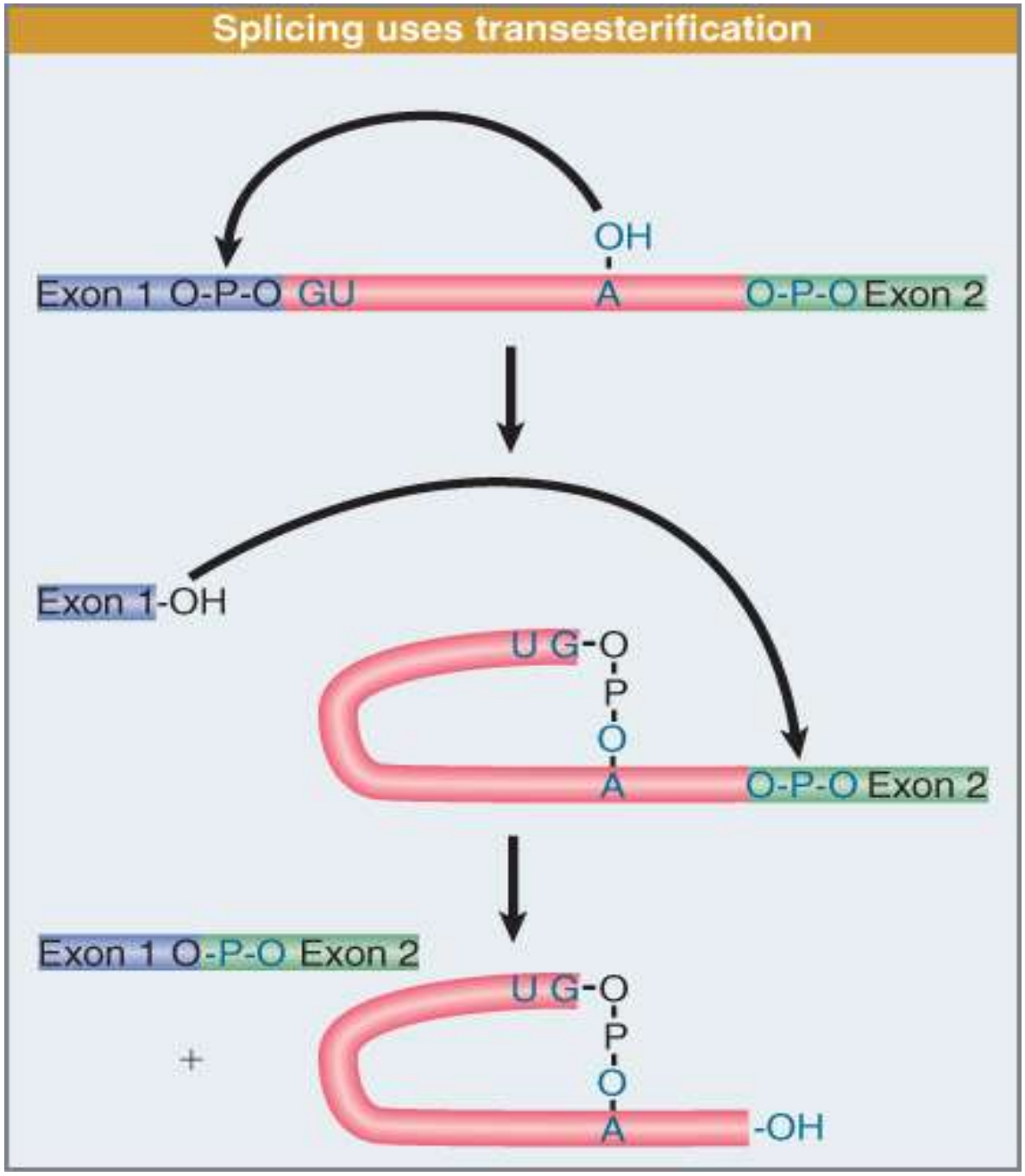


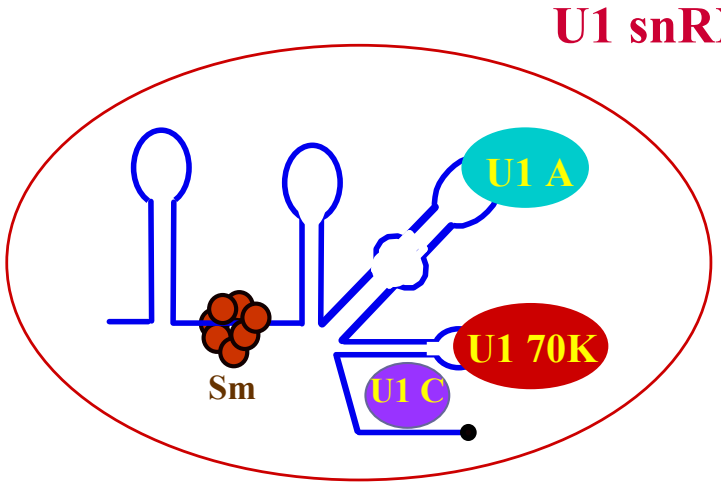
Figure 26.7

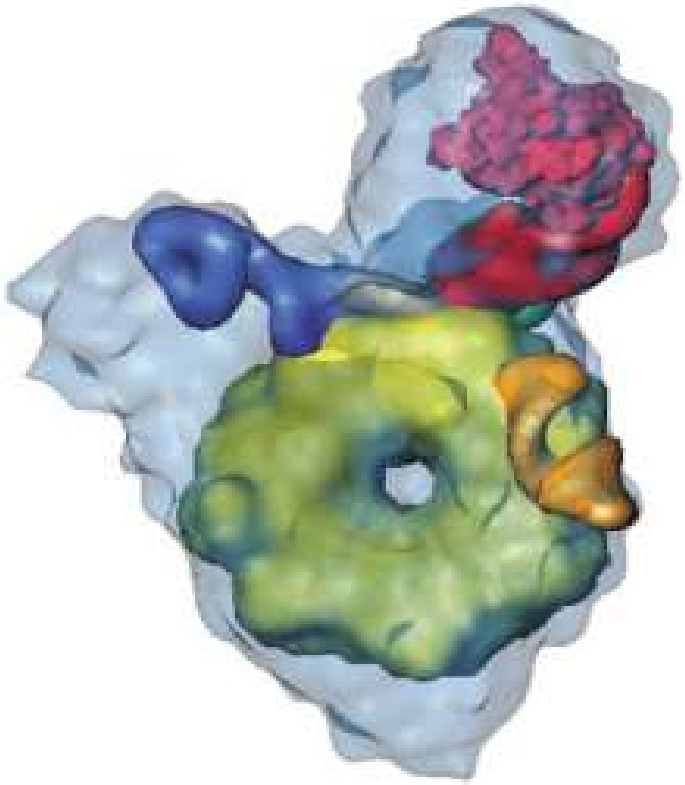
Alkaline lysis of ssRNA.

The reactive 2’ OH groups on riboses can also mediate complete cleavage of 5’-3’ phosphate bonds in single stranded RNA under alkaline conditions

Splicing is catalysed by a series of five small nuclear ribonucleoprotein particles (snRNPs, 'SNURPs'). Conserved small structured RNAs with Sm and other proteins bound.

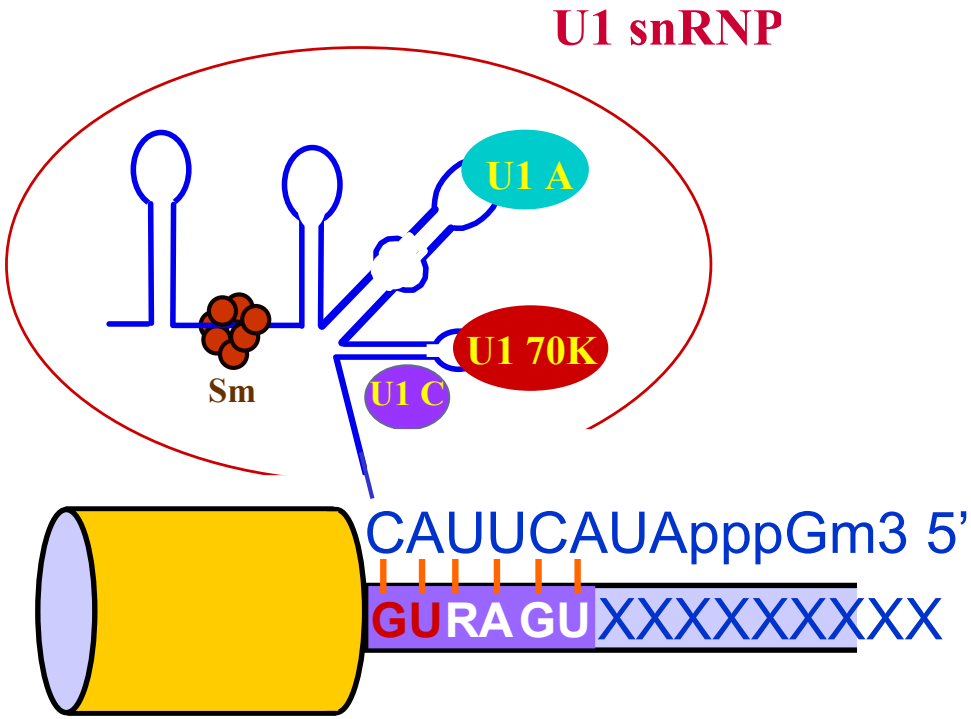
SNURPs assemble sequentially on pre-mRNA to form a spliceosome.



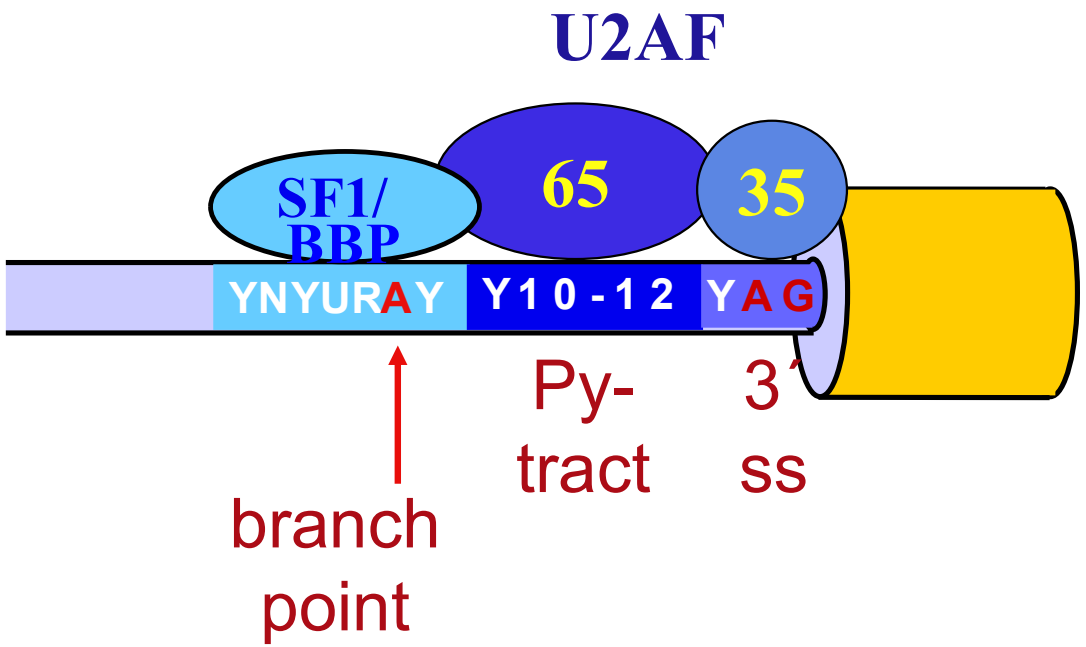


Cryo-EM picture
(Stark et al. Nature 409, 539)

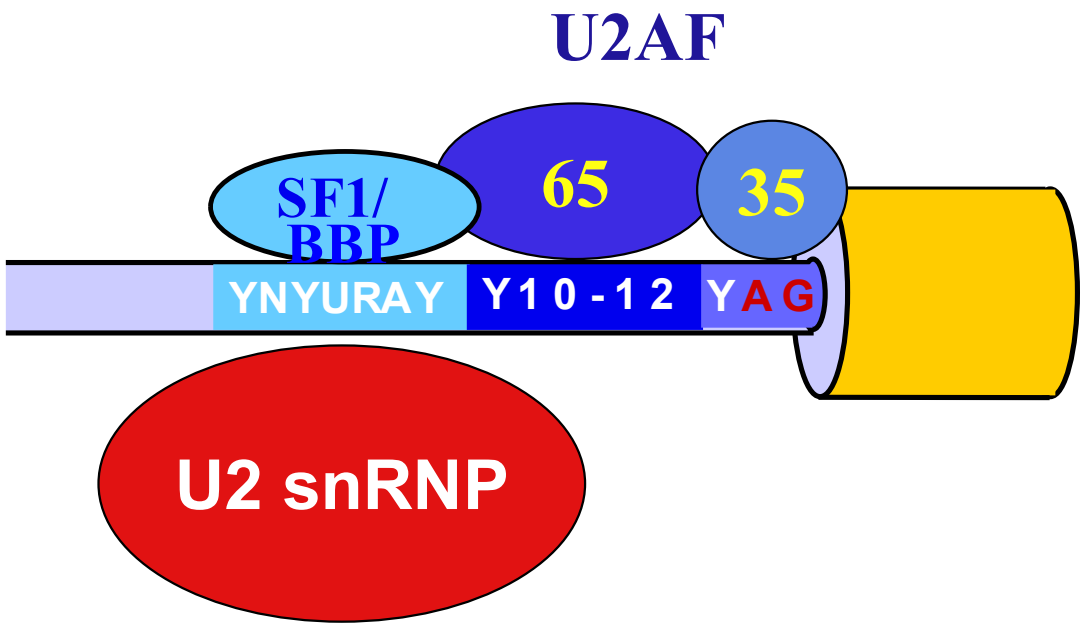
5’ Splice Site Recognition



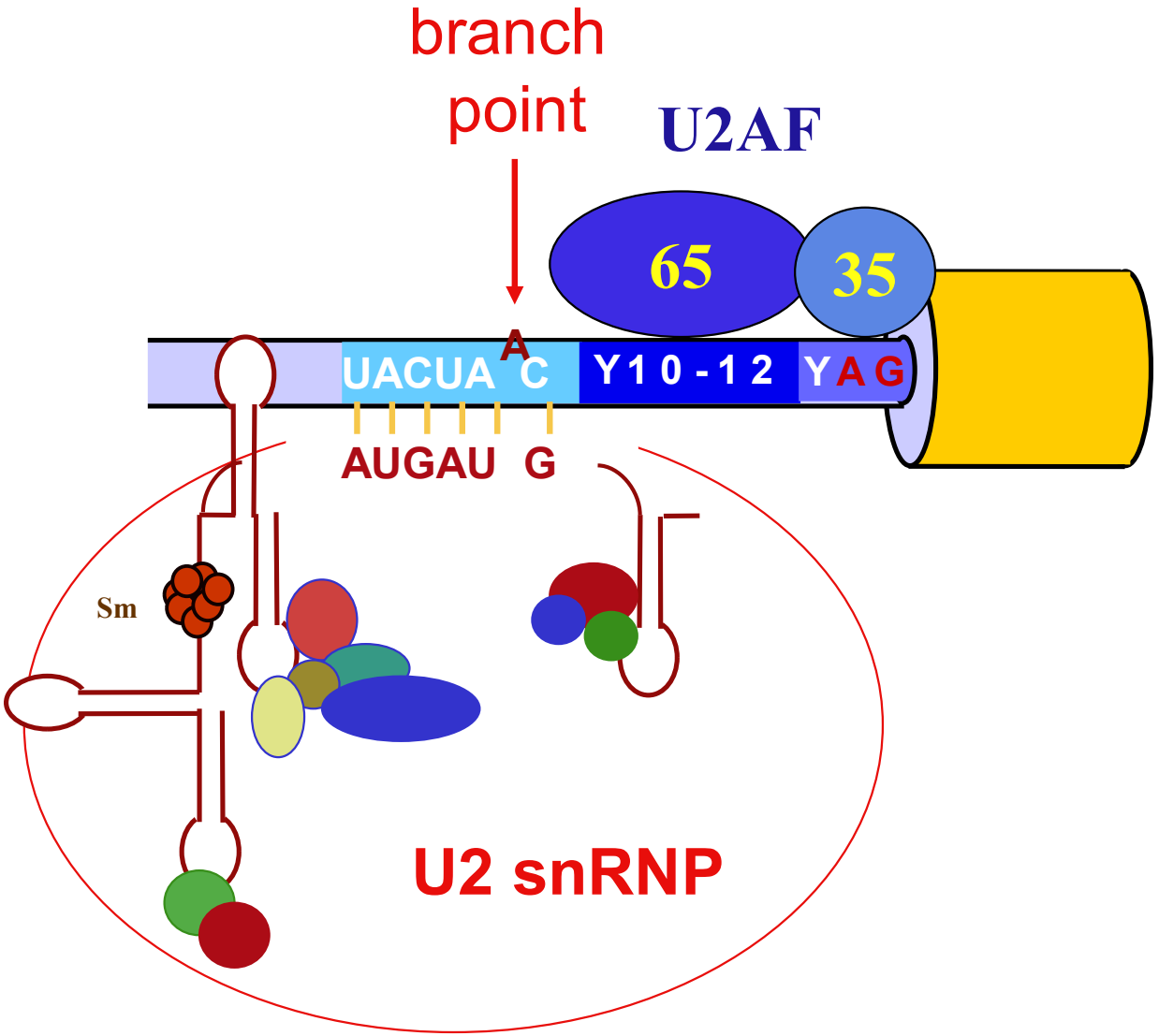
3’ Splice Site Recognition



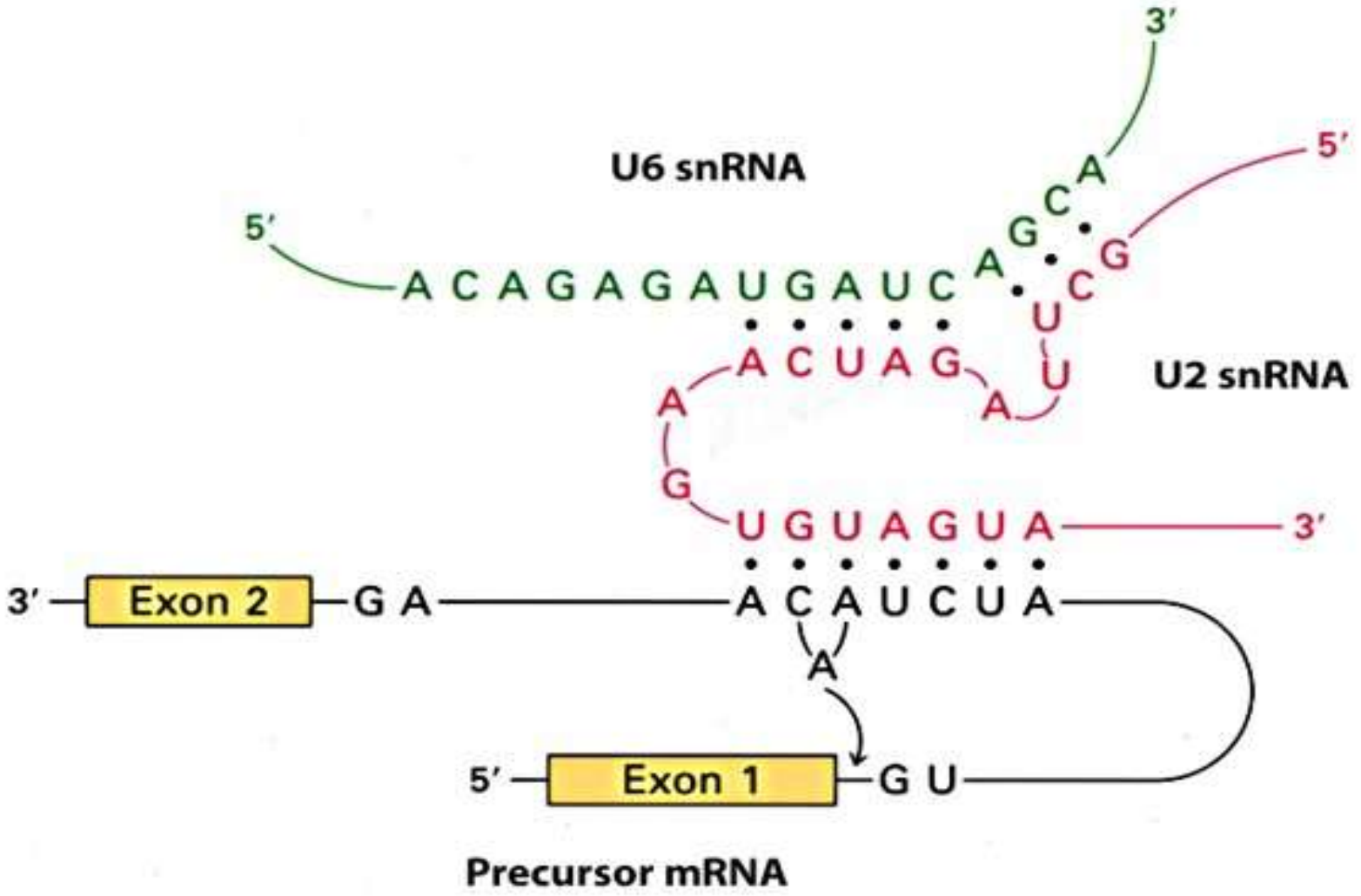
3’ Splice Site Recognition



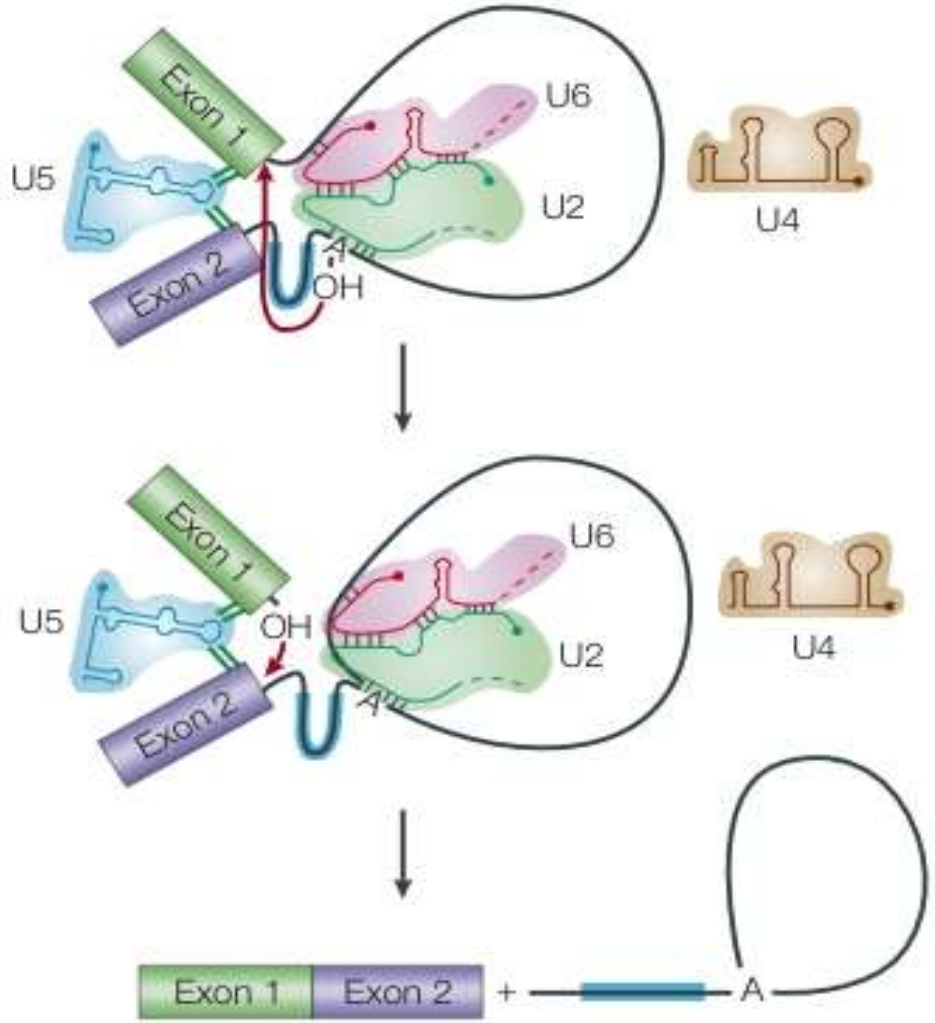
3’ Splice Site Recognition



The U4/U5/U6 tri snRNP arrives last and base pairs to U2.

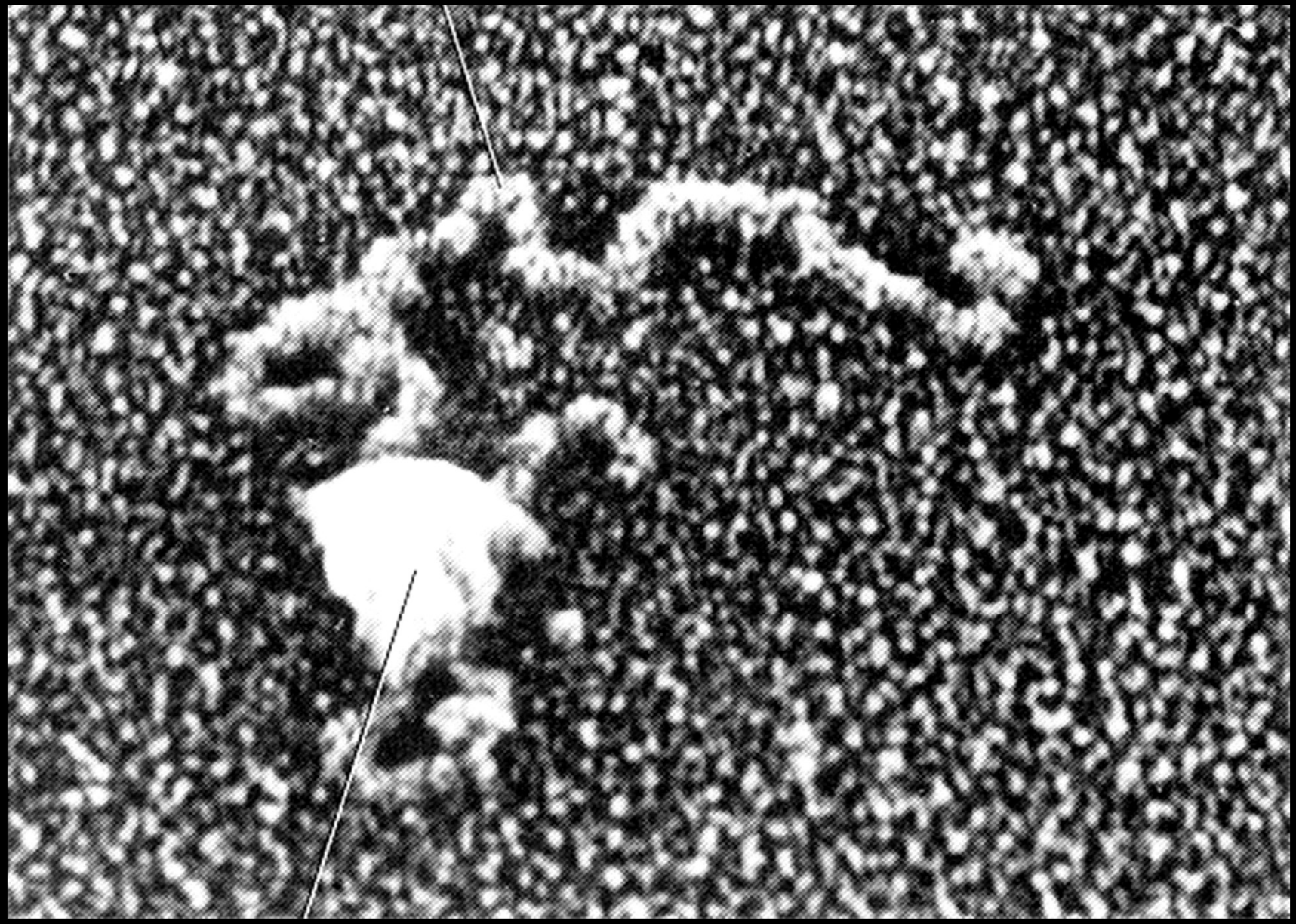


Conformation changes bring splice junctions together.



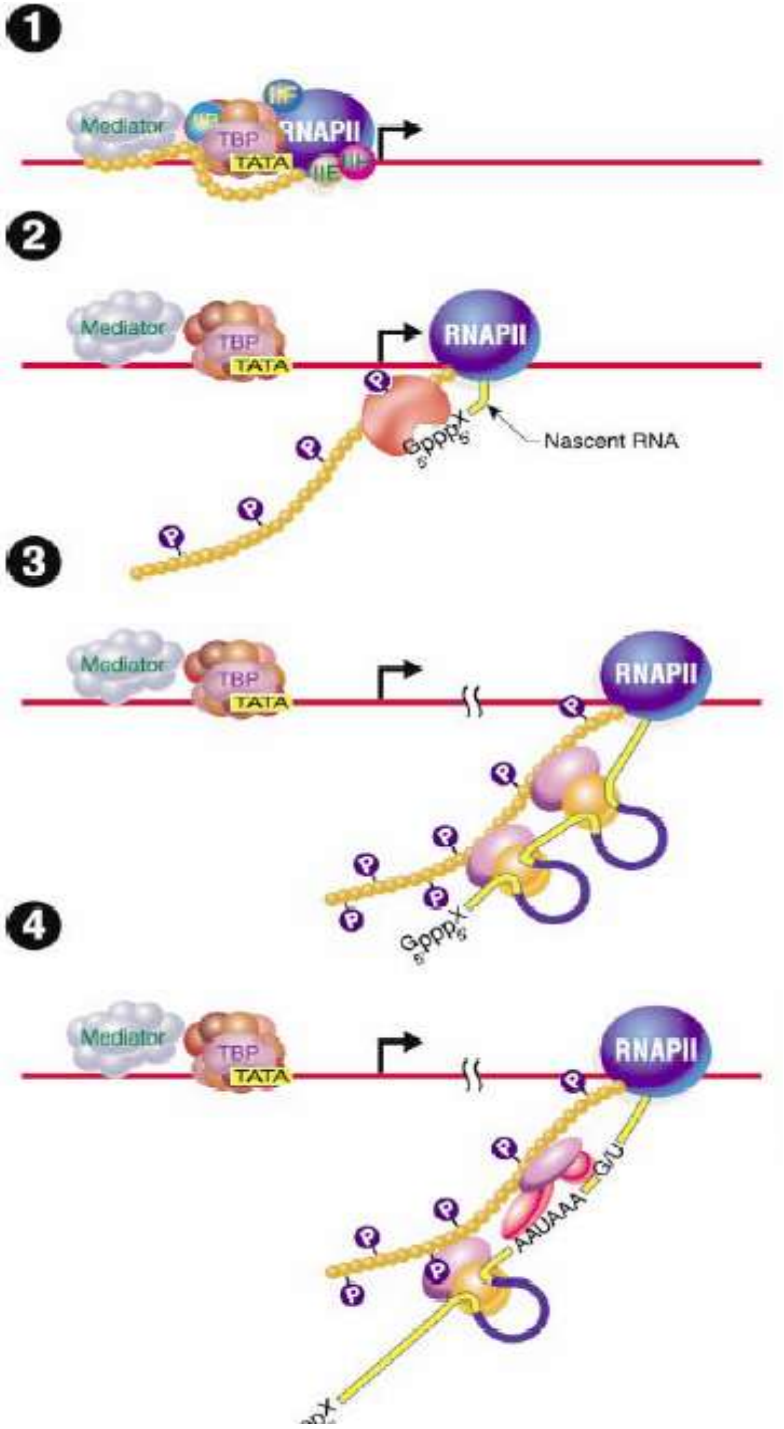
Patel and Steitz, Nature Reviews Molecular Cell Biology (2001) 4, 960-970

Pre-mRNA



spliceosome

RNA processing is coupled to transcription.



Splicing Is Connected to Export and efficient translation of mRNA

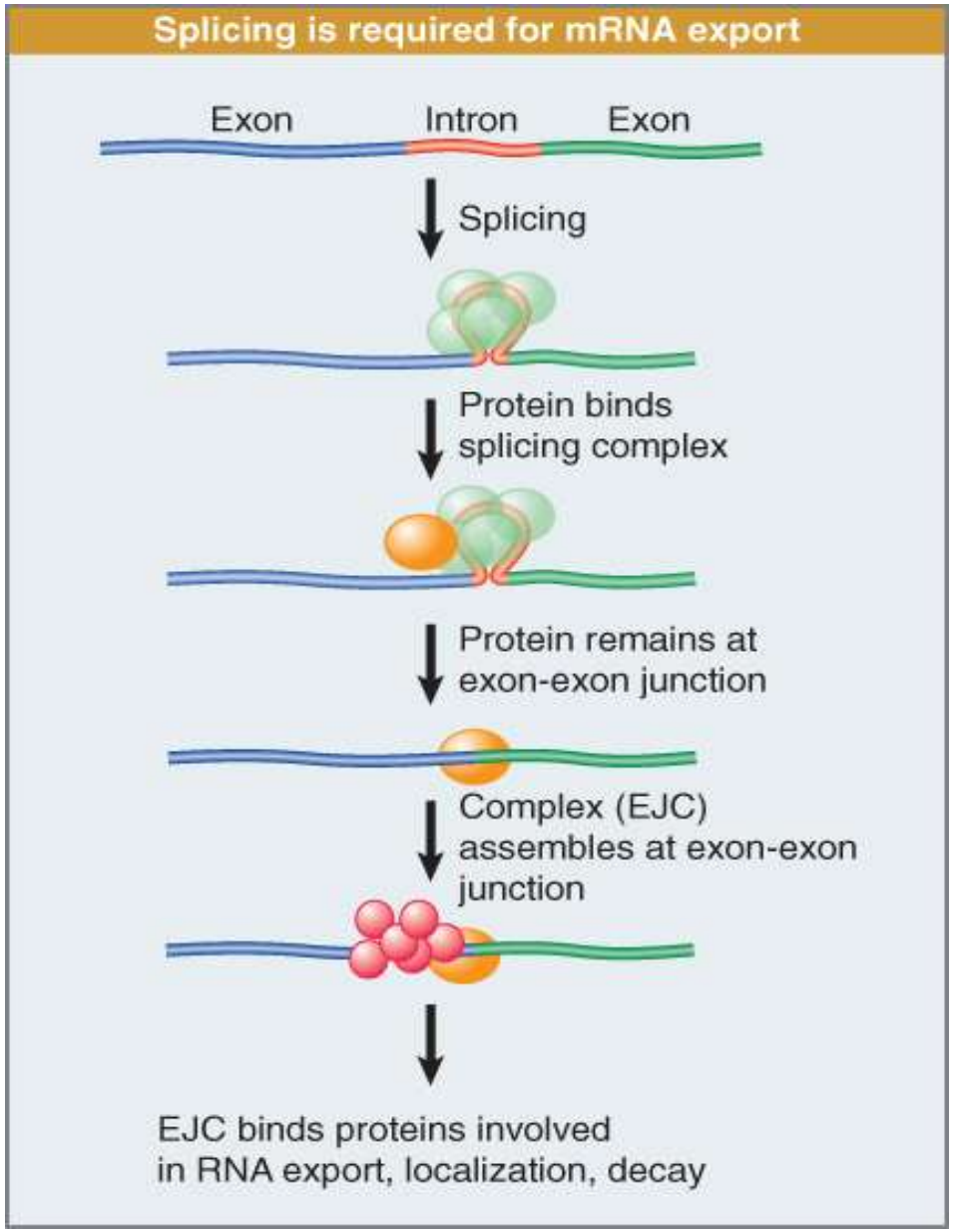


Figure 26.16

Alternative splicing is an important strategy to increase the number of protein isoforms expressed from genes.

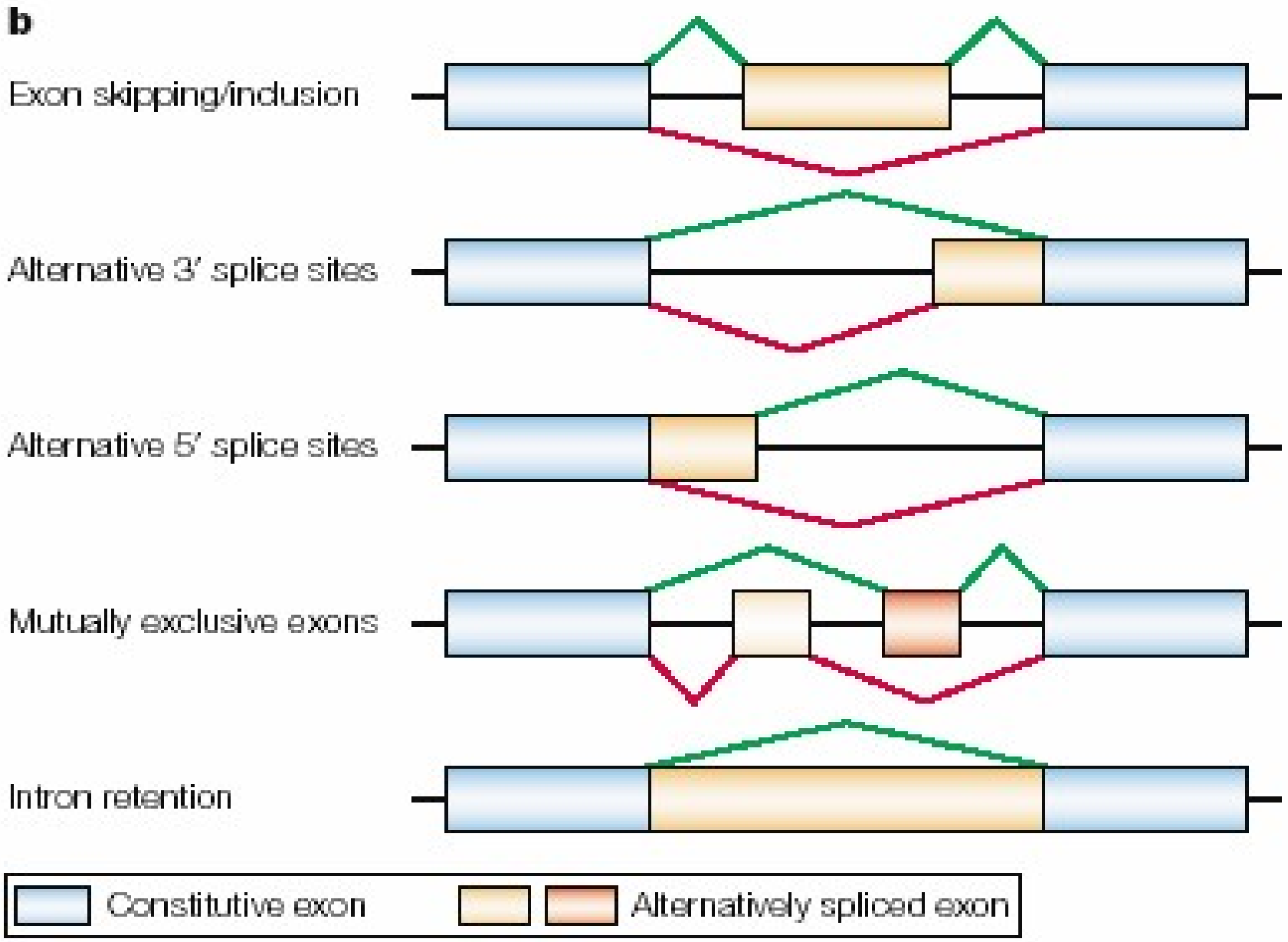
The importance of alternative splicing

It is estimated that 95% of human genes are alternatively spliced from new sequencing information.

Alternative splicing creates protein diversity
26,000 human protein-encoding genes,
yet 90,000 proteins

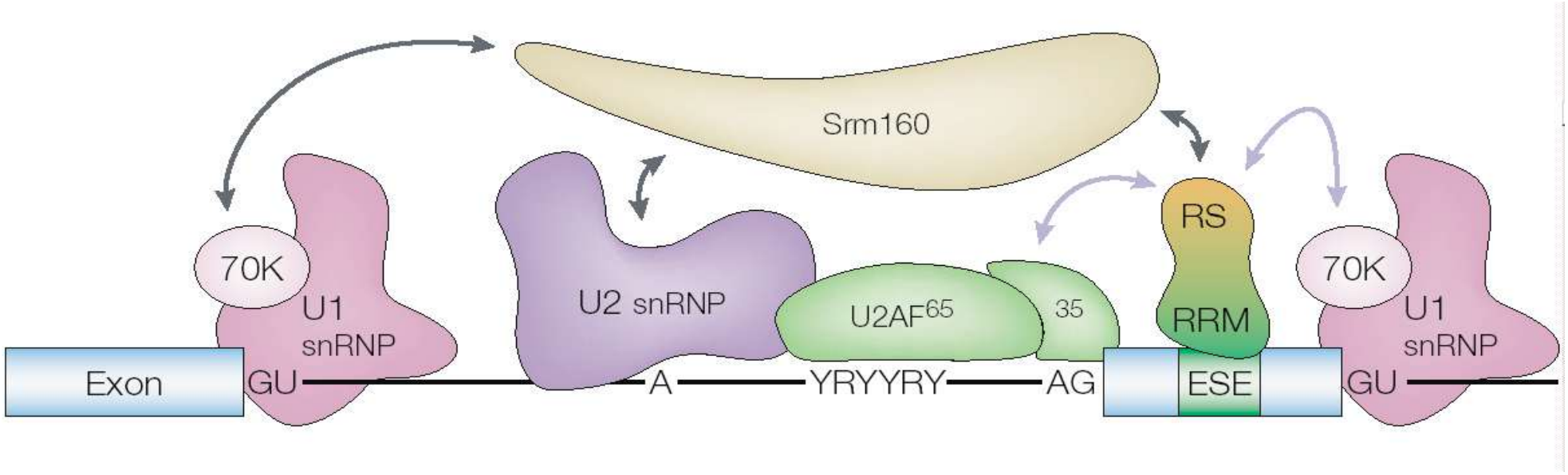
Aberrant splicing is a frequent effect of human disease gene mutations.

Modes of alternative splicing.



Cartegni L. Chew SL. Krainer AR.

Inclusion of alternative exons may be promoted by exonic splicing enhancers that bind serine/arginine-rich (SR) RNA-binding proteins.



Heterogeneous ribonuclear proteins (hnRNP proteins) coat pre-mRNAs and mRNAs and often inhibit splicing events that SR proteins promote.

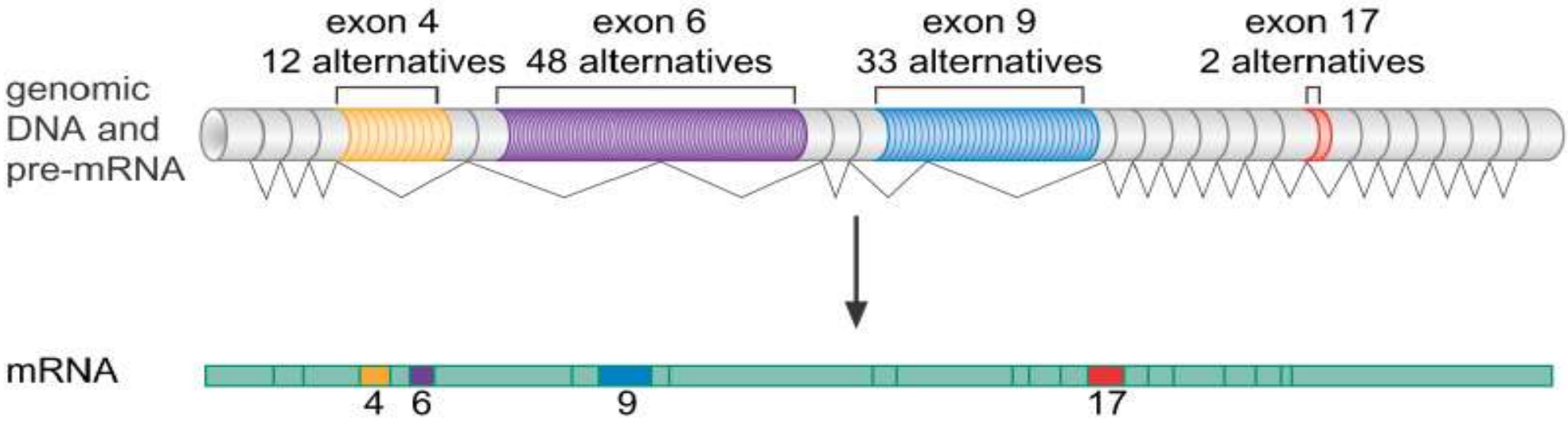
'Ying-yang' balance of these effects on many splicing events.

Splicing complexes Can be Formed by Intron Definition or Exon Definition

- The direct, **intron definition**, way of forming an E complex at short introns is for:
  - U1 snRNP to bind at the 5′ splice site
  - U2AF to bind at a pyrimidine tract between the branch site and the 3′ splice site

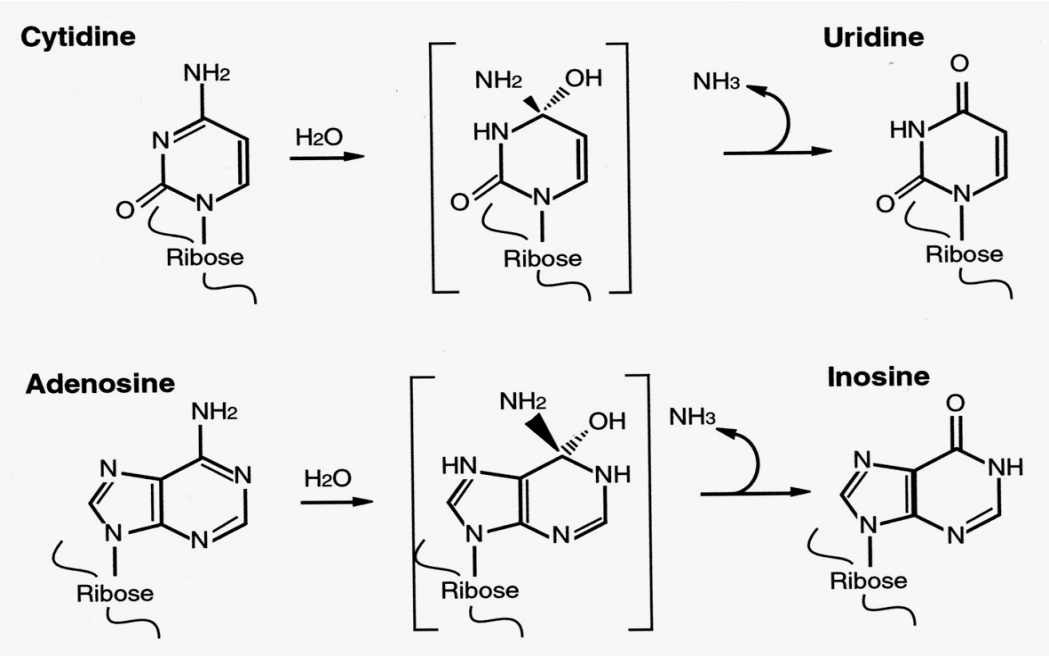
- Another possibility, **exon definition**, for very long introns is for the complex to form between:
  - U2AF at the pyrimidine tract
  - U1 snRNP at a downstream 5′ splice site

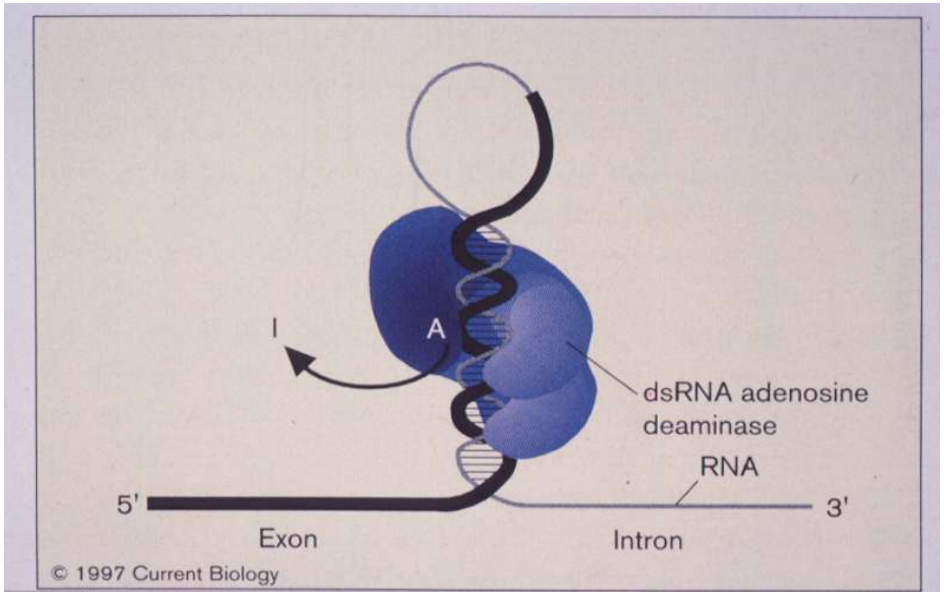
Alternative splicing in *Drosophila Dscam*.



fg13-11

Base deamination RNA editing also occurs site-specifically in some transcripts and changes codon meaning.





Adenosine deaminases acting on RNA (ADARs) edit 4% of *Drosophila* transcripts (972 sites), and ~250 known sites in human transcripts, including those encoding brain ion-channel subunits, particularly GluR B, the dominant subunit of AMPA class glutamate receptors.

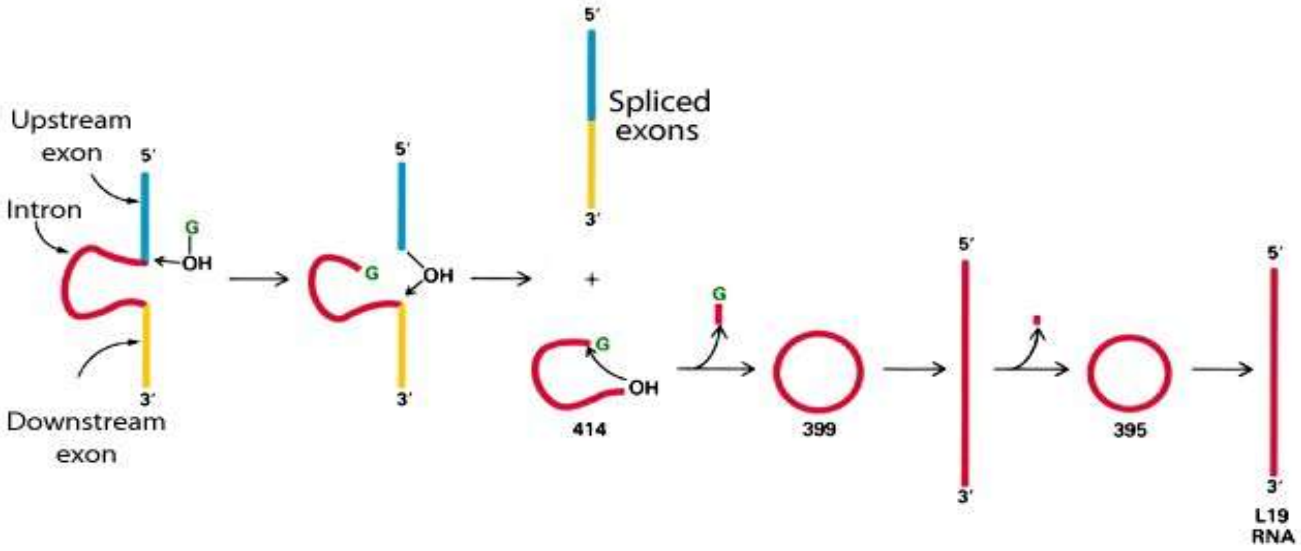
APOBECs are C to U deaminases first identified in RNA ediding but more important as DNA editors. Activation induced deaminase (AID) edits DNA in immunoglobulin genes to initiate Somatic Hypermutation of antibody variable regions. APOBEC3G edits HIV and interferes with virus replication.

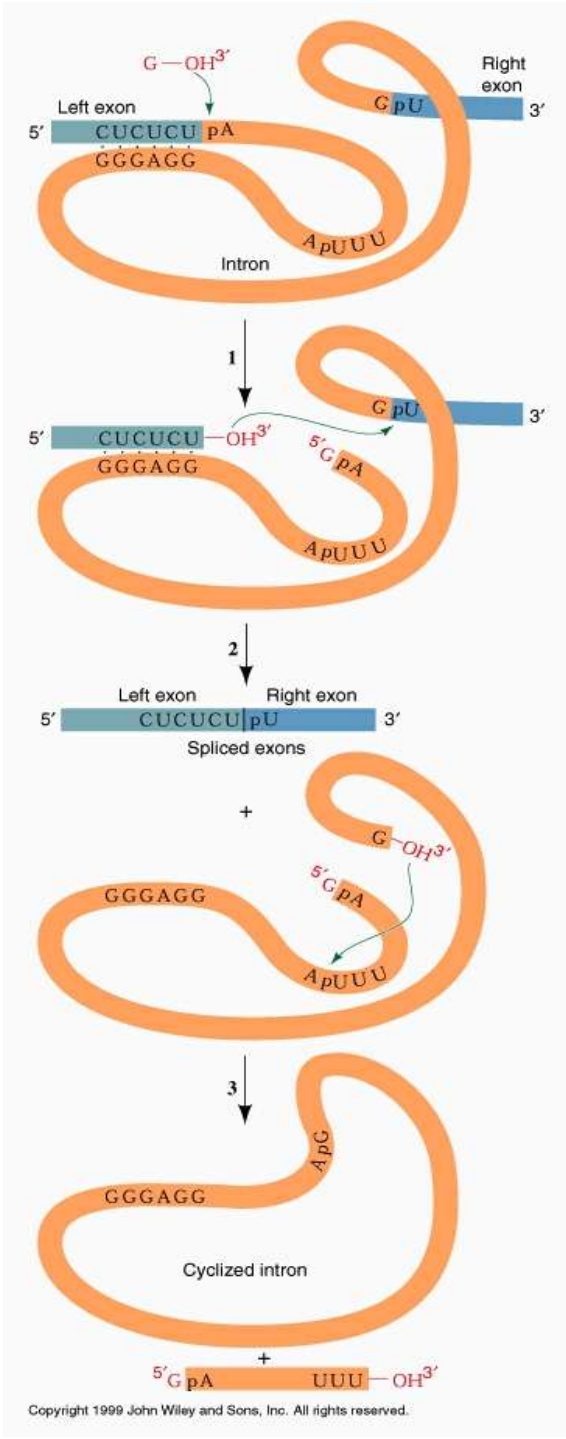
RNA catalysis, Ribozymes and the RNA World.

The RNA World.

Self-splicing introns are small transposable elements that remove themselves cleanly from the transcripts they are inserted in.

Group I intron splicing in 26S rRNA of *Tetrahymena* is catalysed by RNA in the absence of protein (Tom Cech).





The RNA World idea - RIBOZYMES.

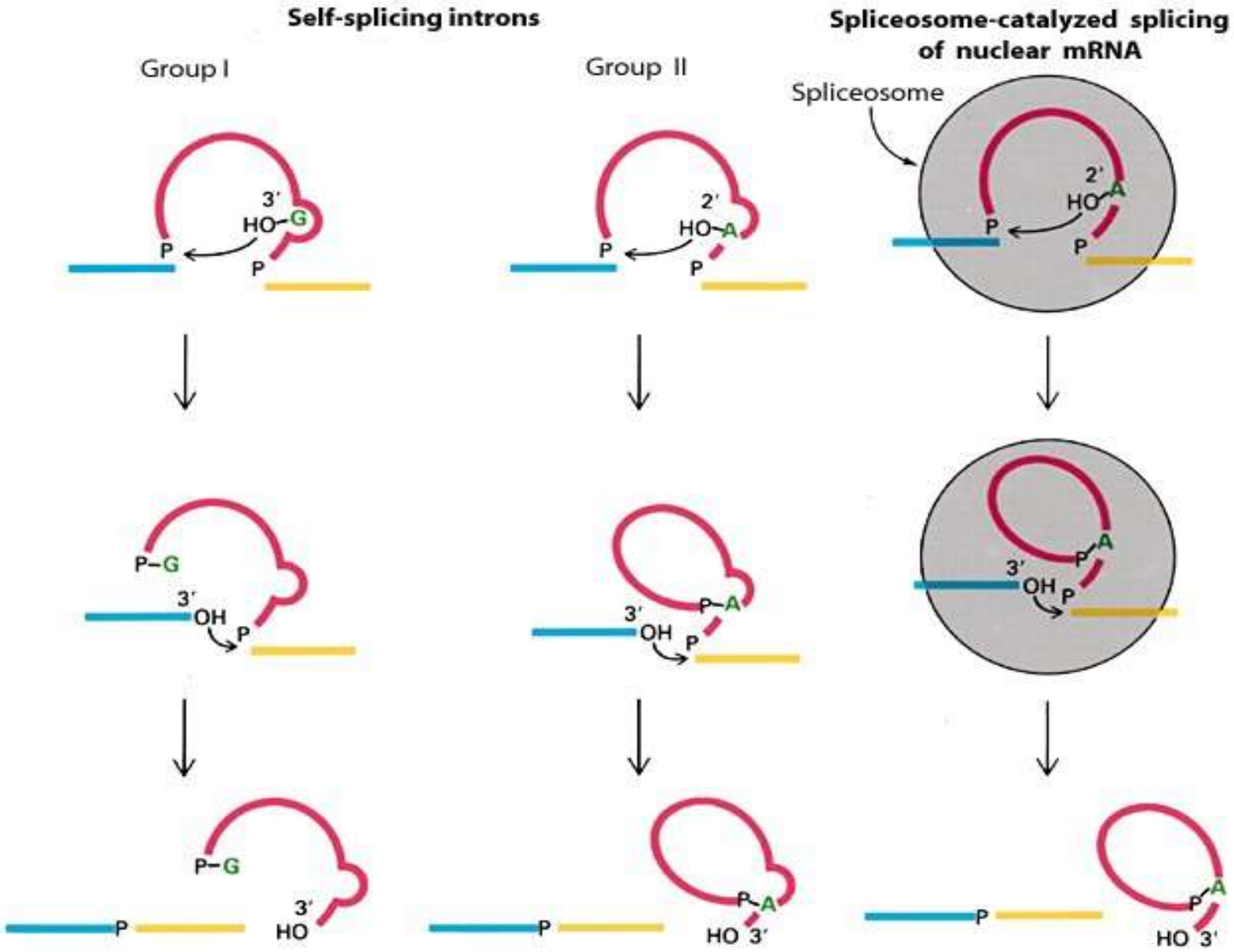
- Ribosomes and spliceosomes function through RNA-RNA interactions. Before the self-splicing introns were discovered it was assumed that catalysis would be caused by the proteins in these complexes.

- The discovery that RNA alone is catalytic in the Group I self-splicing intron showed that the RNA could be catalytic in ribosomes and spliceosomes also. RNA enzymes are called **RIBOZYMES** and many artificial ones have been created.

- This contributes to our understanding of possible origins of life as RNA alone could have carried out its own replication using natural ribozyme activities until translation was invented as a way to make more versatile protein catalysts.

- DNA also may have been a later invention for use as a genomic library. It is more stable to hydrolytic cleavage than RNA and more readable by proteins than dsRNA to allow gene control.

The RNA World.
Group II introns also self-splice by RNA catalysis (Cech).



Nuclear splicing is similar to Group II self-splicing introns

spliceosome evolved from Group II introns.

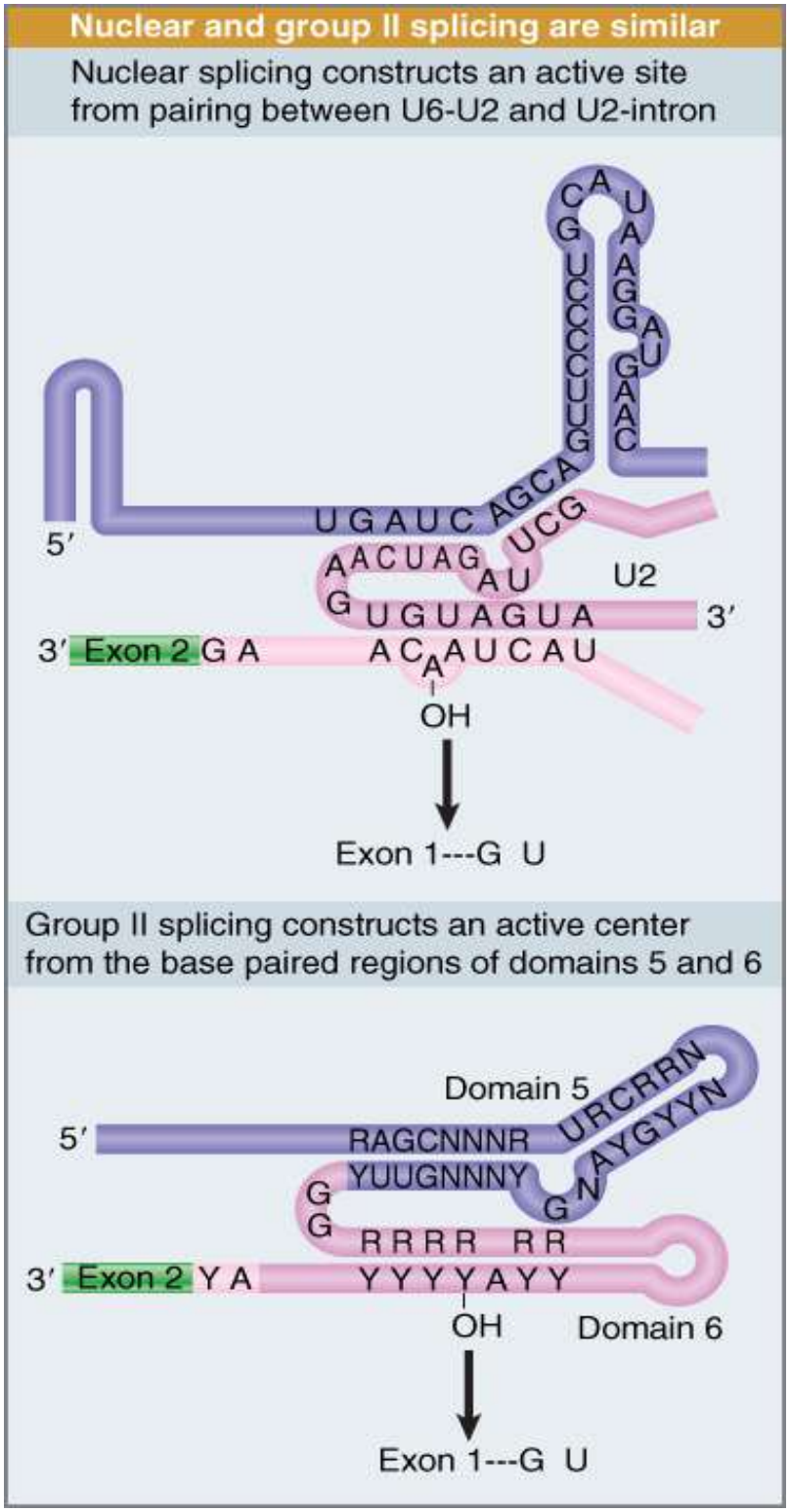


Figure 26.20

How introns might have spread.
Mobile Group II introns.

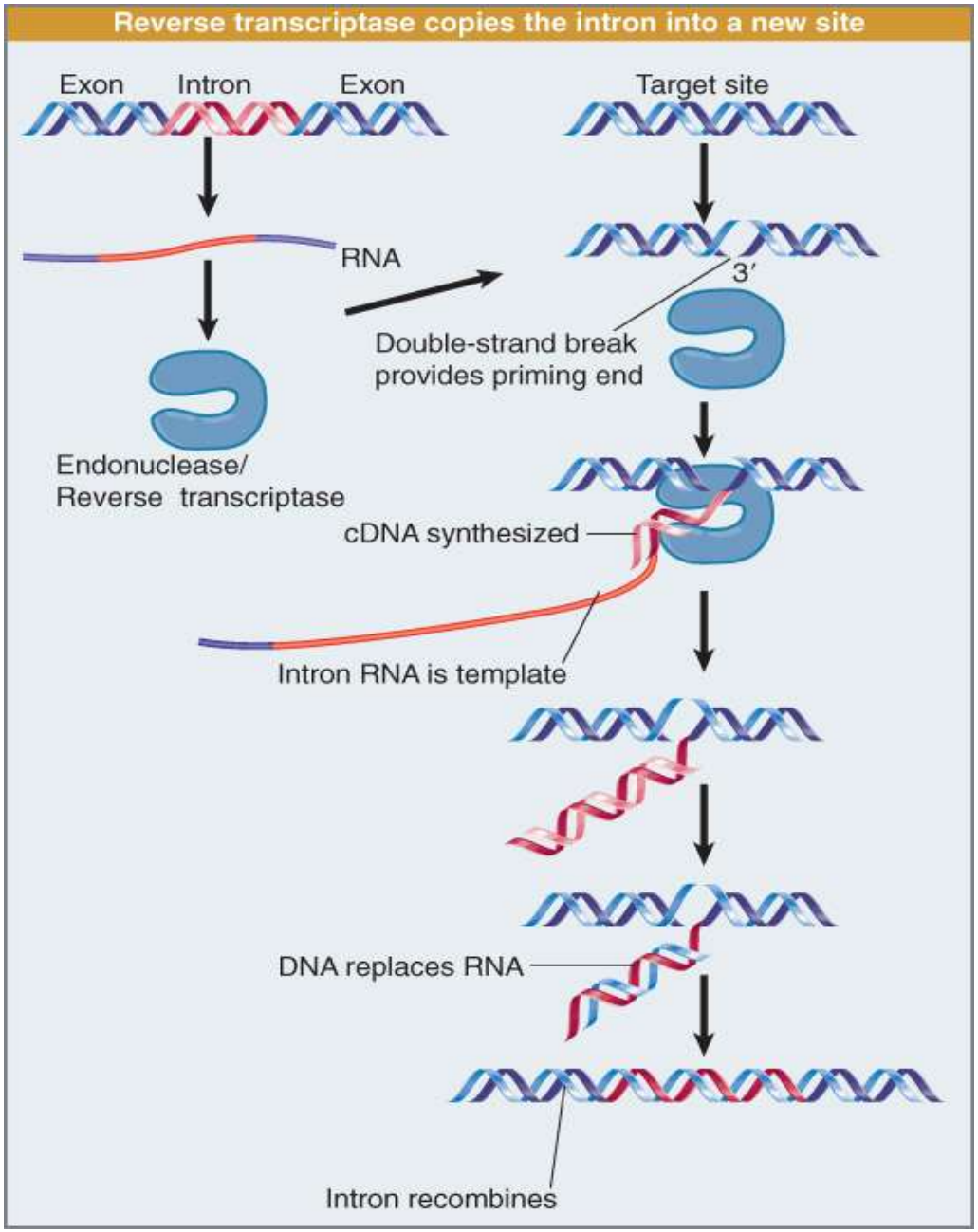
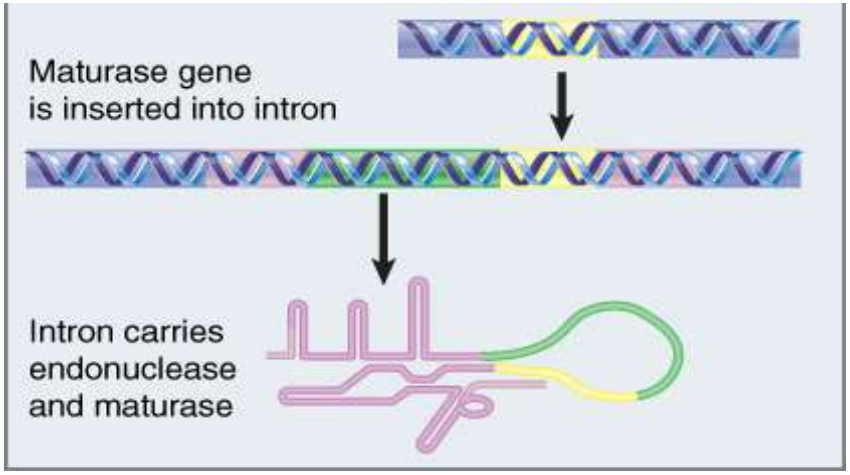


Figure 27.13

How spliceosomal introns might have evolved from self-splicing Group II introns.



- Autosplicing introns may require **maturase** protein activities encoded within the intron.
  - They assist folding into the active catalytic structure.
- Imagine that such an intron loses its maturase but can use maturase provided by other similar introns. **Maturase** becomes trans-acting **spliceosome**.
- This intron might eventually evolve to consist of just the target splice junctions GU..AG with all the catalytic activities supplied by a trans-acting **splicesome**.

Figure 27.15

Summary.

- Eukaryotic mRNA stability, 5’ cap and polyA addition and splicing.
- Mechanism of pre-mRNA splicing and alternative splicing. RNA editing.
- Self-splicing Group II introns, RNA catalysis, Ribozymes and the The RNA World.