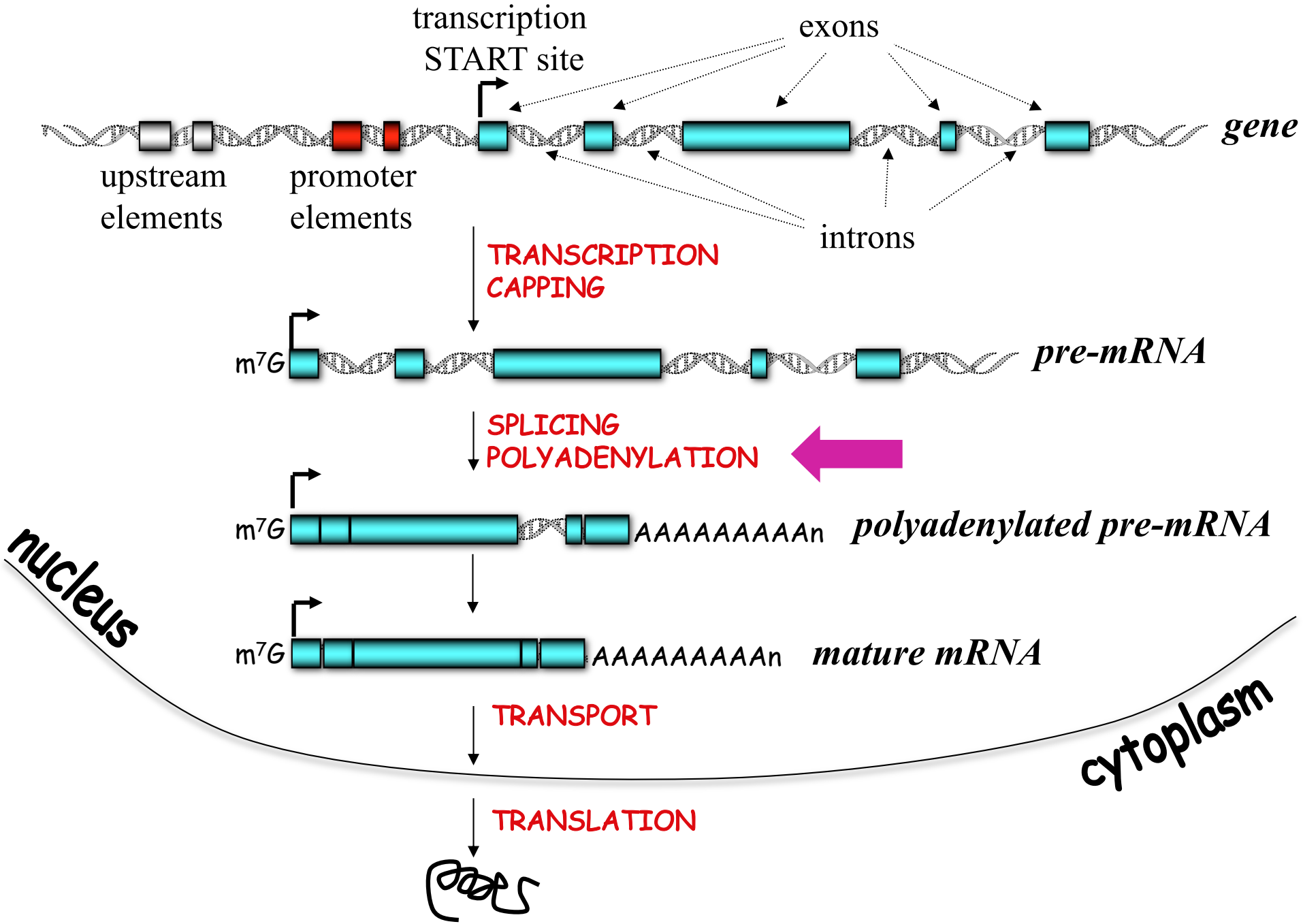
3′-end formation

Eukaryotic gene expression



What is the function of the PAP-dependendent polyA tail?

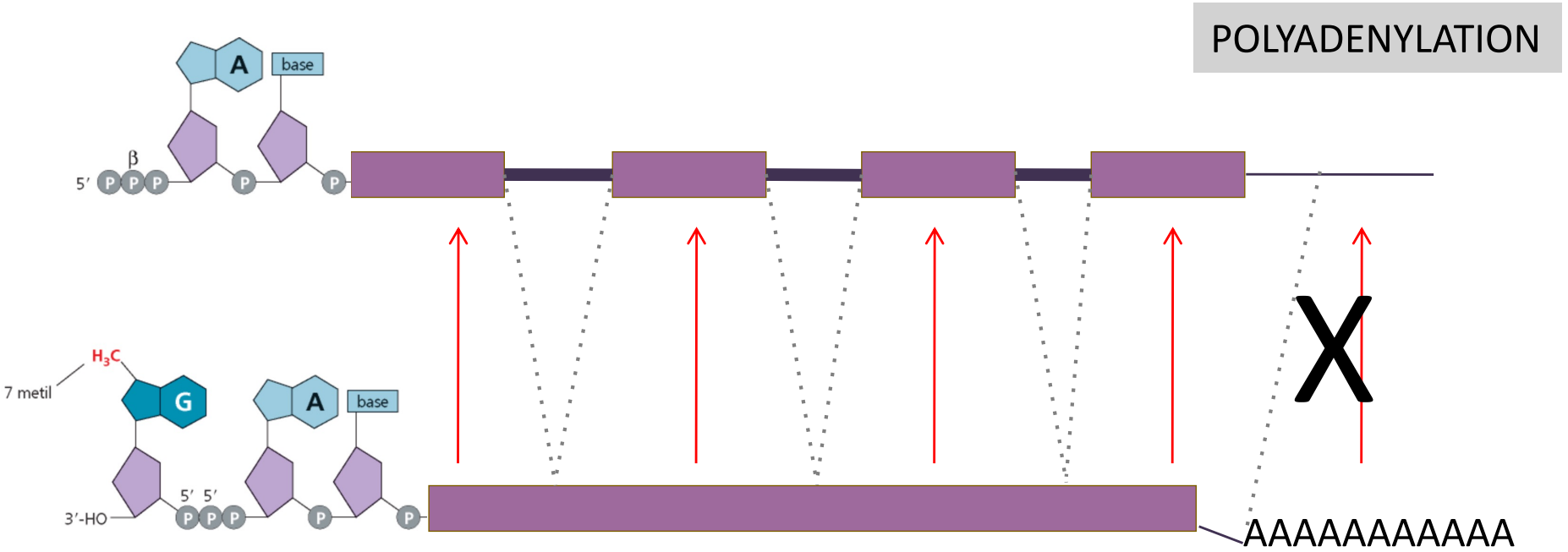
1) Increases RNA stability

2) Favours the mRNA transport to the cytoplasm

3) Increases translation efficiency by favouring the loading of ribosomal 40S subunit

4) Stimulates mRNA transcription termination

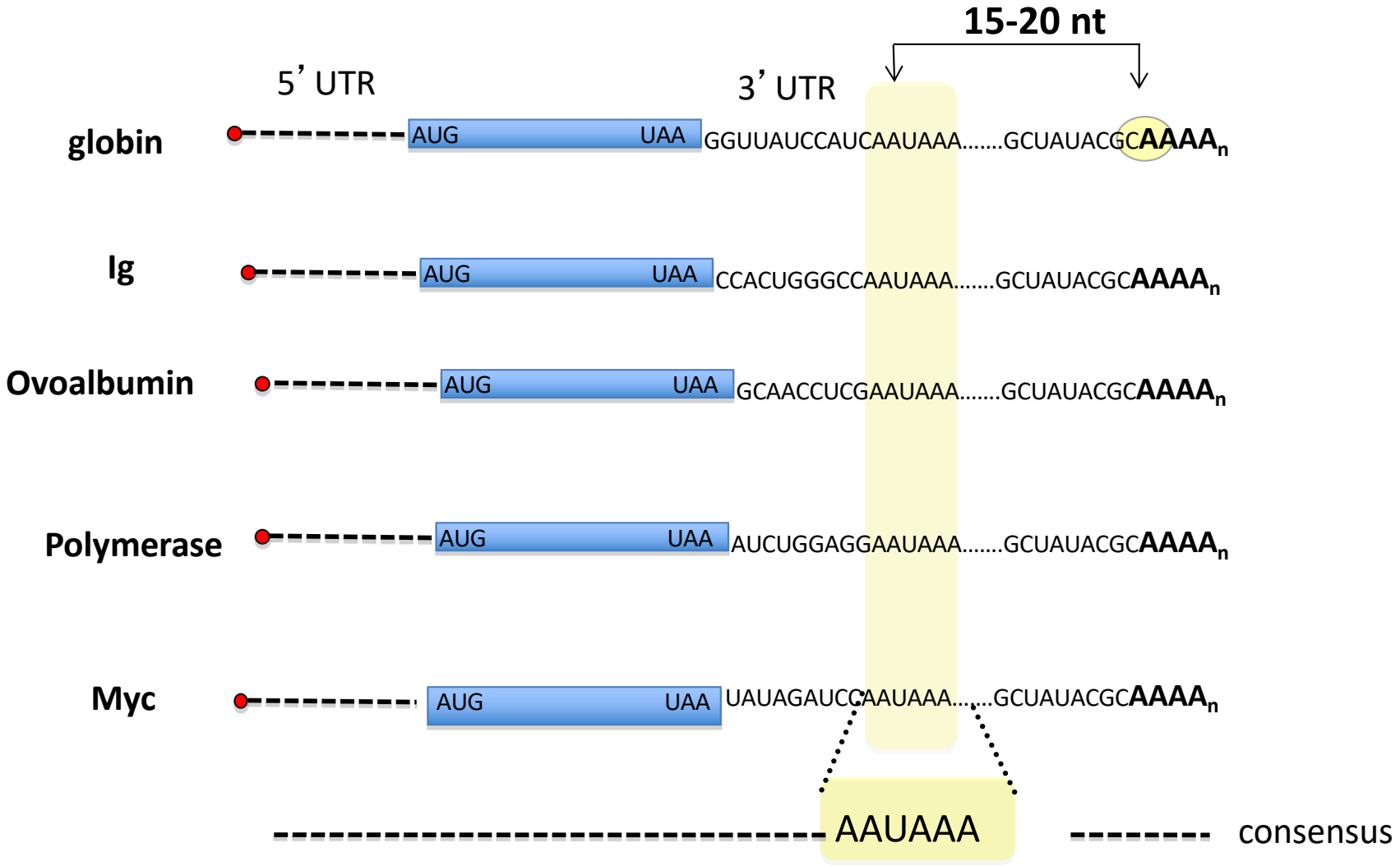
mRNA maturation



The polyA tail is not transcribed from the DNA but it is added to the mRNA during a process named polyadenylation.

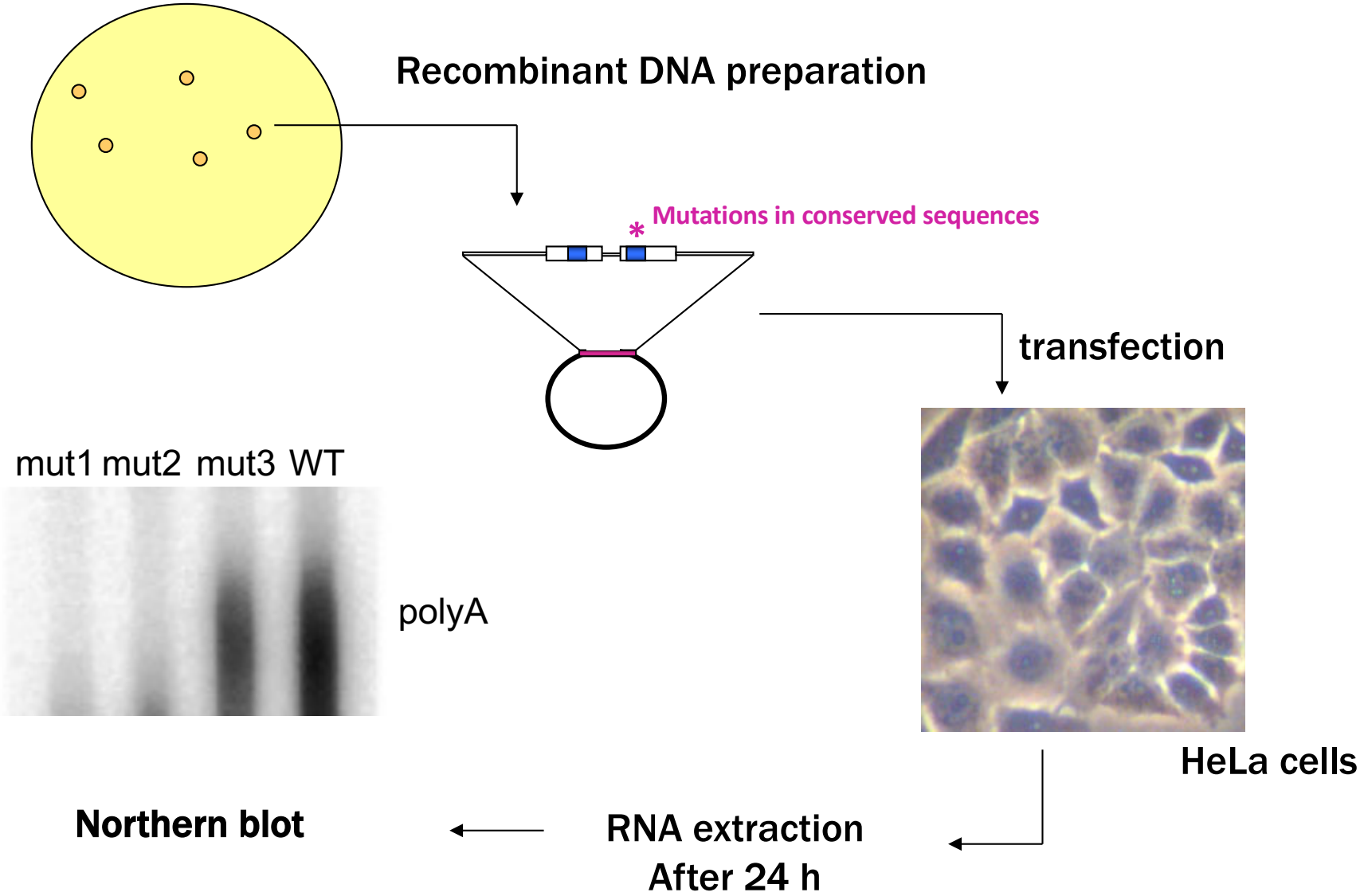
Looking for consensus sequences in mammals

SEQUENCE ALLINEAMENT OF cDNAs STARTING FROM THE POLYA TAIL

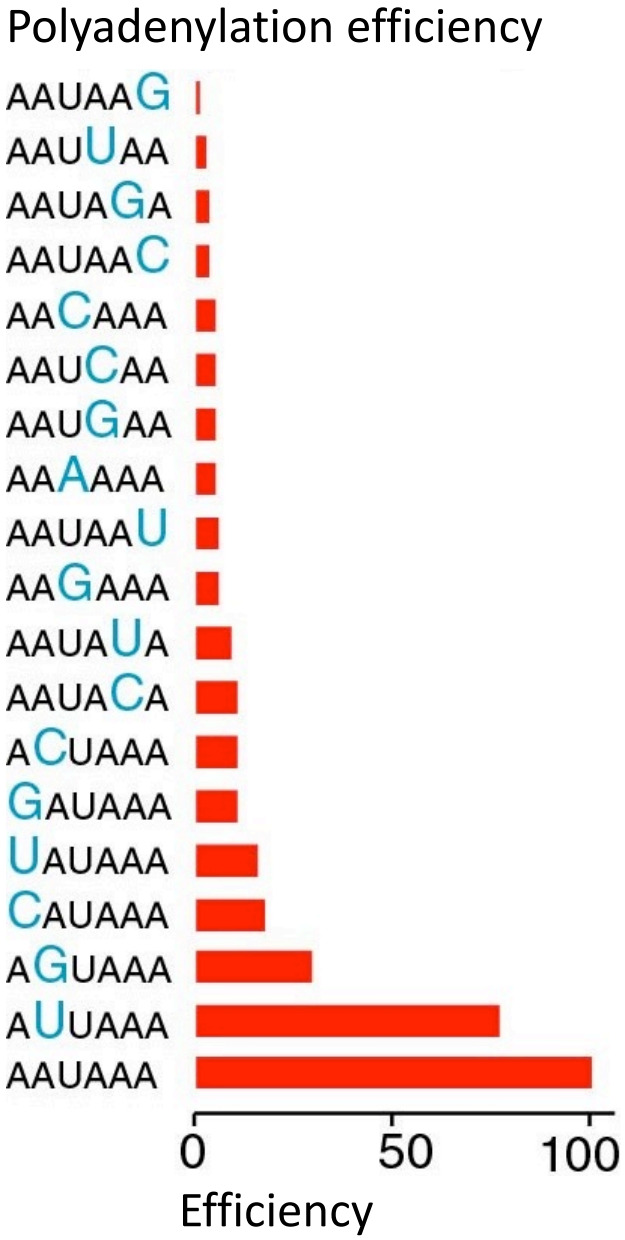


LOOKING FOR CONSENSUS SEQUENCES IMPORTANT FOR POLYADENYLATION

In vivo polyadelynation analysis



In vivo polyadelynation analysis



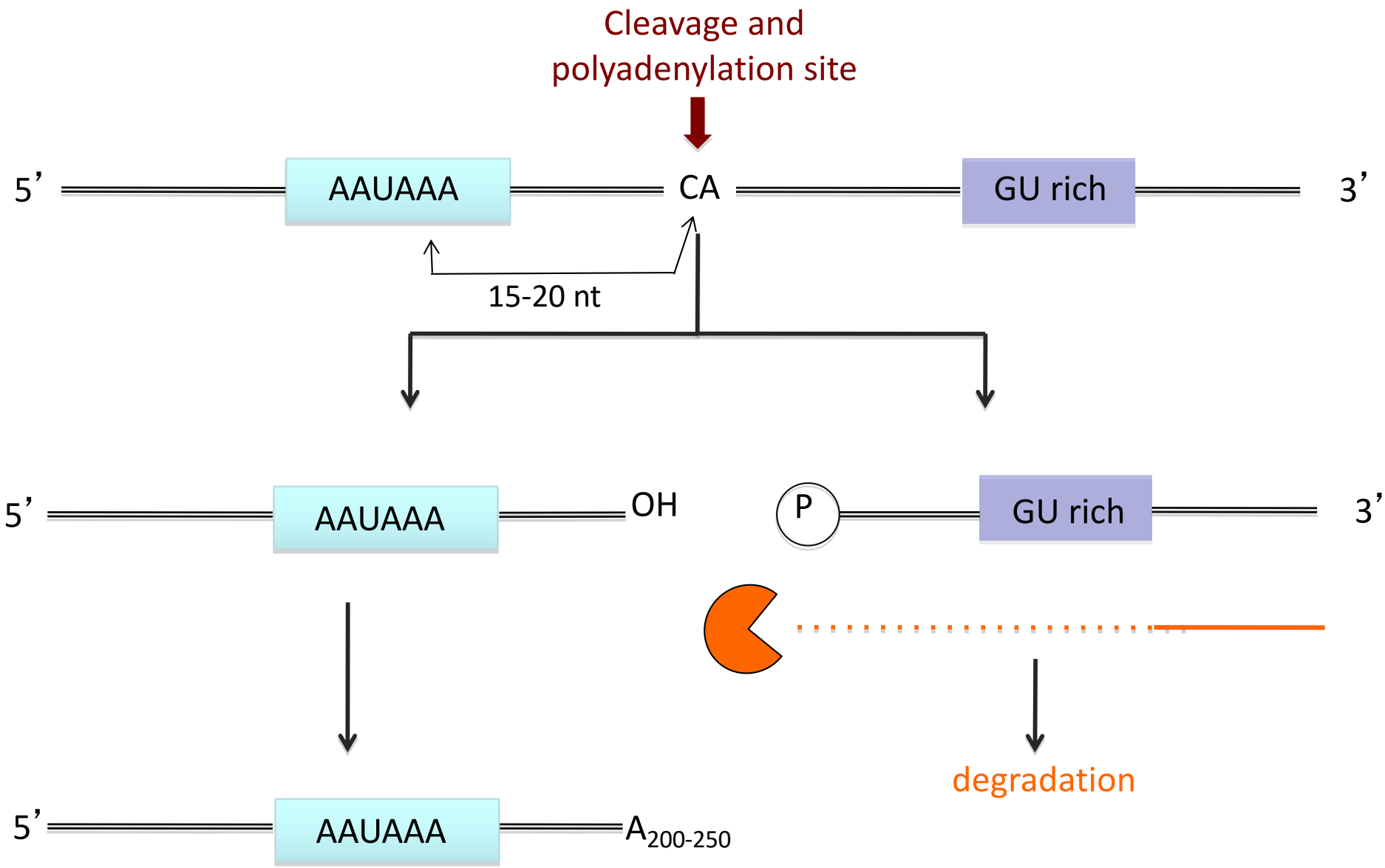
mammals

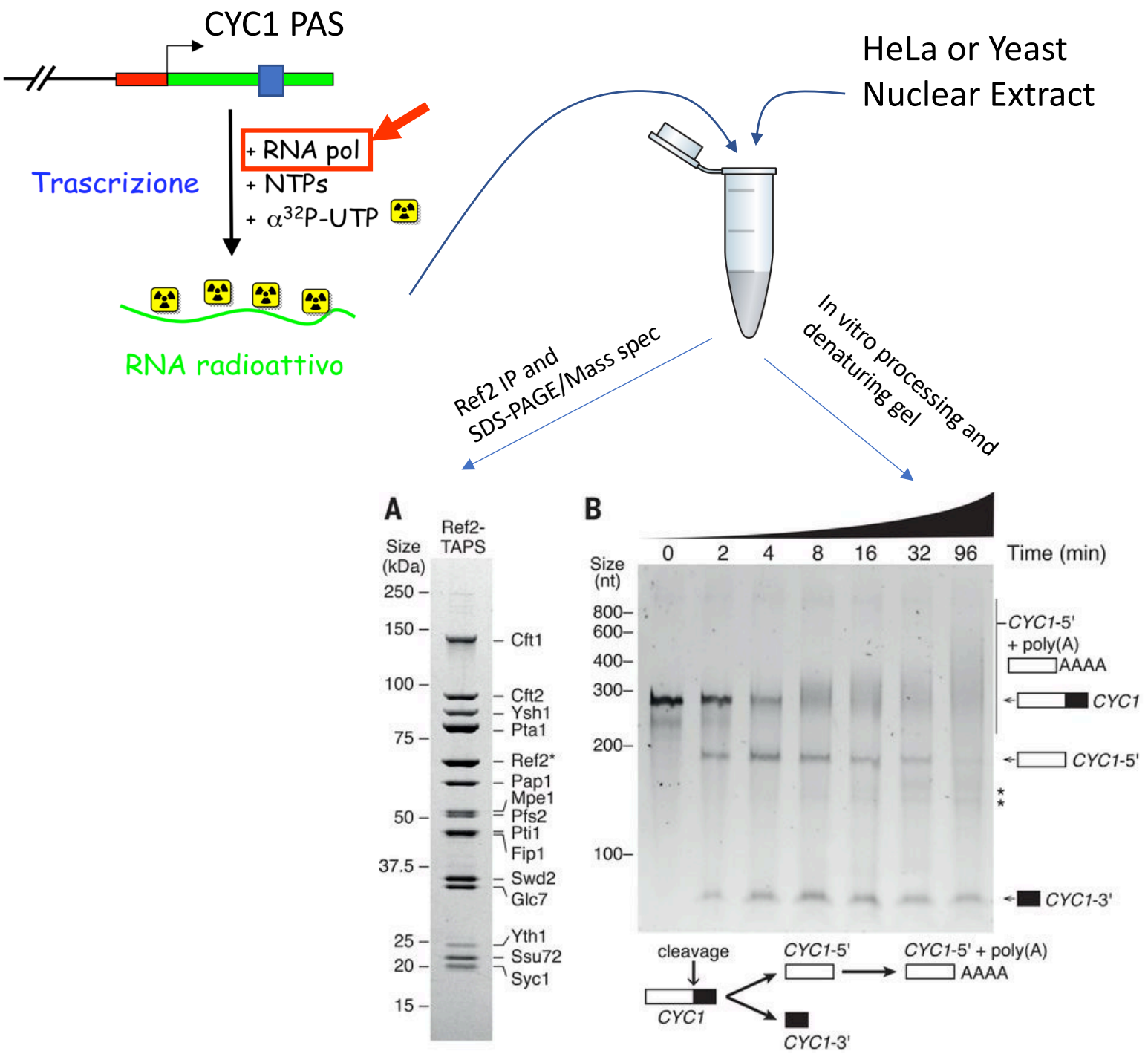
consensus

$A_{98}A_{86}U_{98}A_{98}A_{95}A_{96}$
U_{12}

- AAUAAA in mammals Located ~20-30 bp from the polyA site
  - Other hexamers less efficient but are used
- Mutagenesis and *in vivo* expression studies reveal also the importance of the GU-rich downstream of AAUAAA sequence.

3′-end formation in mammalian cells



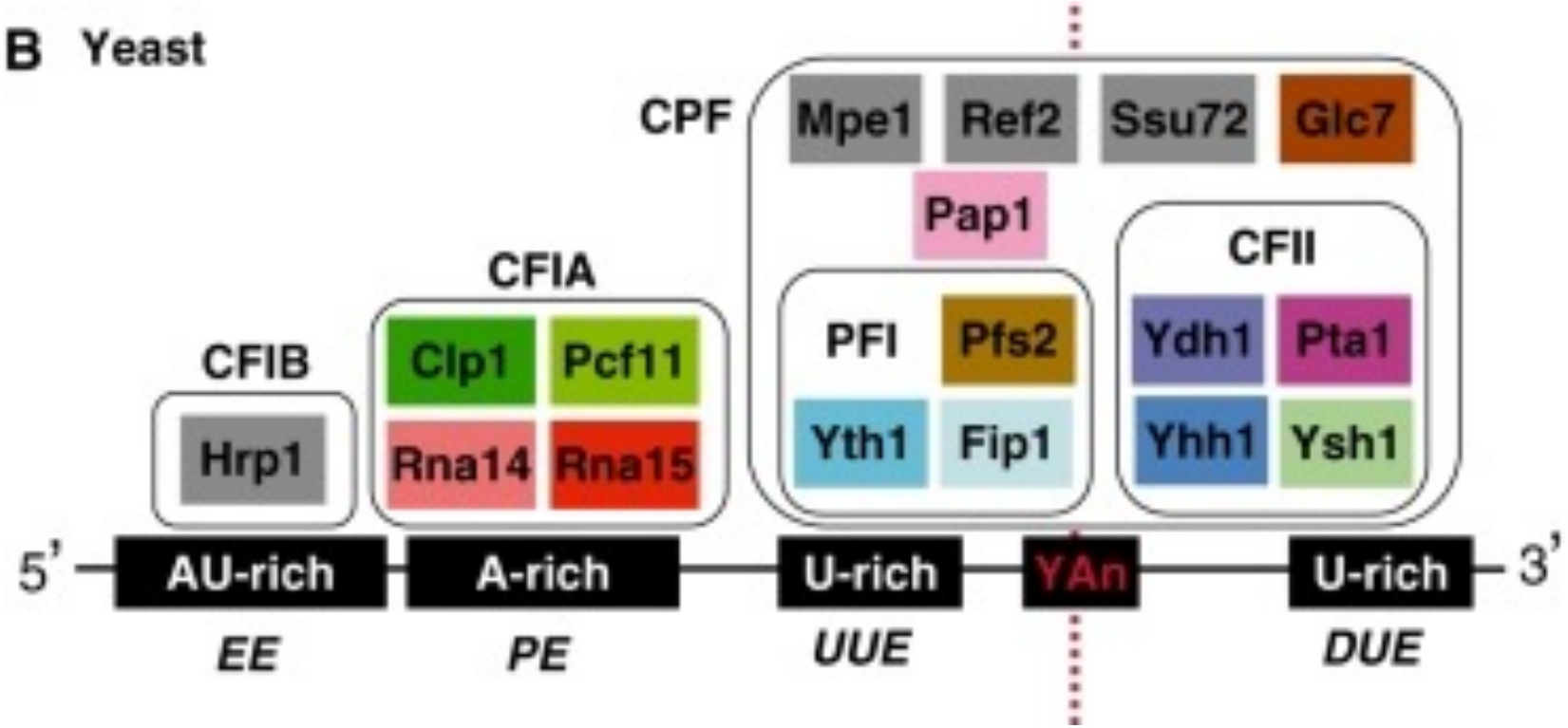

CYC1 PAS
Trascrizione
+ RNA pol
+ NTPs
+ α32P-UTP
RNA radioattivo
HeLa or Yeast
Nuclear Extract
Ref2 IP and
SDS-PAGE/Mass spec
In vitro processing and
denaturing gel
A
Size
(kDa)
Ref2-
TAPS
250
150
100
75
50
37.5
25
20
15
Cft1
Cft2
Ysh1
Pta1
Ref2*
Pap1
Mpe1
Pfs2
Pti1
Fip1
Swd2
Glc7
Yth1
Ssu72
Syc1
B
0
2
4
8
16
32
96
Time (min)
Size
(nt)
800
600
400
300
200
100
CYC1-5'
+ poly(A)
AAAA
CYC1
CYC1-5'
CYC1-3'
cleavage
CYC1-5'
CYC1-5' + poly(A)
AAAA
CYC1
CYC1-3'

PAS sequences are not conserved in evolution

Cis elements around the PAS in simpler species differ from those in mammals:

- downstream GU-rich elements are not present in nematode PASs.
- AAUAAA is prominent in *Schizosaccharomyces pombe*, while a general A-rich sequence is present in the same region in *Saccharomyces cerevisiae*. Yeast PASs do not have downstream GU-rich elements. However, UAUA elements are highly enriched in the upstream region.
- a typical plant PAS is more similar to that of yeast than of metazoan, with a loosely defined upstream AAUAAA element flanked by U-rich sequences.

Recognition of polyadenylation sites in yeast pre-mRNAs by cleavage and polyadenylation factor



In contrast to higher eukaryotes, *Saccharomyces cerevisiae* uses degenerate and complex signals to direct the reaction. Site-specific cleavage requires cleavage and polyadenylation factor IA (CF IA), cleavage and polyadenylation factor IB (CF IB) and cleavage factor II (CF II). Specific polyadenylation occurs when CF IA, CF IB, poly(A)-binding protein (Pab1p) and polyadenylation factor I (PF I) are present. A factor harbouring PF I and CF II activities (designated cleavage and polyadenylation factor, CPF) was isolated from yeast extracts by affinity purification

The polyadenylation machinery in metazoans

The polyadenylation machinery in metazoans is composed of ~20 core proteins, including four protein complexes and several single proteins:

- **cleavage and polyadenylation specificity factor (CPSF)**, which contains **CPSF160** (also known as CPSF1), **CPSF100** (also known as CPSF2), **CPSF73** (also known as CPSF3), **CPSF30** (also known as CPSF4), **FIP1** (factor interacting with PAP) and **WDR33**

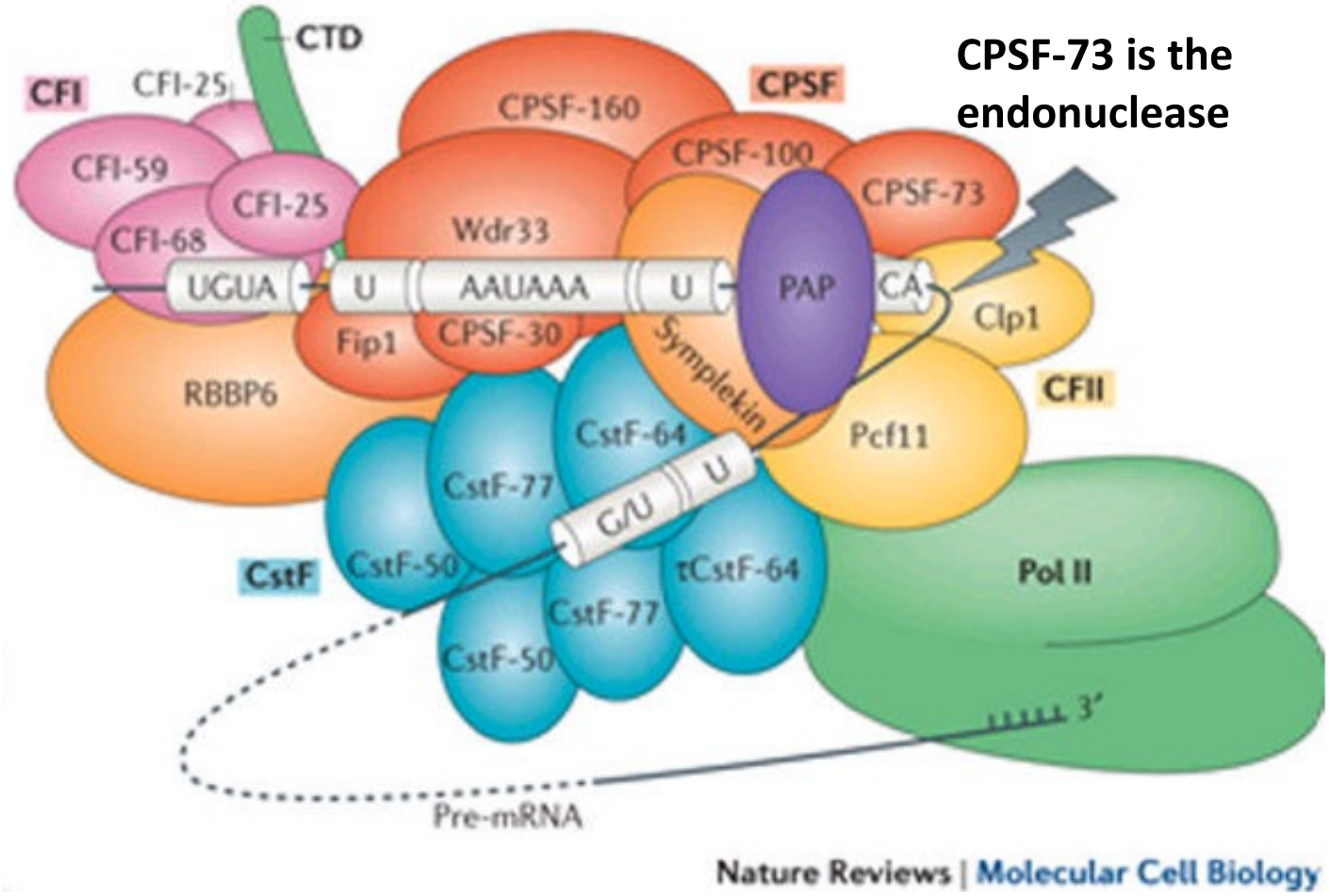
- **cleavage stimulation factor (CSTF)**, which contains **CSTF 77**, **CSTF50** (also known as CSTF1) and **CSTF64**

- **cleavage factor I (mCFI)**, which contains **CFI25** and either **CFI68** or **CFI59**;

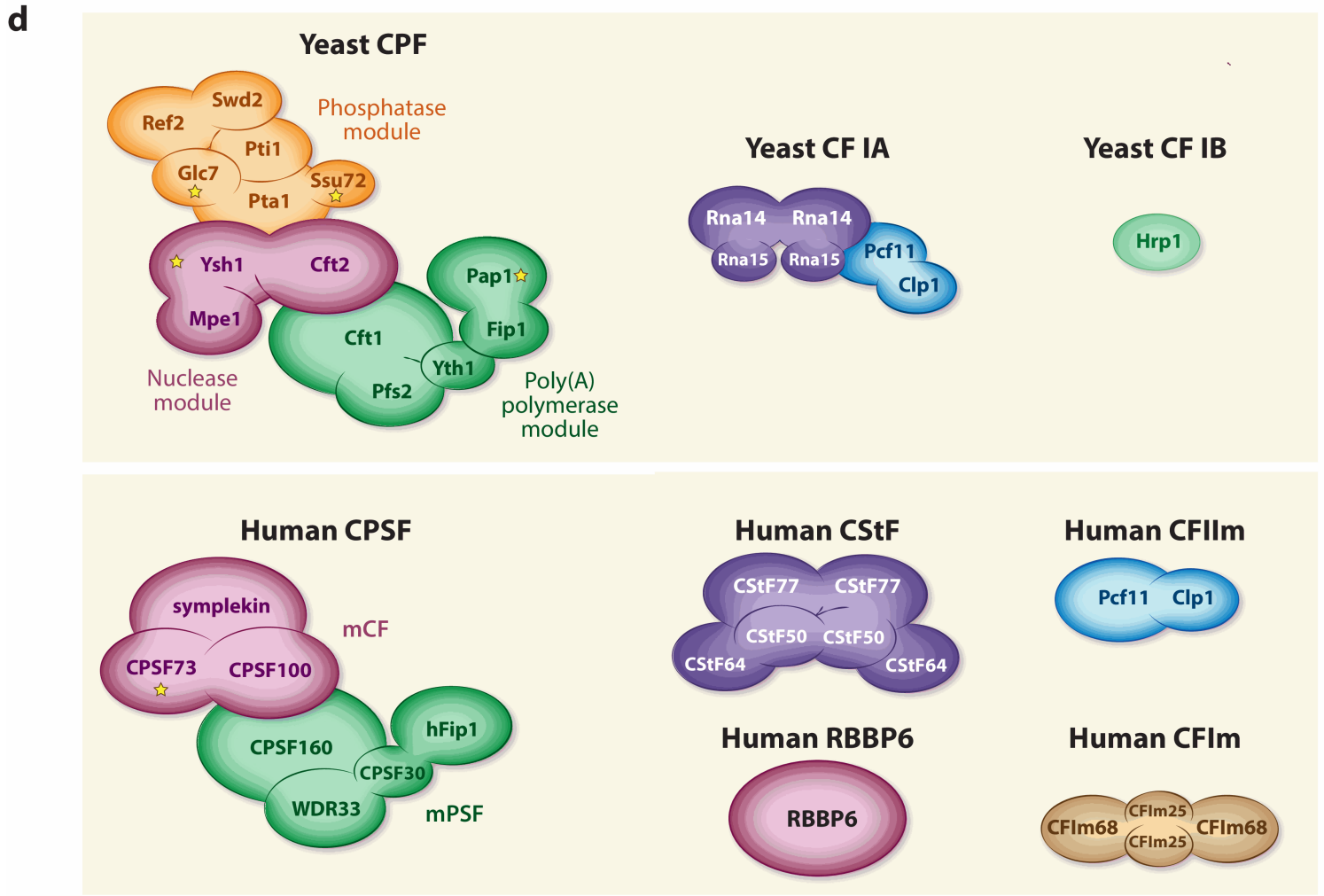
- **cleavage factor II (mCFII)**, which contains **PCF11** and **CLP1**

- Single proteins include **symplexin**, **poly(A) polymerase (PAP)**, **retinoblastoma-binding protein 6 (RBBP6)**.

Specific complexes are involved in mRNA 3'end-processing



CPSF-73 is the endonuclease



Boreikaitė V, Passmore LA. 2023
Annu. Rev. Biochem. 92:199–225

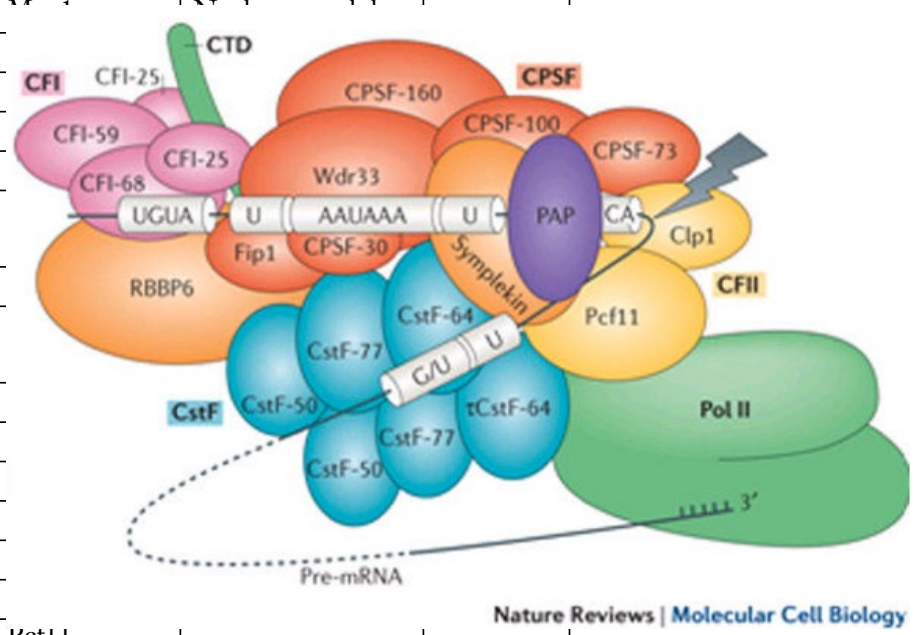
Despite considerable divergence between yeast and mammals in the core RNA sequences that constitute the PAS, nearly all mammalian polyadenylation factors have homologues in yeast, with the exception of the CFI proteins and CSTF50. Moreover, the yeast polyadenylation factor Hrp1p, which interacts with UA-rich elements, is missing from metazoans. Human PAP is not constitutively associated with the mPSF complex, but interestingly, Pap1 is a stable subunit in yeast.

Table 1 Canonical pre-mRNA 3′-end-processing factors in humans and budding yeast, their functions, and the multi-subunit protein complexes to which they belong. Alternative names are in brackets

Human complex		Human protein	Function	Yeast protein	Yeast complex		
CPSF	mPSF	CPSF160 (CPSF1)	Scaffold	Cft1	Poly(A) polymerase module	Core CPF	CPF
		WDR33	Scaffold, RNA binding	Pfs2			
		CPSF30 (CPSF4)	RNA binding	Yth1			
		hFip1	PAP/Pap1 recruitment	Fip1			
	N/A	PAP	Poly(A) polymerase	Pap1			
		RBBP6	Endonuclease activation	Mpe1	Nuclease module		
	mCF	CPSF100 (CPSF2)	Pseudonuclease	Cft2			
		CPSF73 (CPSF3)	Endonuclease	Ysh1			
		Symplekin	Scaffold	Pta1	Phosphatase module	APT	
N/A	N/A	SSU72	Protein phosphatase	Ssu72			
	Phosphatase complex	WDR82	Transcription termination	Swd2			
		PP1	Protein phosphatase	Glc7			
		PNUTS	Scaffold, PP1/Glc7 activator	Ref2			
		N/A	Scaffold	Pti1			
		Tox4	DNA binding	N/A	N/A		N/A
	N/A	N/A	Ysh1 antagonizing	Syc1			
	CStF	CStF50 (CSTF1)	Complex stabilizing	N/A		N/A	
		CStF64 (CSTF2)	RNA binding	Rna15	CF IA		
		CStF77 (CSTF3)	CPSF160/Cft1 binding	Rna14			
	CFIIm	Pcf11	Pol II CTD binding	Pcf11			
		Clp1	Polynucleotide kinase	Clp1			
		N/A	Cleavage fidelity	Hrp1	CF IB		
	CFIm	CFIm25 (CPSF5)	RNA binding	N/A	N/A		
		CFIm68 (CPSF6)	hFip1 binding	N/A			
		CFIm59 (CPSF7)	hFip1 binding	N/A			
	N/A	PABPN1	Poly(A) tail binding	N/A			
		N/A	Poly(A) tail binding	Pab1			
		ZC3H14	Poly(A) tail binding	Nab2			

Table 1 Canonical pre-mRNA 3′-end-processing factors in humans and budding yeast, their functions, and the multi-subunit protein complexes to which they belong. Alternative names are in brackets

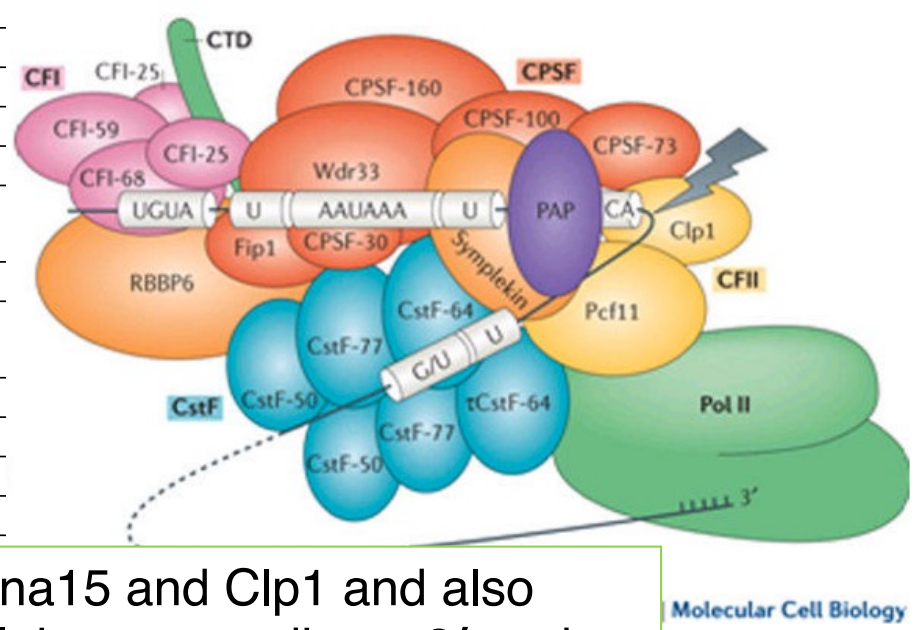
Human complex		Human protein	Function	Yeast protein	Yeast complex		
CPSF	mPSF	CPSF160 (CPSF1)	Scaffold	Cft1	Poly(A) polymerase module	Core CPF	CPF
		WDR33	Scaffold, RNA binding	Pfs2			
		CPSF30 (CPSF4)	RNA binding	Yth1			
		hFip1	PAP/Pap1 recruitment	Fip1			
	N/A	PAP	Poly(A) polymerase	Pap1			
		RBBP6	Endonuclease activation				
	mCF	CPSF100 (CPSF2)	Pseudonuclease				
		CPSF73 (CPSF3)	Endonuclease				
		Symplekin	Scaffold				
N/A	N/A	SSU72	Protein phosphatase				
	Phosphatase complex	WDR82	Transcription termination				
		PP1	Protein phosphatase				
		PNUTS	Scaffold, PP1/Glc7 activator				
		N/A	Scaffold				
		Tox4	DNA binding				
	N/A	N/A	Ysh1 antagonizing				
	CStF	CStF50 (CSTF1)	Complex stabilizing				
		CStF64 (CSTF2)	RNA binding				
		CStF77 (CSTF3)	CPSF160/Cft1 binding				
	CFIIm	Pcf11	Pol II CTD binding	Pcf11			
		Clp1	Polynucleotide kinase	Clp1			
		N/A	Cleavage fidelity	Hrp1	CF IB		
	CFIm	CFIm25 (CPSF5)	RNA binding	N/A	N/A		
		CFIm68 (CPSF6)	hFip1 binding	N/A			
		CFIm59 (CPSF7)	hFip1 binding	N/A			
	N/A	PABPN1	Poly(A) tail binding	N/A			
		N/A	Poly(A) tail binding	Pab1			
		ZC3H14	Poly(A) tail binding	Nab2			



Factors important for RNA binding and complex interaction for integration of 3’end processing with other maturation processes

Table 1 Canonical pre-mRNA 3′-end-processing factors in humans and budding yeast, their functions, and the multi-subunit protein complexes to which they belong. Alternative names are in brackets

Human complex		Human protein	Function	Yeast protein	Yeast complex		
CPSF	mPSF	CPSF160 (CPSF1)	Scaffold	Cft1	Poly(A) polymerase module	Core CPF	CPF
		WDR33	Scaffold, RNA binding	Pfs2			
		CPSF30 (CPSF4)	RNA binding	Yth1			
		hFip1	PAP/Pap1 recruitment	Fip1			
	N/A	PAP	Poly(A) polymerase	Pap1			
		RBBP6	Endonuclease activation				
	mCF	CPSF100 (CPSF2)	Pseudonuclease				
		CPSF73 (CPSF3)	Endonuclease				
		Symplekin	Scaffold				
N/A	N/A	SSU72	Protein phosphatase				
	Phosphatase complex	WDR82	Transcription termination				
		PP1	Protein phosphatase				
		PNUTS	Scaffold, PP1/Glc7 activator				
		N/A	Scaffold				
		Tox4	DNA binding				
	N/A	N/A	Ysh1 antagonizing				
	CStF	CStF50 (CSTF1)	Complex stabilizing				
		ZC3H14	Poly(A) tail binding	Nab2			



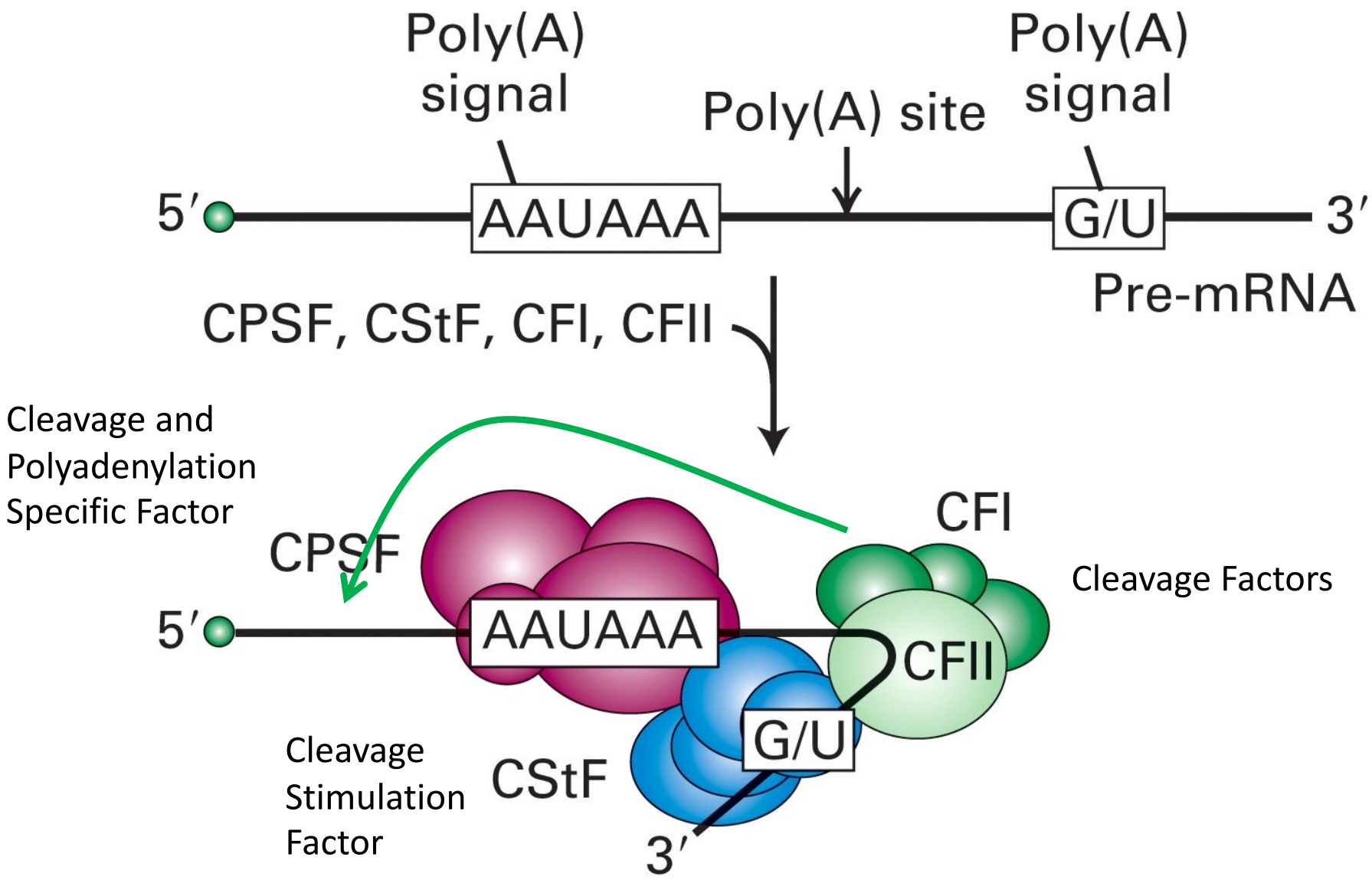
Molecular Cell Biology

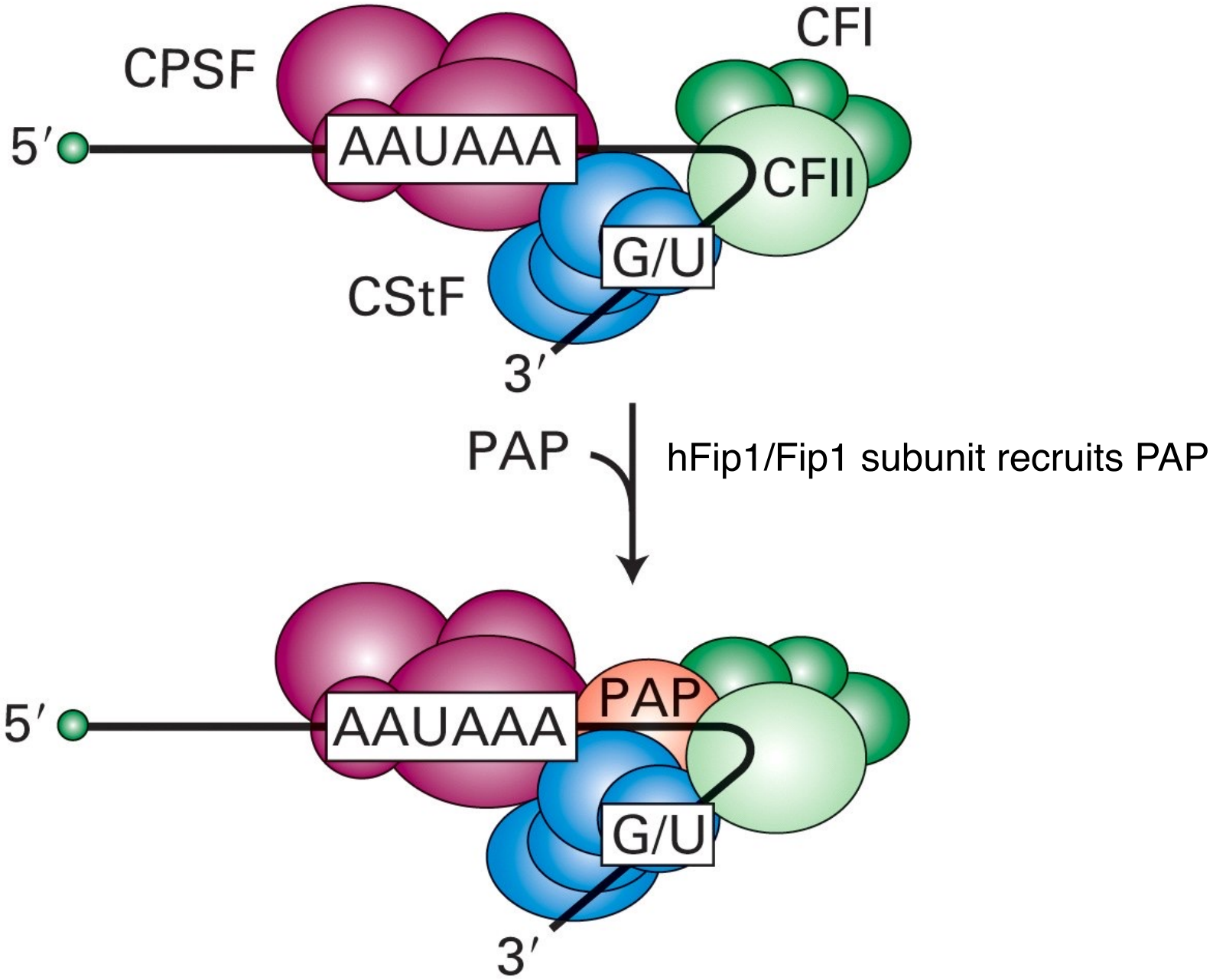
Pcf11 physically bridges CStF/Rna14–Rna15 and Clp1 and also interacts with the CTD of Pol II, likely helping to coordinate 3′-end processing with transcription

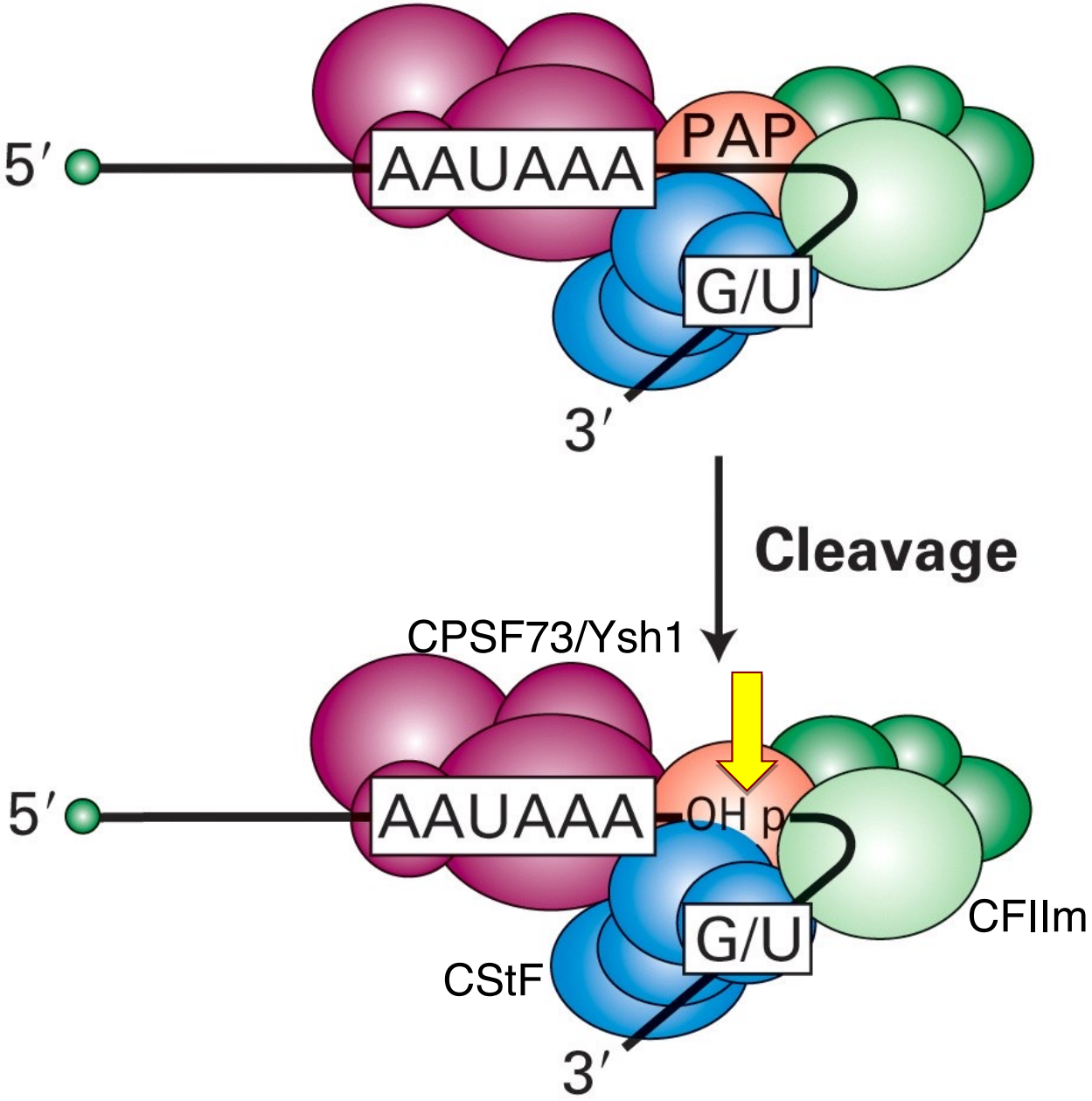
Human RBBP6 is not required for activation of cleavage but may interact with transcription factors and splicing regulators

Factors important for RNA binding and complex interaction for integration of 3’end processing with other maturation processes

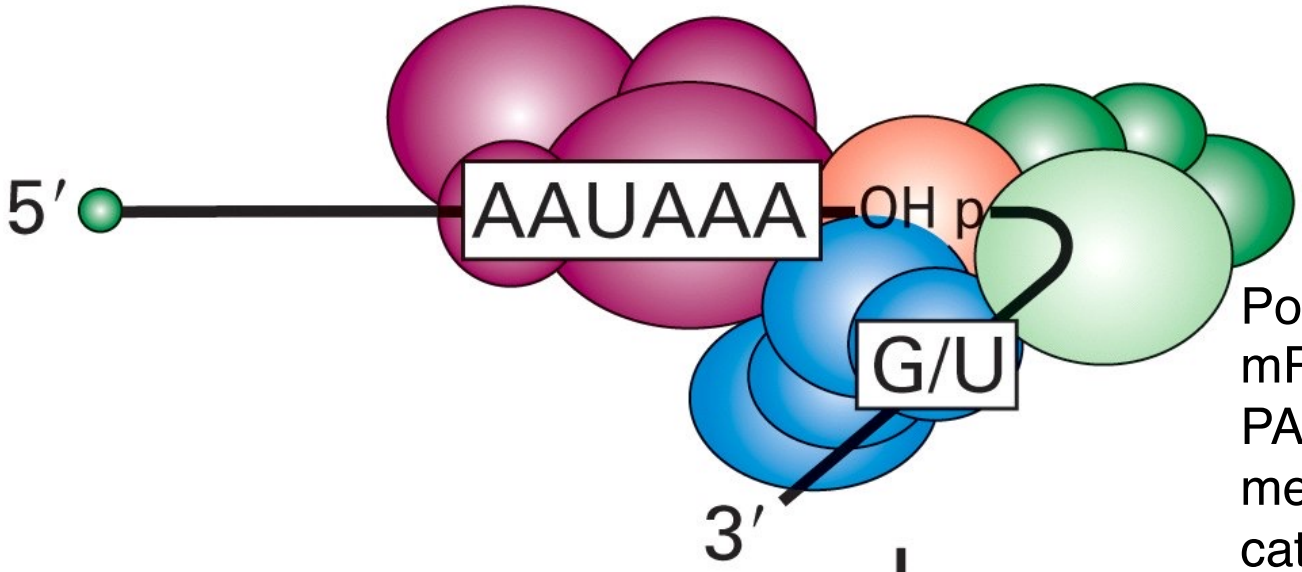
3′-end formation: RNA Processing




CPSF
CFI
5′
AAUAAA
CFII
G/U
CStF
3′
PAP
hFip1/Fip1 subunit recruits PAP
5′
AAUAAA
PAP
G/U
3′



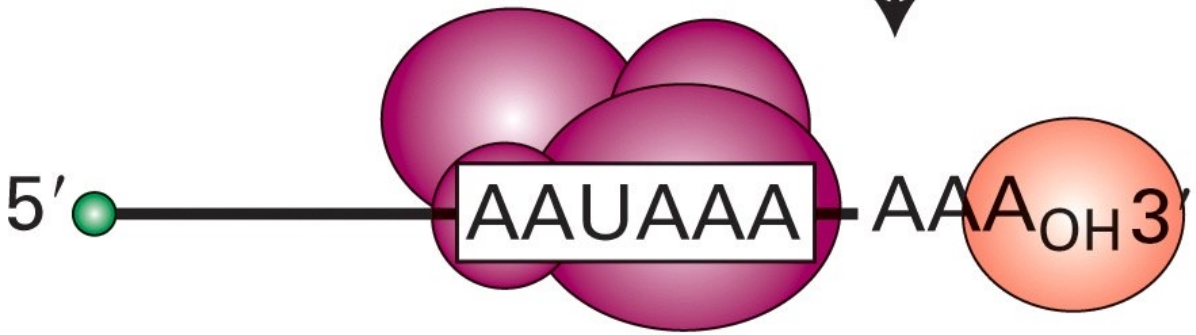
CPSF73/Ysh1adopts an inactive closed conformation and has only weak and nonspecific nuclease activity. CF IA/CStF and CFIIm are required to activate the 3′-end-processing endonuclease



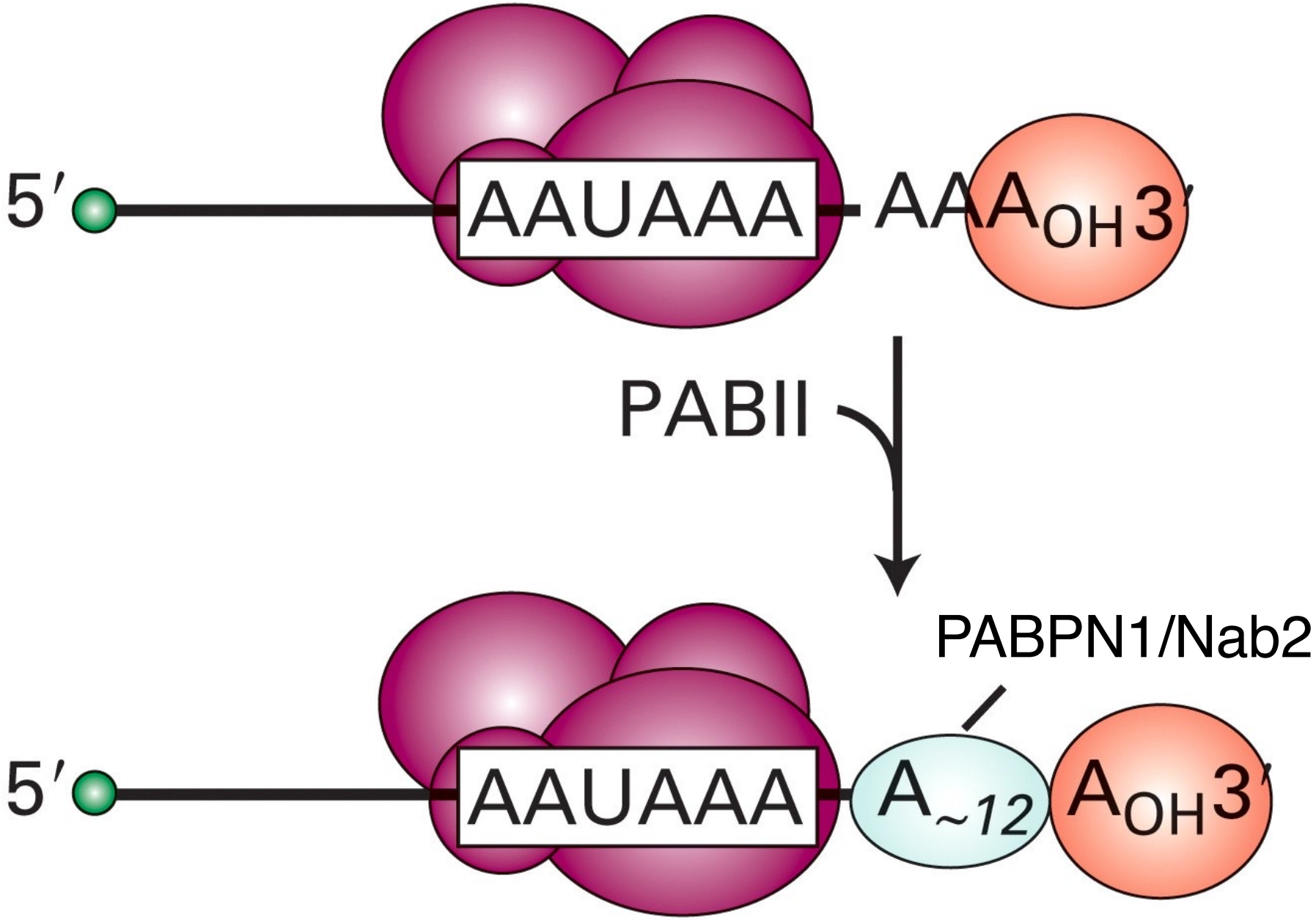
Polyadenylation of cleaved pre-mRNAs is catalyzed by PAP/Pap1 using a polymerization mechanism dependent on two catalytic magnesium ions

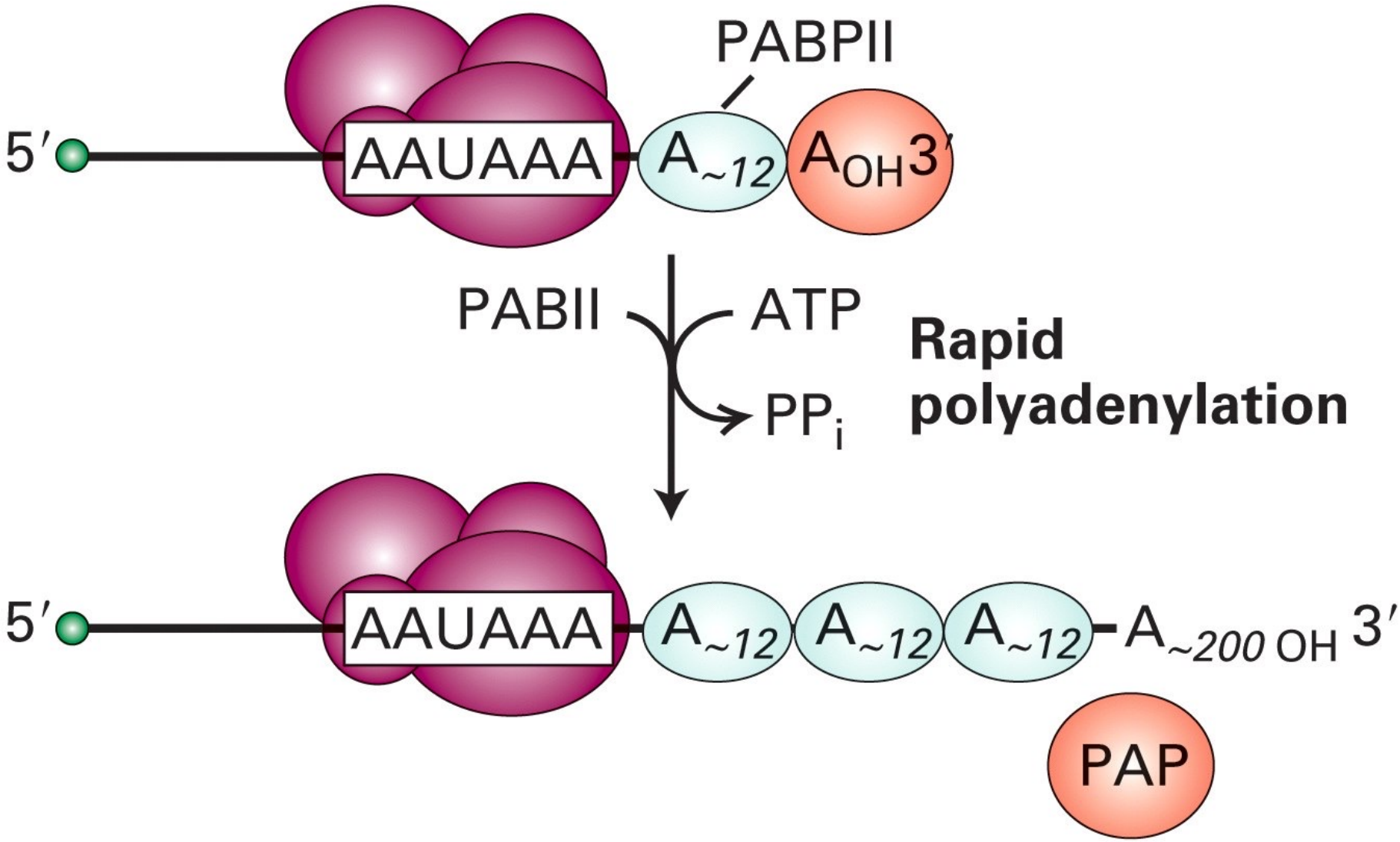
Slow polyadenylation

likely because of its low affinity for RNA

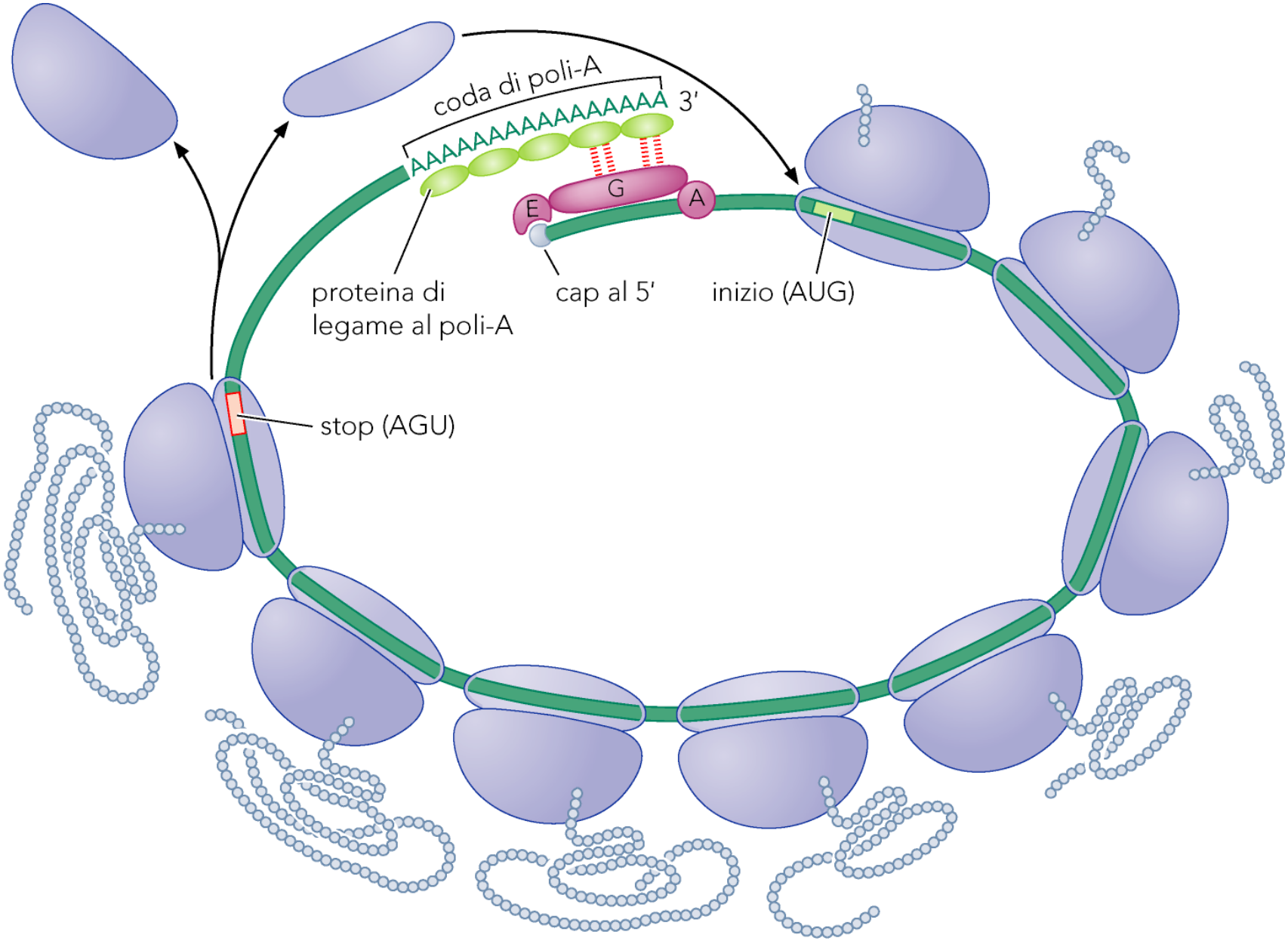


degradation


5′
AAUAAA
AAA$_{OH}$3′
PABII
PABPN1/Nab2
5′
AAUAAA
A$_{\sim 12}$
A$_{OH}$3′

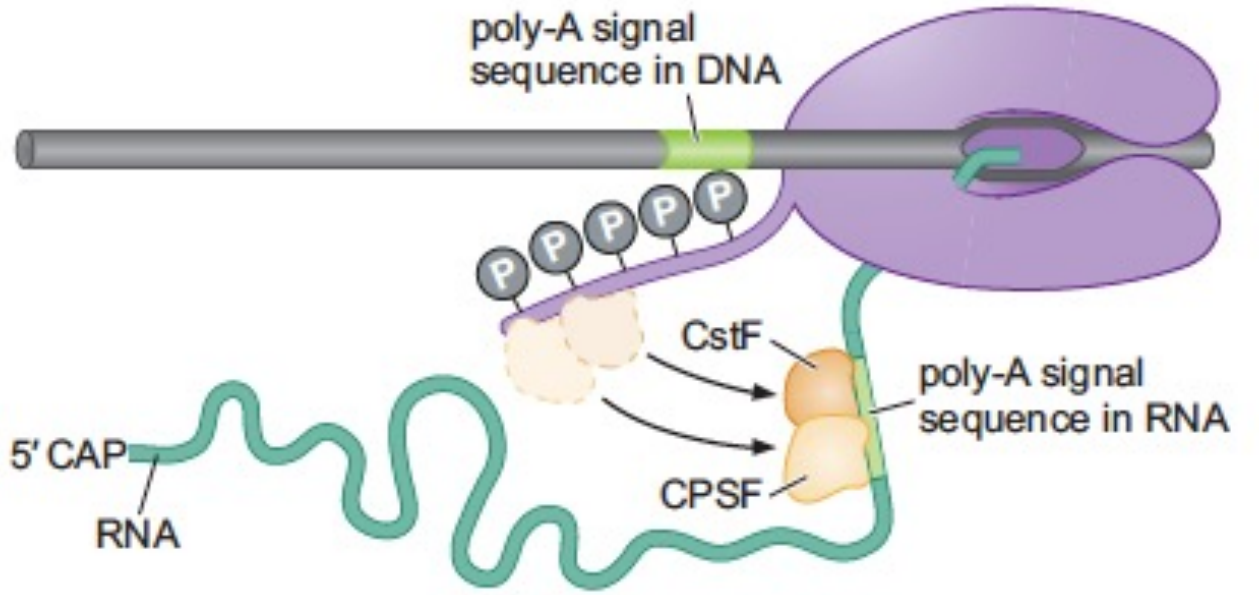

PABPII
5′
AAUAAA
A~12
AOH3′
PABII
ATP
PPi
Rapid
polyadenylation
5′
AAUAAA
A~12
A~12
A~12
A~200 OH 3′
PAP

CAP and polyA tail promote mRNA translation

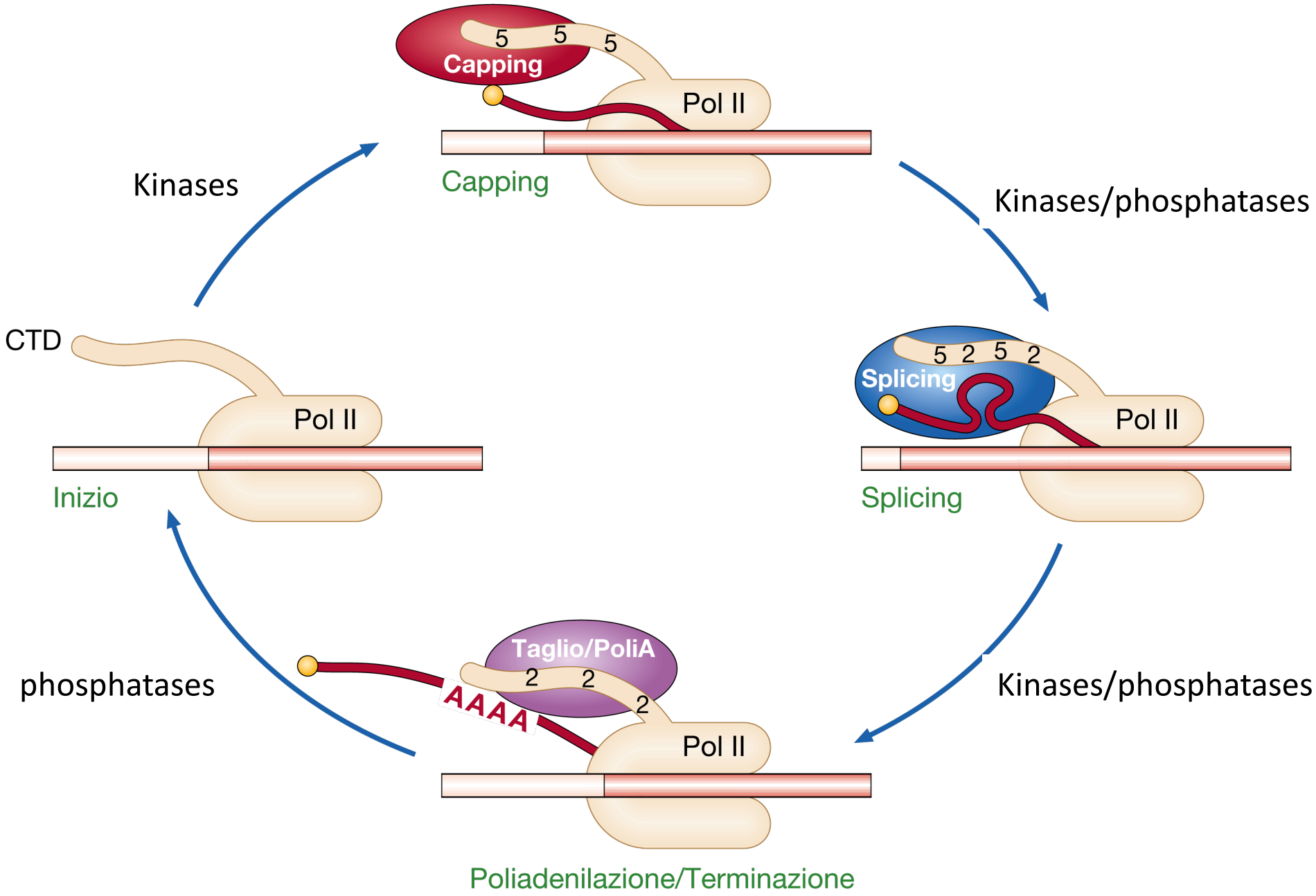


3′-end formation occurs during transcritpion

- TFIID associates with CPSF in the preinitiation complex (PIC)
- After transcription initiation CPSF dissociates from the PIC and associates with the CTD of elongating RNA Pol II together with CstF



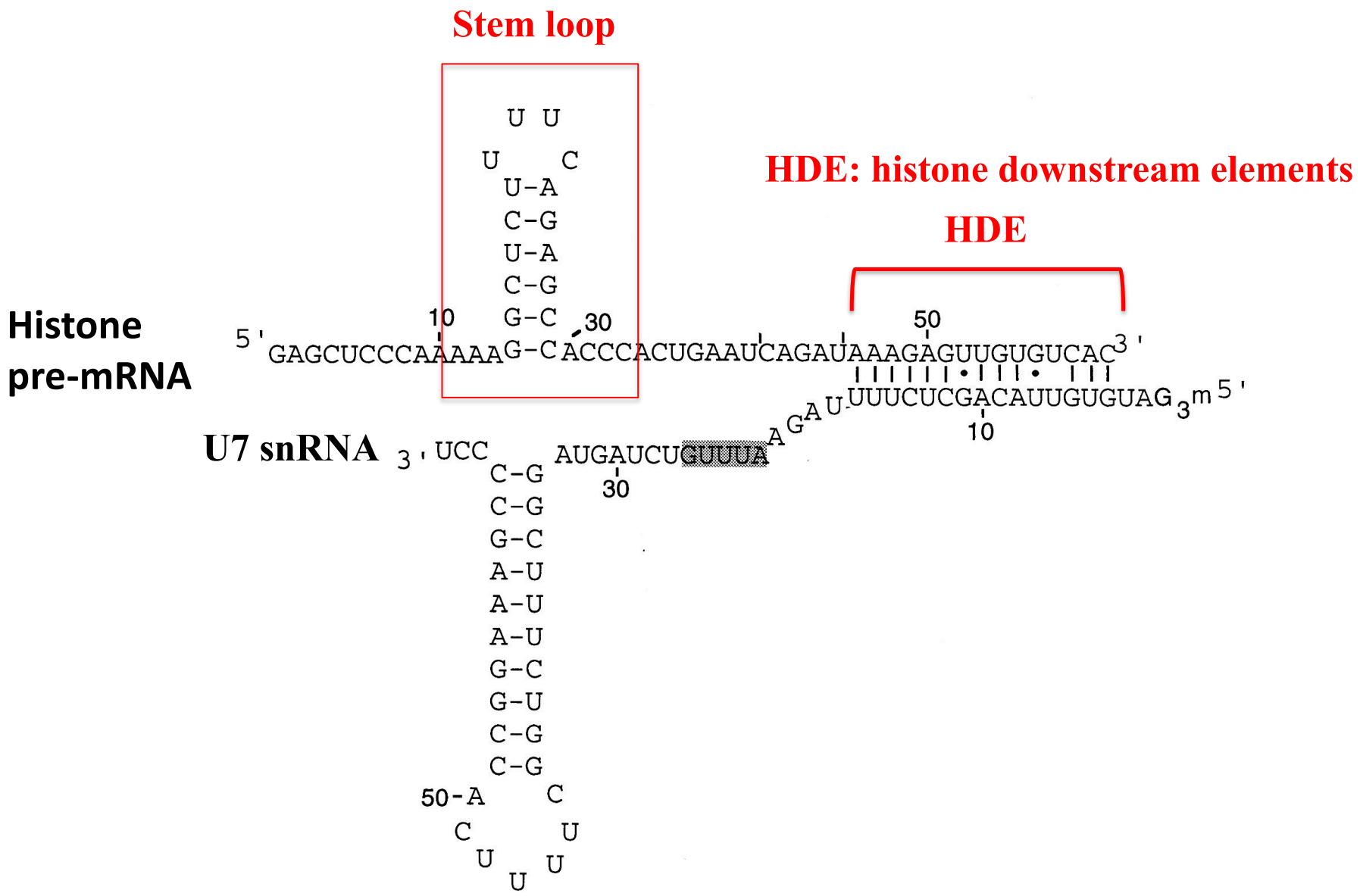
capping, splicing and polyadenylation processes are tightly coupled with to transcription



Characteristics of Histone Messenger RNA

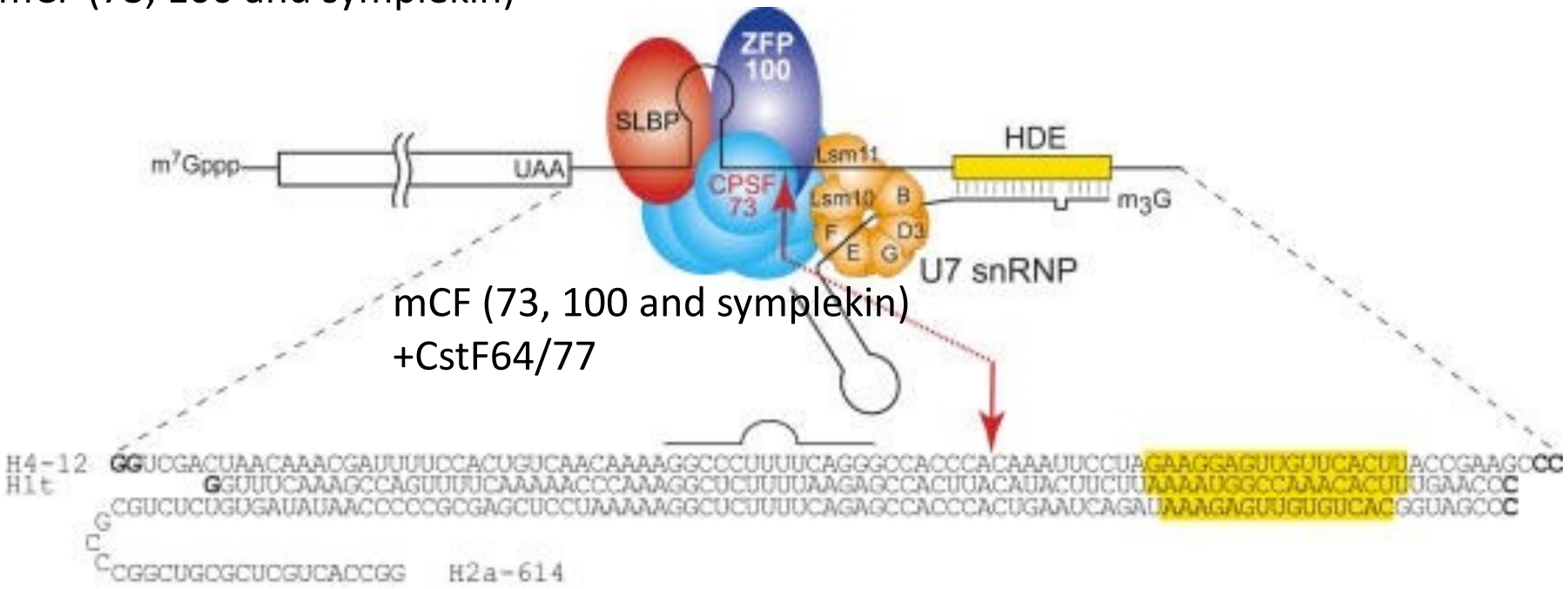
- Lacks introns: No splicing
- Lacks a polyA tail: cleavage not followed by polyadenylation
- Export to the cytoplasm is independent of the polyA tail
- Degradation occurs via uridylation
- Regulation of expression is dependent on the cell cycle

The maturation of histone mRNA requires cis elements and trans factors, including U7 snRNA



Histone mRNA processing is polyadenylation independent

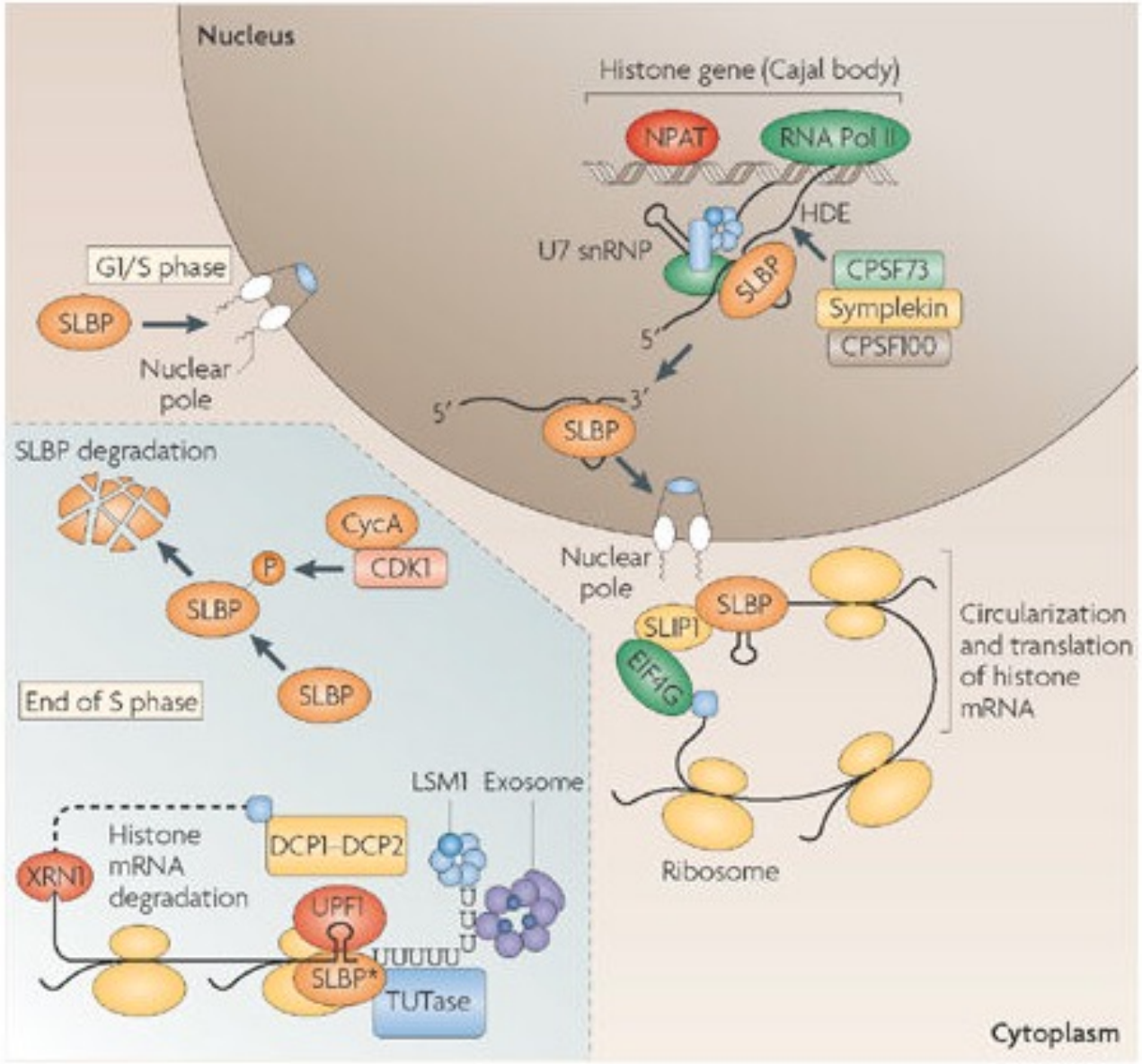
Histone cleavage complex (HCC)
mCF (73, 100 and symplekin)



The histone pre-mRNA 3′-end-processing complex contains a seven subunit Sm ring bound to U7 snRNA; subunits Lsm10, Lsm11, ZFP100, and SLBP; the HCCsubcomplex that is composed of the same subunits as mCF, including endonuclease CPSF73

The HDE duplex is~15 nucleotides long, which, along with the recognition of the stem loop structure by SLBP, ensures highly specific recognition of histone pre-mRNAs.
The U7 snRNP and stem–loop binding protein (SLBP), and the cleavage complex are responsible for the cleavage of the pre-mRNA from the DNA template, forming the mature histone mRNA

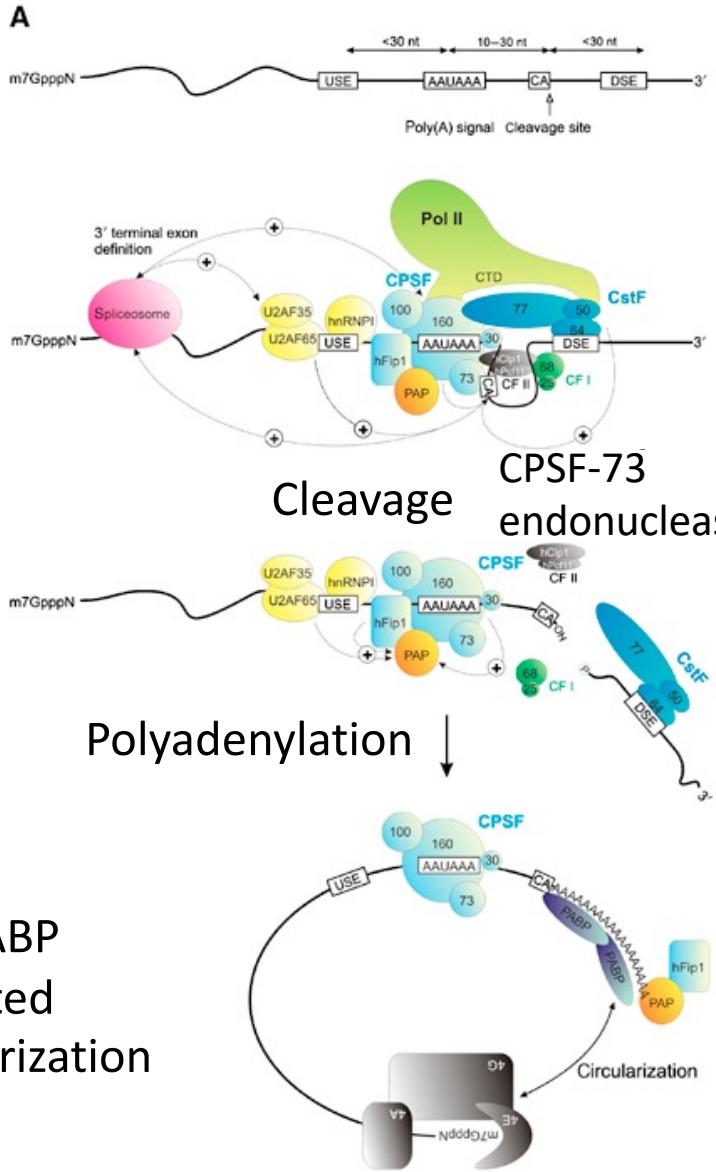
Histone mRNAs: life without a poly(A) tail



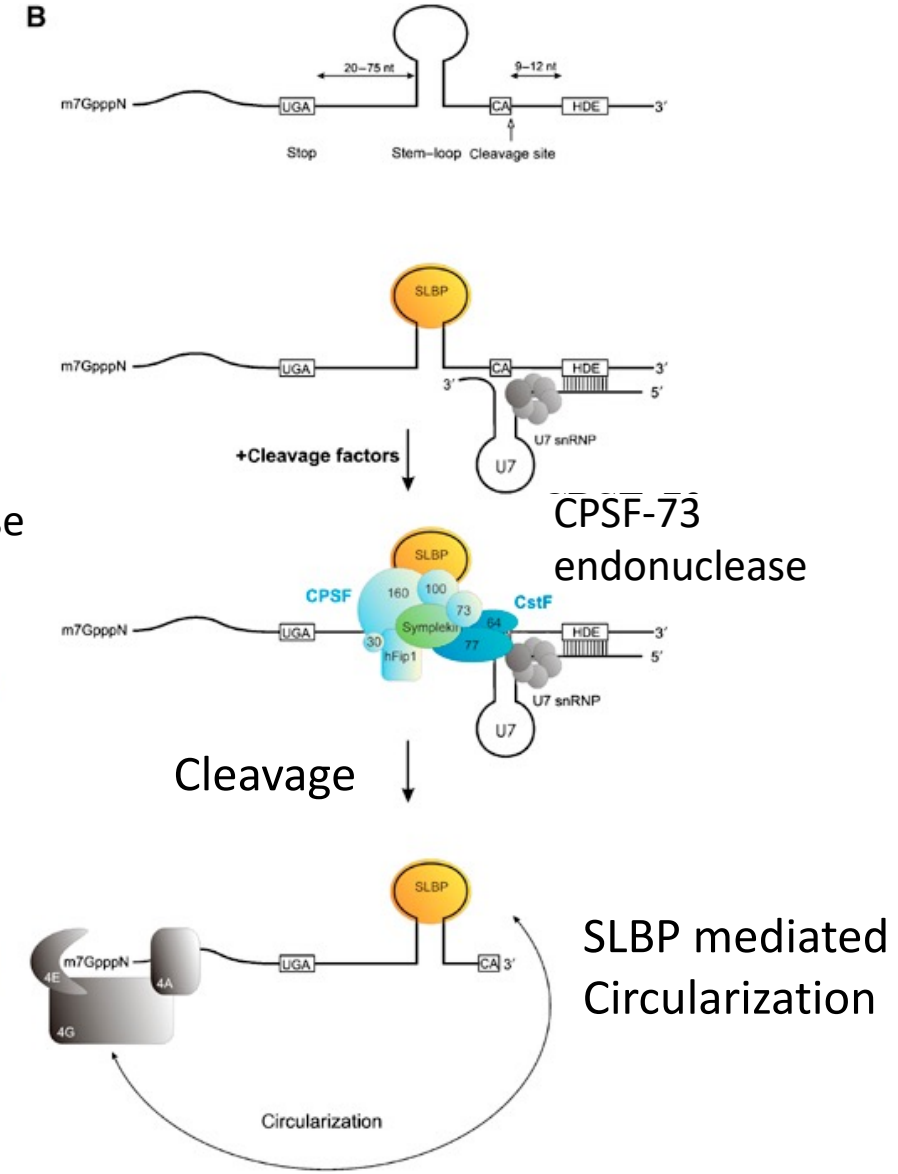
SLBP remains bound to the histone mRNA as it goes to the cytoplasm, where histone mRNA is circularized through a complex of proteins mediating translation of histone mRNA. At the end of S phase, a short U tail is added to histone mRNA in the cytoplasm. The LSM1–7 ring binds the oligo(U) to cooperate in the recruitment of the decapping complex and the exosome to degrade the mRNA. In addition, cyclin A (CycA)–CDK1 (cyclin-dependent kinase 1) phosphorylates SLBP to trigger its degradation at the end of S phase, preventing further histone mRNA synthesis.

Processing and translation of mRNAs

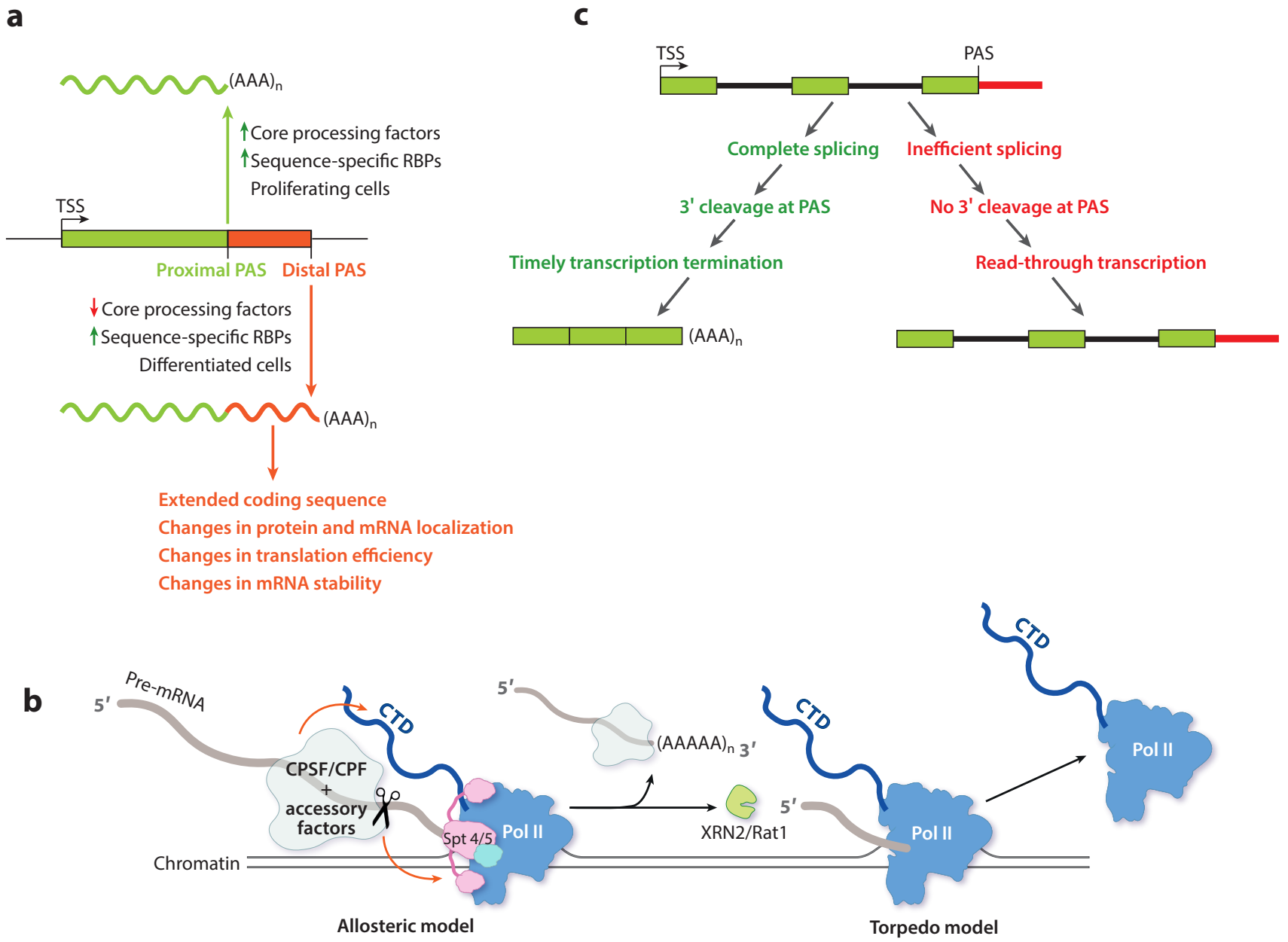
Poly(A)$^+$ mRNAs



Histone mRNAs



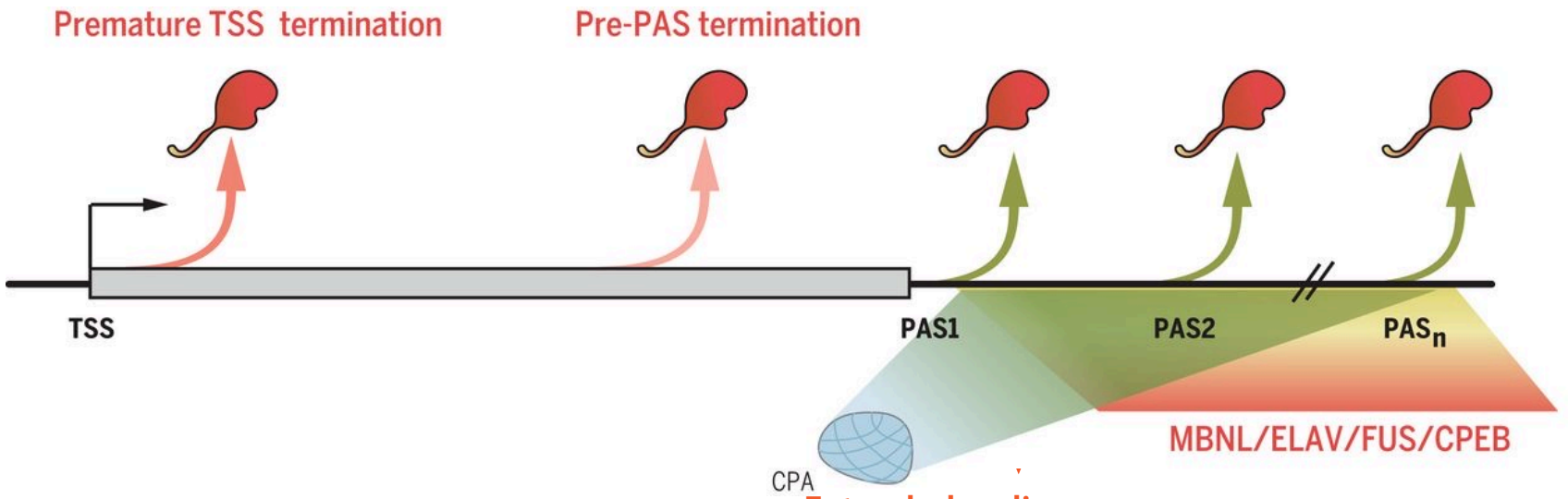
Eukaryotic pre-mRNA 3′-end processing is tightly regulated and coordinated with splicing and transcription termination.



RNA Pol II alternative 3′-end formation.

Termination at gene 3′-ends is also subject to intense regulation. Many mRNAs possess variable lengths of 3′-untranslated sequence defined by the selective usage of different PASs. Because mRNA 3′UTRs define mRNA cytoplasmic functions, the use of alternative poly(A) sites (APA) can constitute a key regulatory process in gene expression.

Polymerase speed and in trans elements can affect APA choice



Extended coding sequence

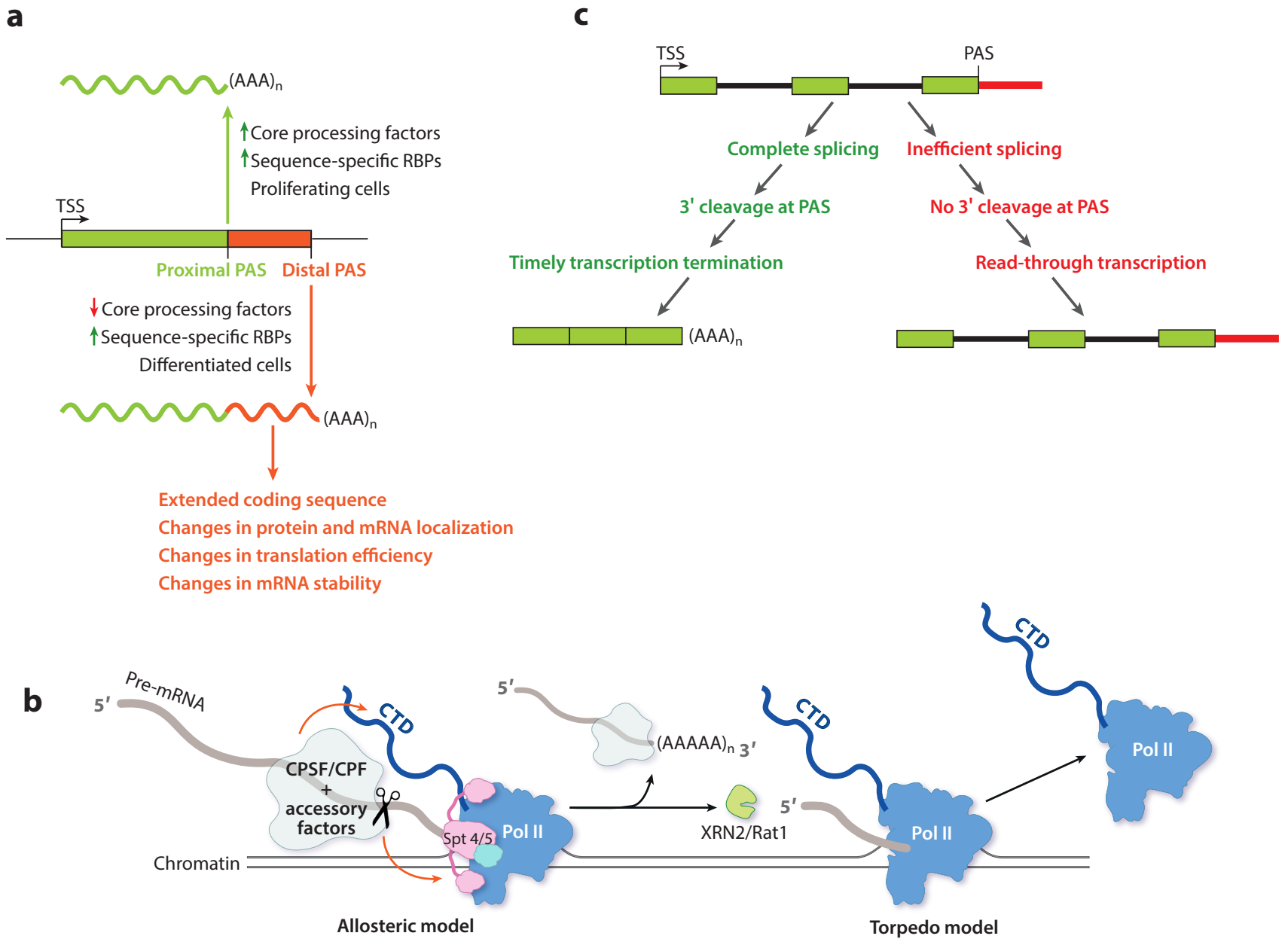
Changes in protein and mRNA localization

Changes in translation efficiency

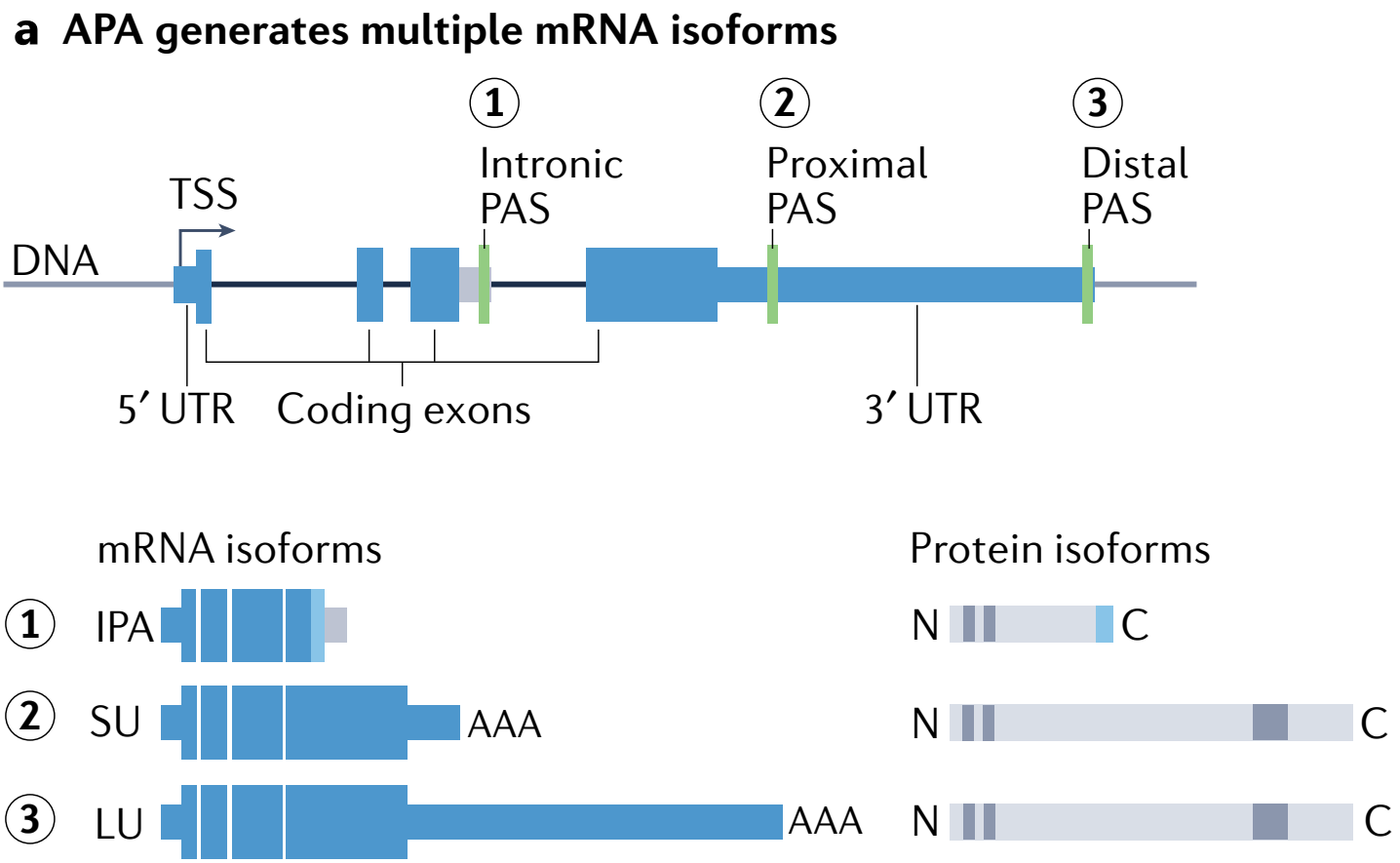
Changes in mRNA stability

Science AAAS

Eukaryotic pre-mRNA 3′-end processing is tightly regulated and coordinated with splicing and transcription termination.

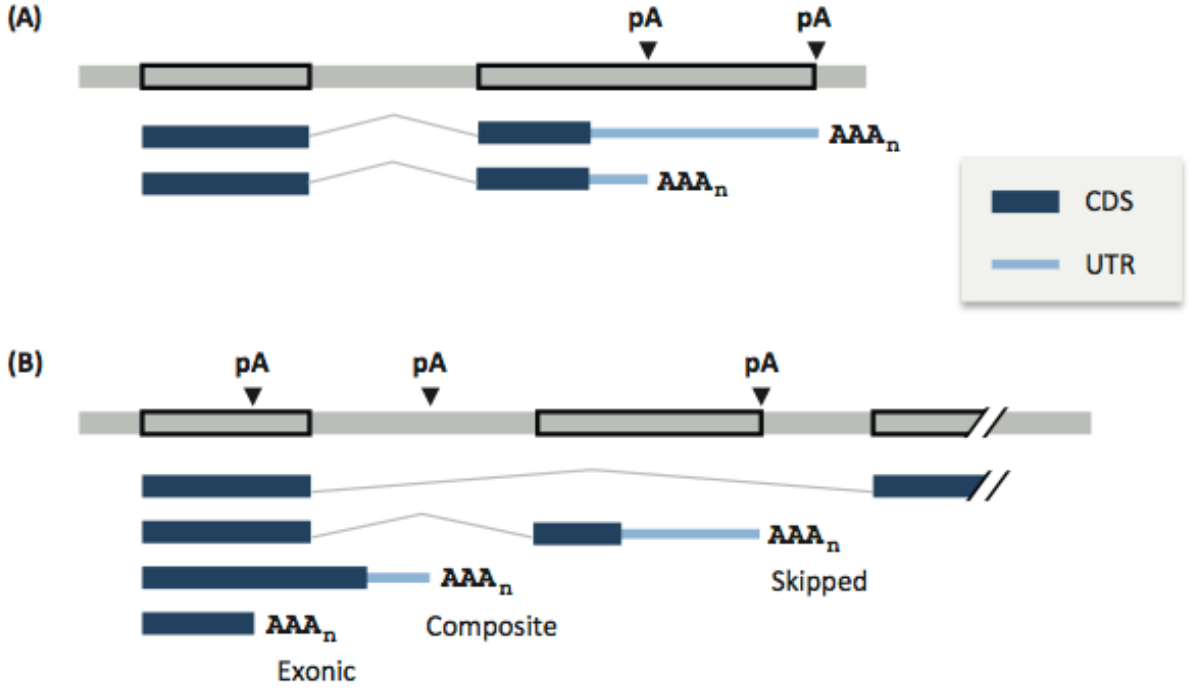


Alternative polyadenylation



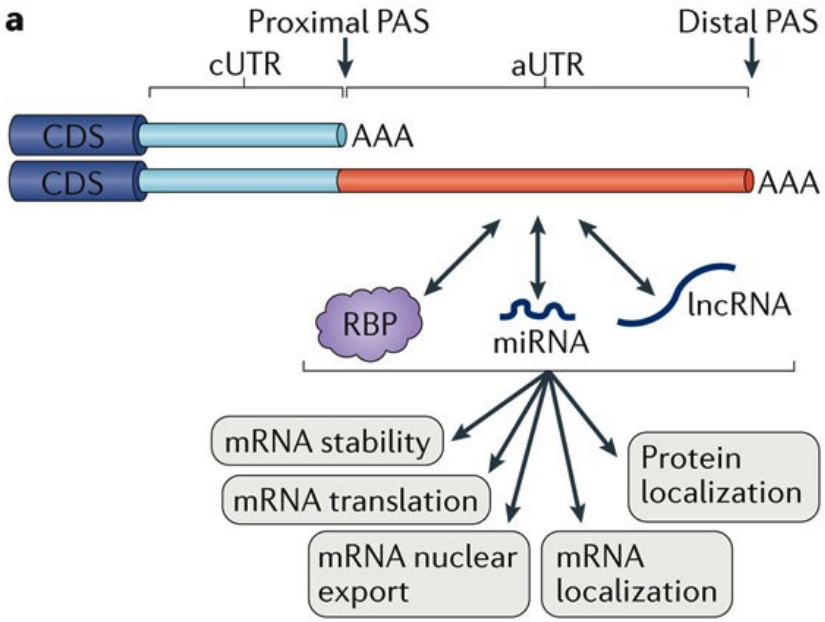
Alternative polyadenylation

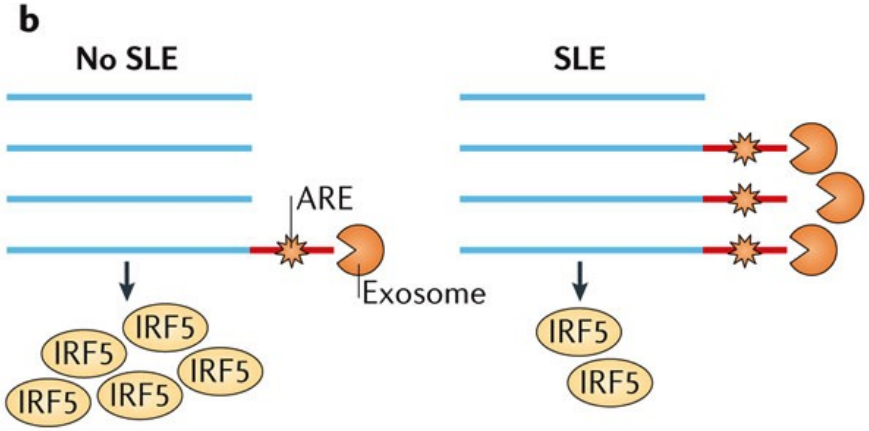
- 70–80% of mammalian genes have been reported to display alternative cleavage and polyadenylation site (APA) ,
- alternative pAs in the 3′-most exon (A) typically leads to variable 3′ UTRs, whereas pAs in upstream introns and exons (B) cause both coding sequence (CDS) and 3′ UTR changes,
- APA is dynamic under different biological conditions (cell growth and development, cancer)



Alternative polyadenylation

Most alternative polyadenylation (APA) sites are located in 3′ UTRs. As 3′ UTRs contain *cis* elements that are involved in various aspects of mRNA metabolism, 3′ UTR-APA can considerably affect post-transcriptional gene regulation in various ways, including through the modulation of mRNA stability, translation, nuclear export and cellular localization, and even through effects on the localization of the encoded protein.

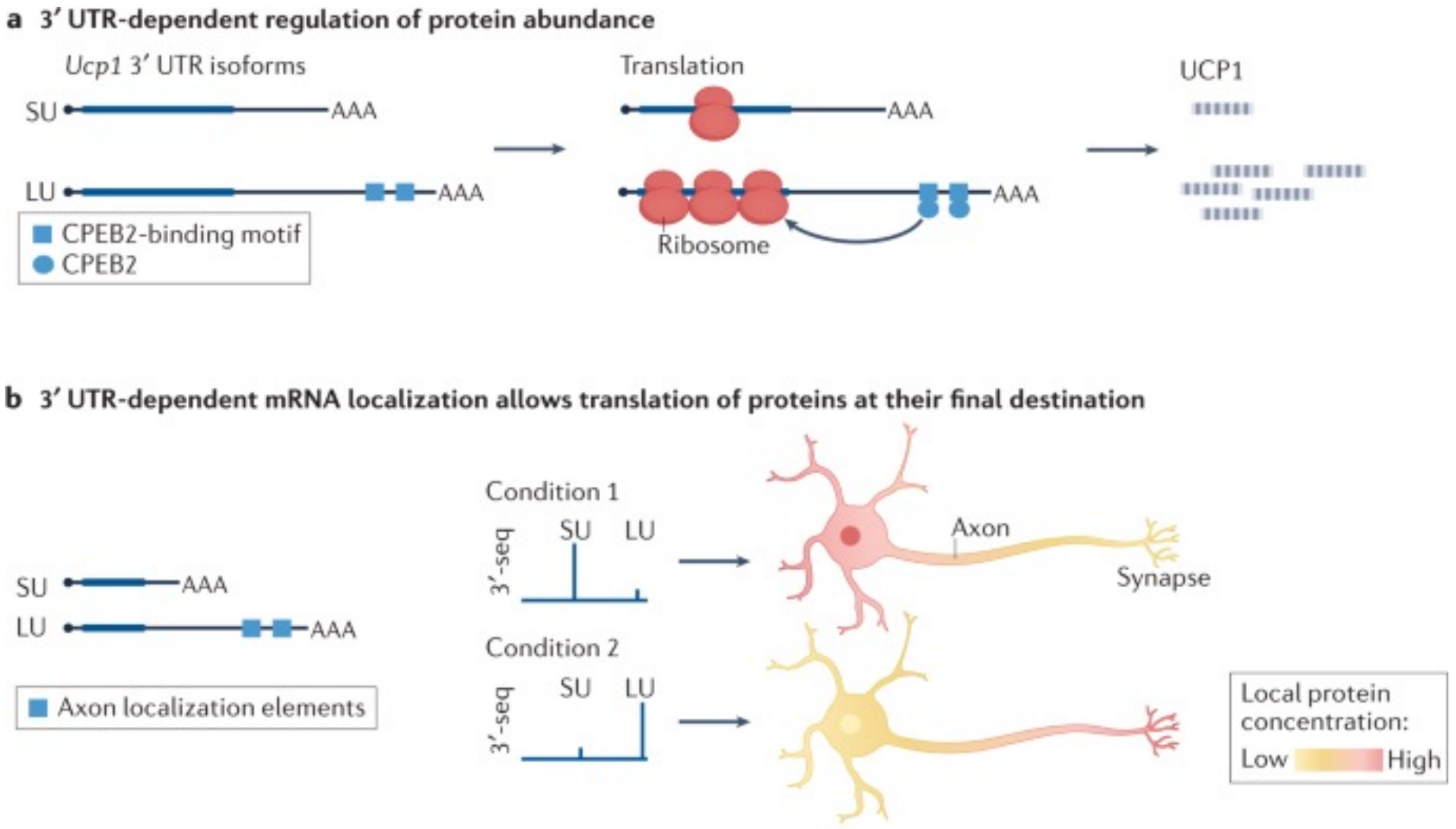




In patients with systemic lupus erythematosus (SLE), a single nucleotide polymorphism reducing the use of the proximal PAS leads to the production of long isoforms at the expense of short isoforms. The long 3′UTR contains a destabilization sequence (ARE) which results in reduced IFN-regulatory factor 5 (IRF5) levels.

Nature Reviews | Molecular Cell Biology

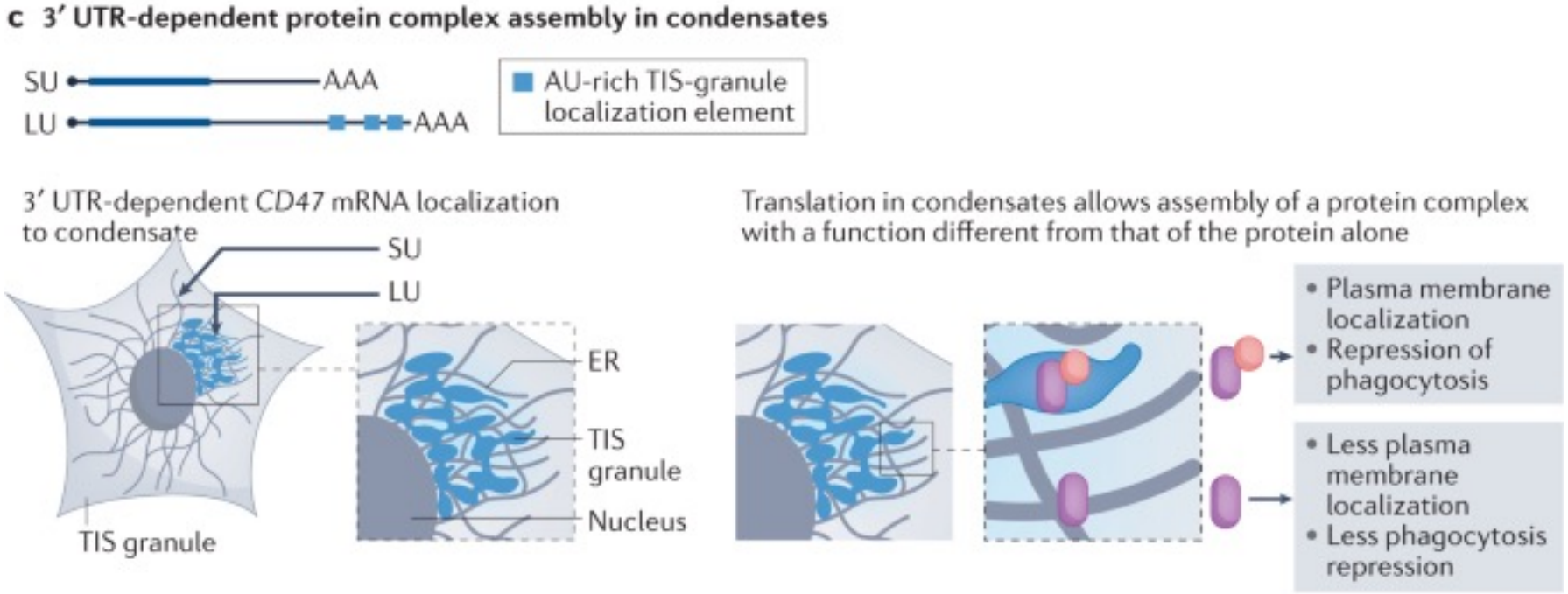
Alternative polyadenylation

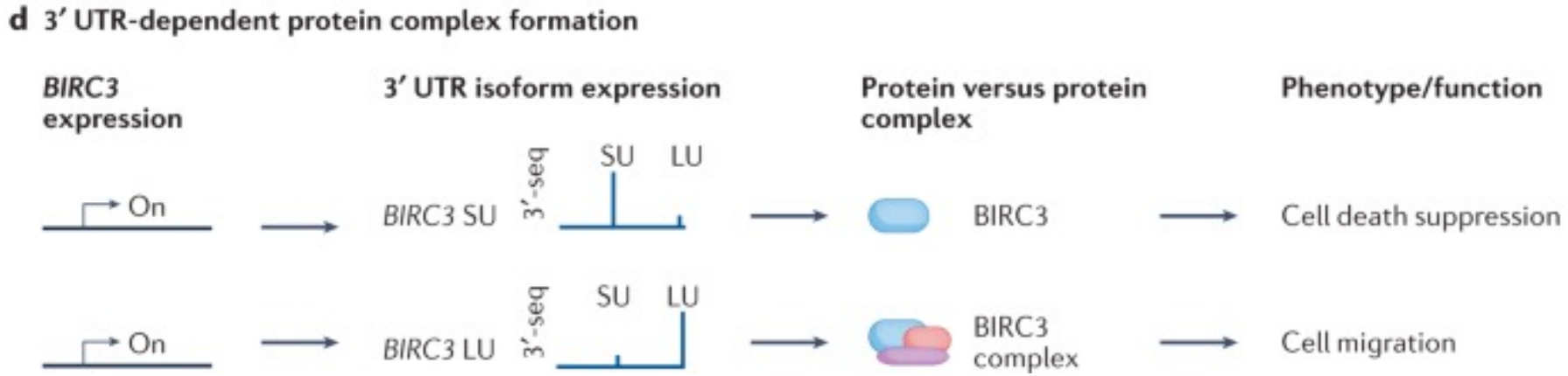


SU: short UTR

LU: long UTR

Alternative polyadenylation





Regulation of Alternative cleavage and polyadenylation

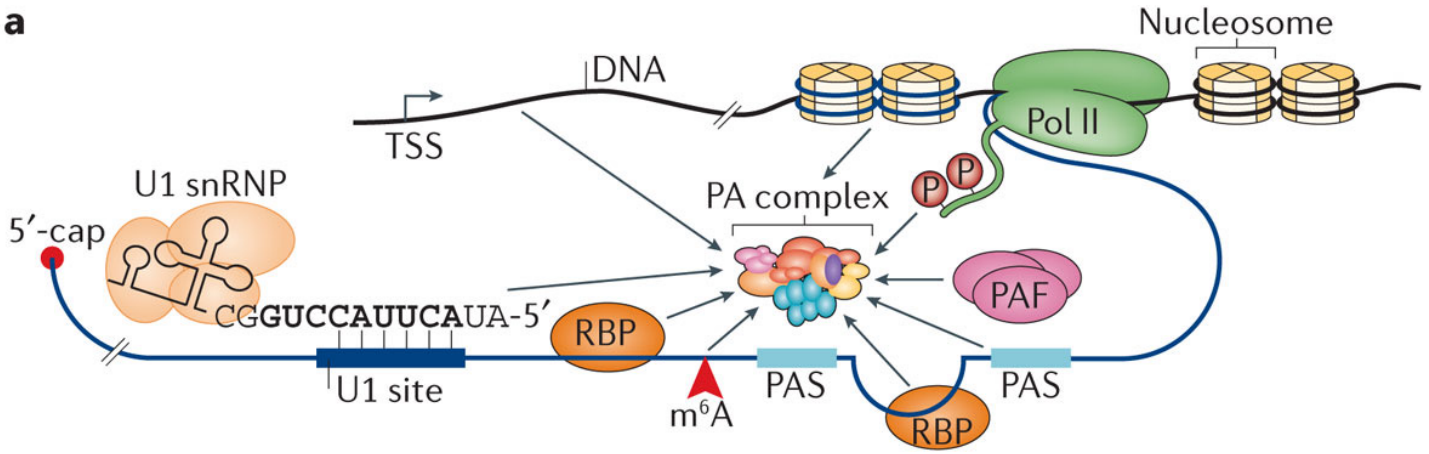
1. U1 can exert a negative role in pA usage

2. RNA binding proteins may inhibits pA usage by occluding the binding of core C/P factors, and some enhance pA usage by recruiting core factors.

3. Transcription activity impact pA choice by :

   - **cis elements** that cause pausing of RNA polymerase (RNAP) II, such as G-rich elements, facilitate pA usage,

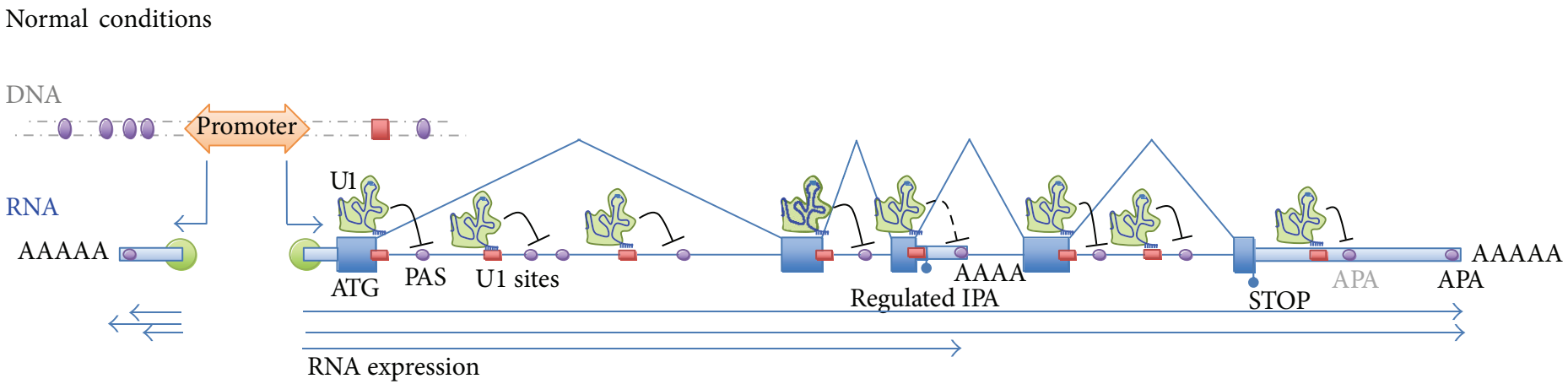
   - **factor recruitment** at the promoter can influence downstream pA usage.

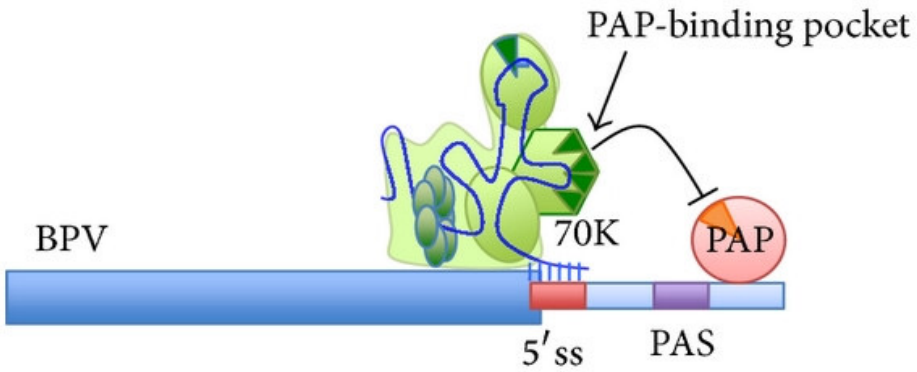


Nature Reviews | Molecular Cell Biology

Regulation of Alternative cleavage and polyadenylation

1. U1 can exert a negative role in pA usage



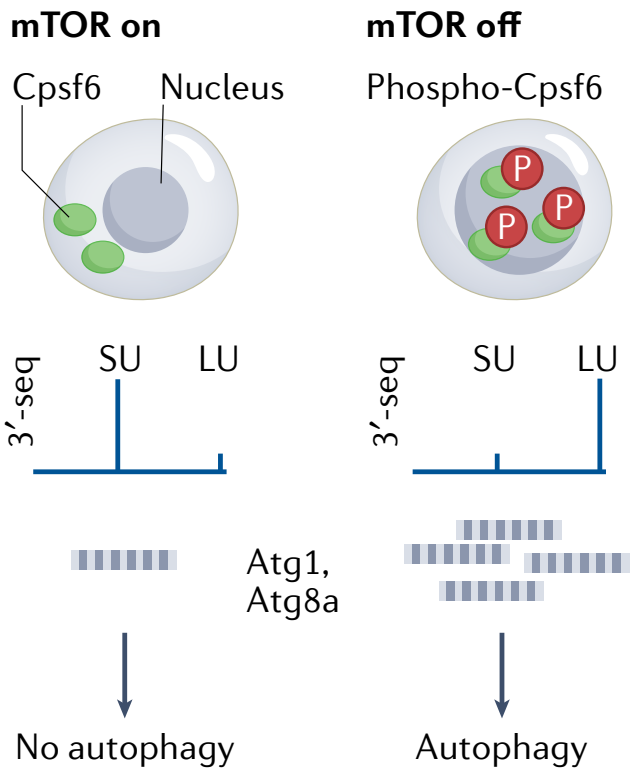


Nature Reviews | Molecular Cell Biology

Regulation of Alternative cleavage and polyadenylation

1. U1 can exert a negative role in pA usage

2. RNA binding proteins may inhibit pA usage by occluding the binding of core C/P factors, and some enhance pA usage by recruiting core factors.



Signalling-induced phosphorylation Cpsf6 (CFIm68), subunit of CFI, increases its activity and promotes autophagy

In flies, inactivation of the mTOR pathway allows the P of Cpsf6 in the cytoplasm, thereby promoting its translocation to the nucleus and RNA-binding activity.

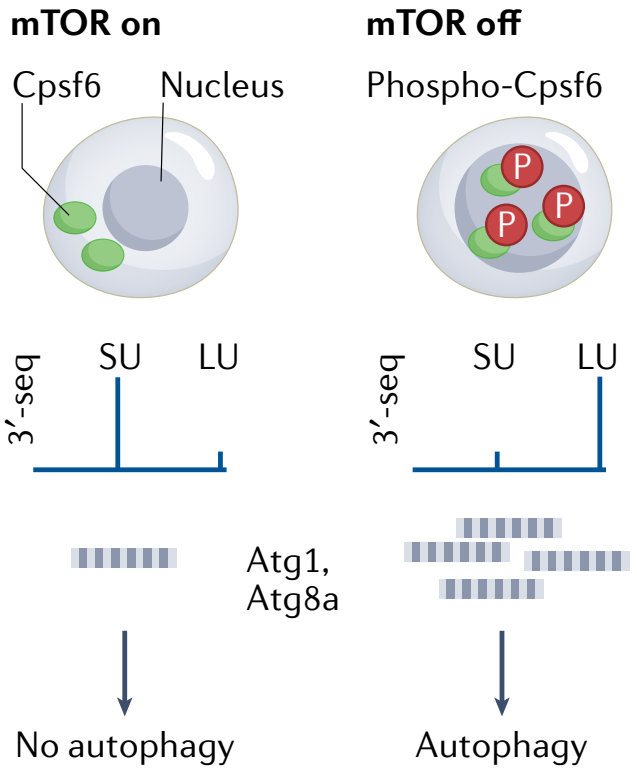
Cpsf6 changes the APA pattern of two master regulators of autophagy, Atg1 and Atg8a. Increased expression of their long 3′ UTR (LU) isoforms supports high- level protein expression.

Nature Reviews | Molecular Cell Biology

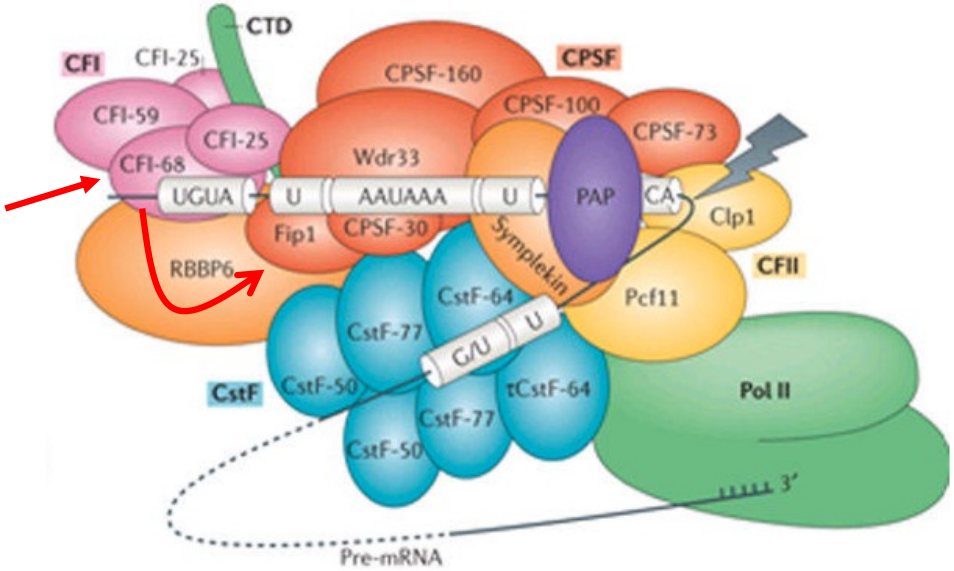
Regulation of Alternative cleavage and polyadenylation

1. U1 can exert a negative role in pA usage

2. RNA binding proteins may inhibit pA usage by occluding the binding of core C/P factors, and some enhance pA usage by recruiting core factors.

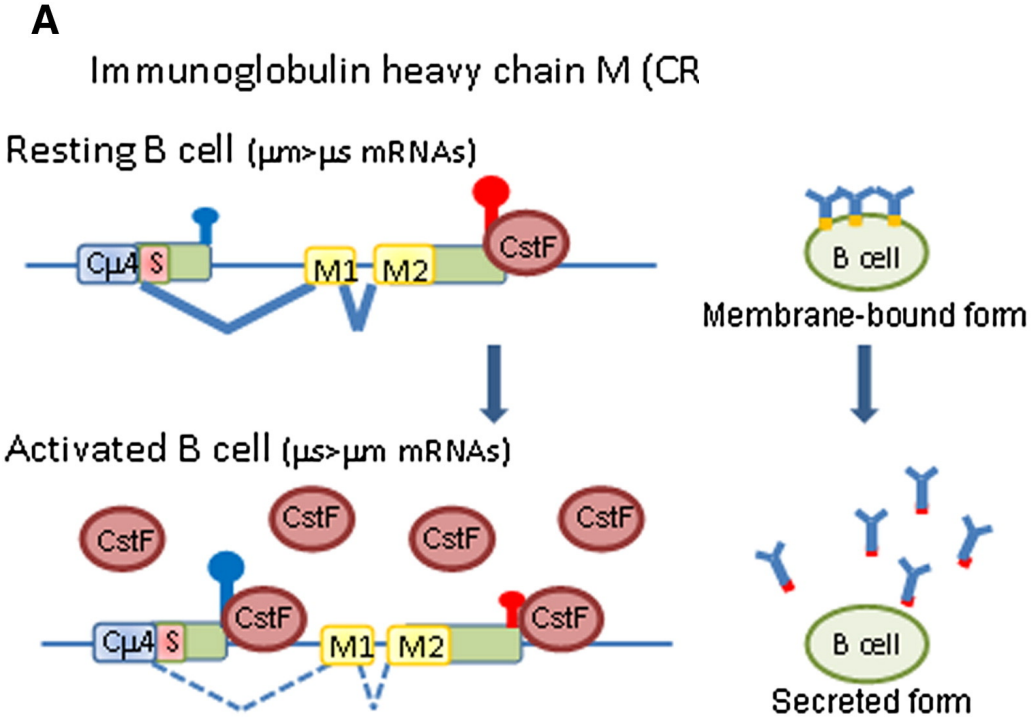


Signalling-induced phosphorylation Cpsf6 (CFIm68), subunit of CFI, increases its activity and promotes autophagy



Nature Reviews | Molecular Cell Biology

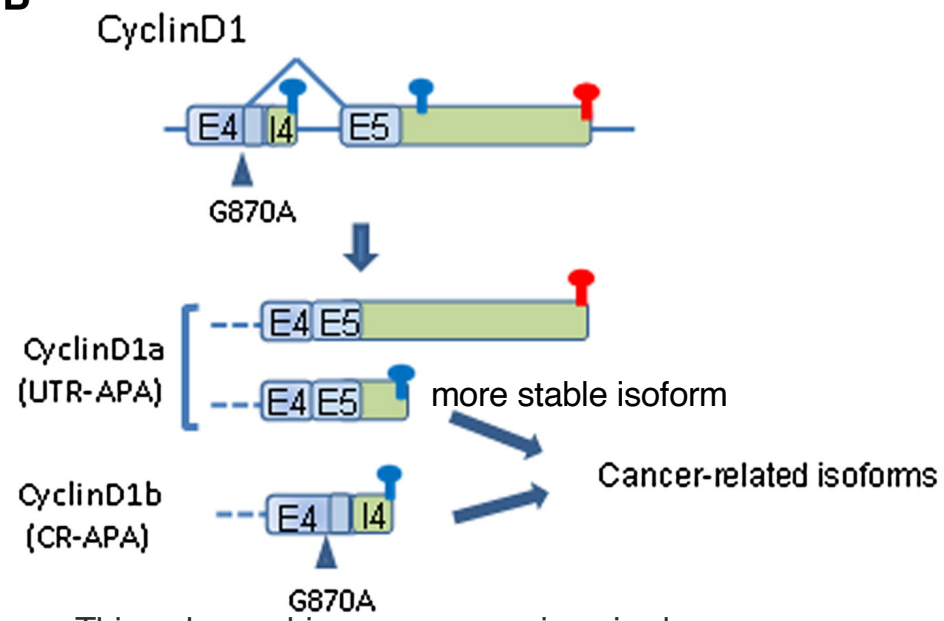
A



During the transition of a B cell to a plasma cell, the IgM protein switches from a membrane-bound form to a secreted form.

In resting B cells, the amount of CstF64 is limiting, and the distal poly(A) site, which binds Cst64 more avidly, is preferentially used, resulting in production of the membrane-bound form of IgM (µm). In activated B cells, the concentration of CstF is elevated and no longer limiting, so the proximal, first transcribed poly(A) site is preferentially selected, leading to production of secreted-form IgM (µs).

B



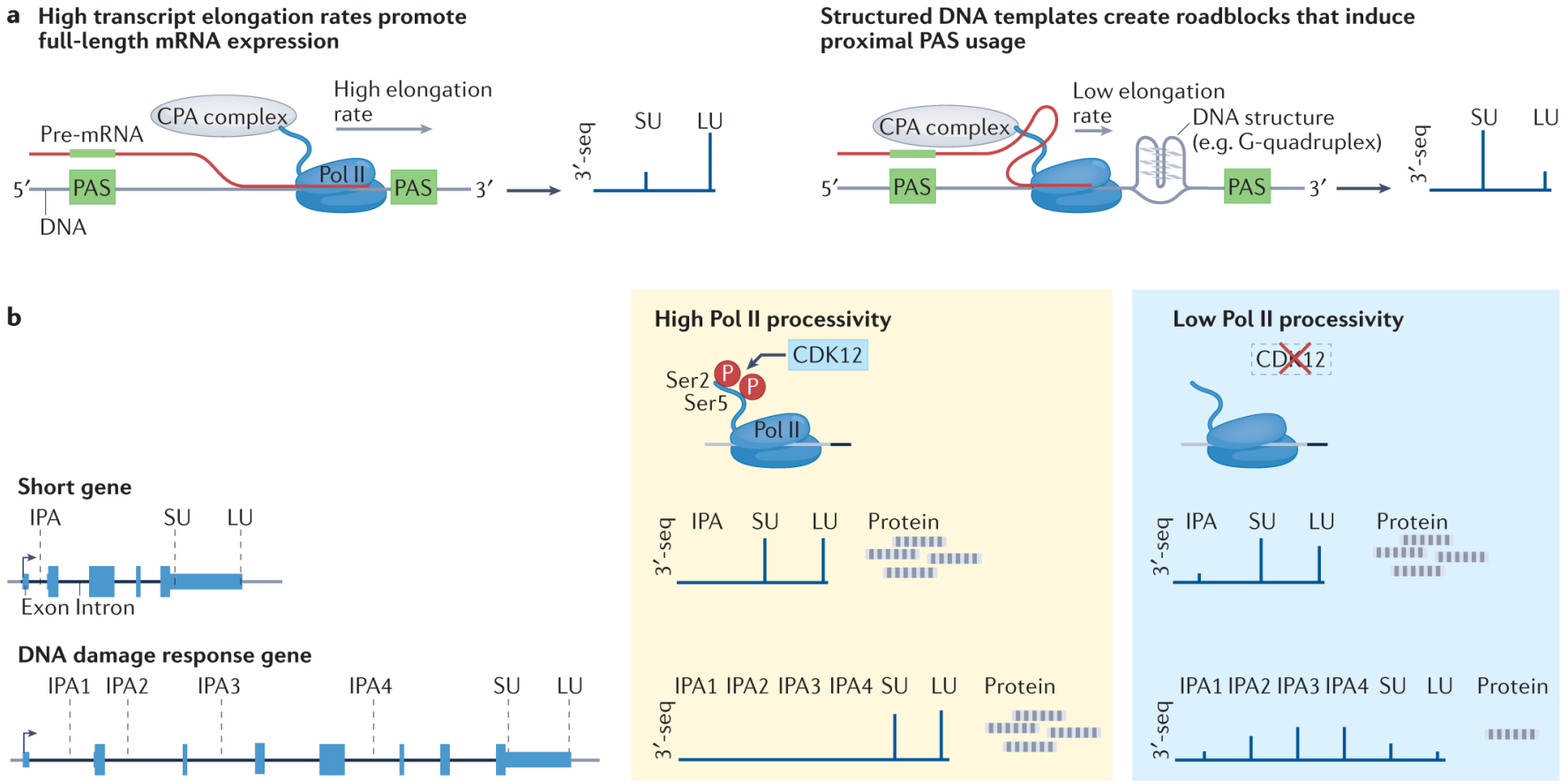
This polymorphism may cause impaired recognition by the splicing machinery, resulting in APA using the intron 4 PAS

Two major isoforms, cyclin D1a and b, are created by alternative splicing/polyadenylation (CR-APA)

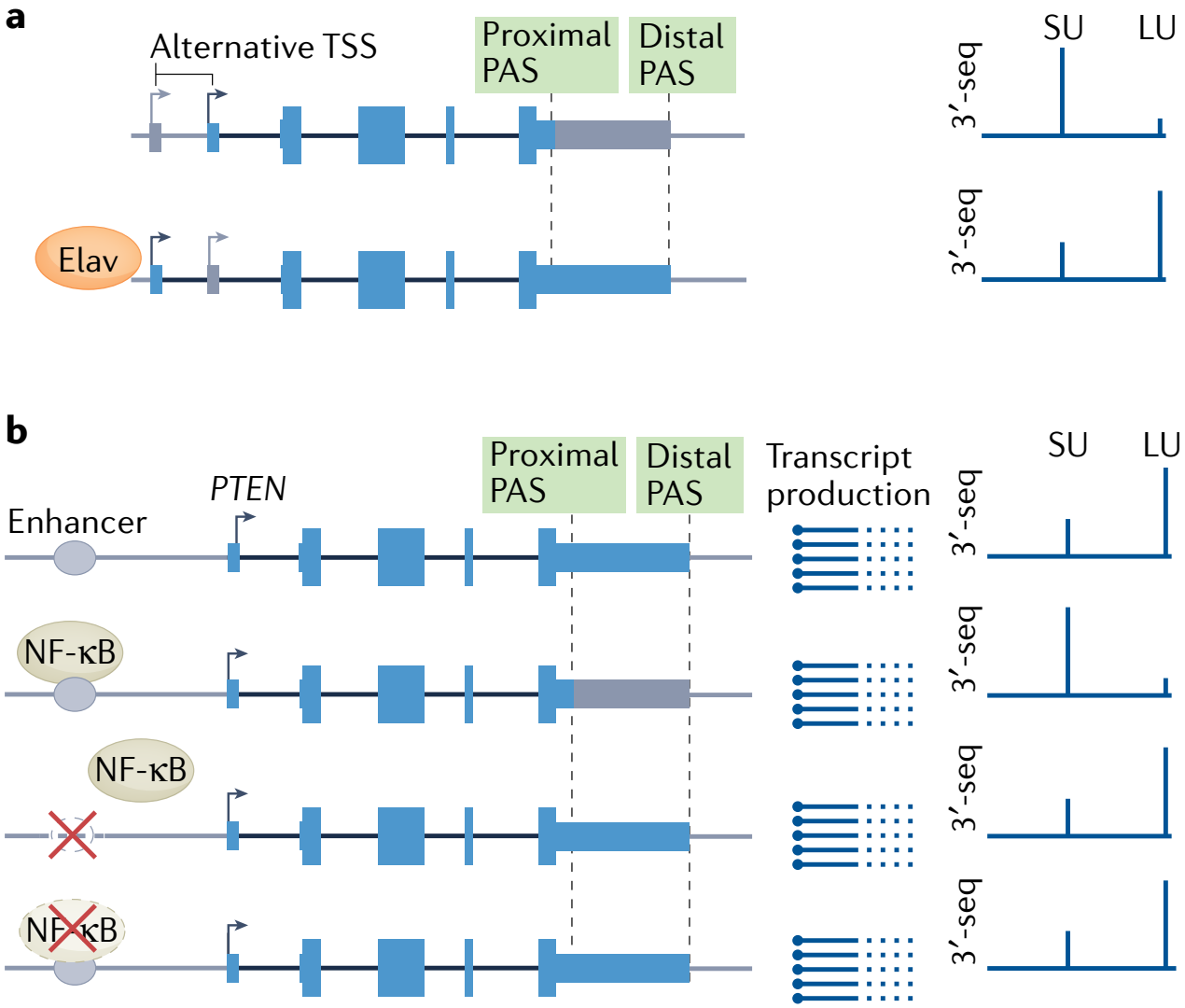
cyclin D1b mRNA is cleaved at an APA site within an intron. Cyclin D1b protein is constitutively nuclear, resulting in increased transforming capability. High expression of cyclin D1b is observed in several human cancers, including breast and prostate cancer. A G870A polymorphism at the end of exon 4 has been associated with production of the cyclin D1b isoform. This polymorphism may cause impaired recognition by the splicing machinery, resulting in APA using the intron 4 poly(A) signal

3. Transcription activity impact pA choice by :

- **cis elements** that cause pausing of RNA polymerase (RNAP) II, such as G-rich elements, facilitate pA usage,



- **factor recruitment** at the promoter can influence downstream pA usage.

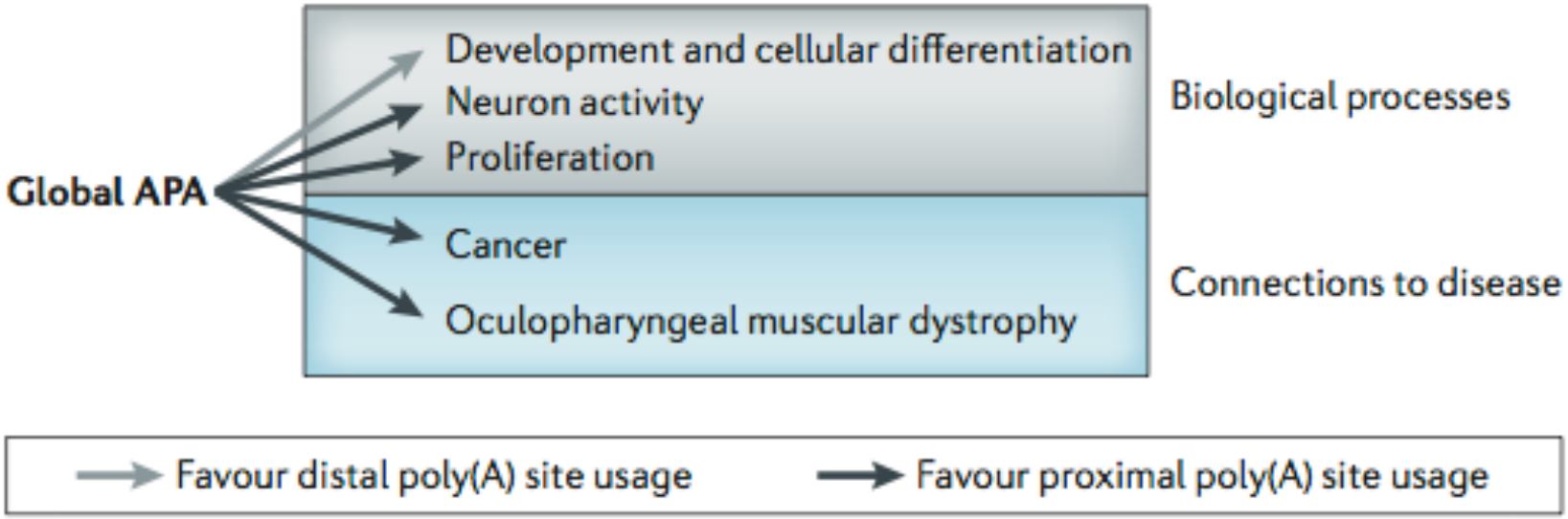


Mutations in poly(A) *cis*-elements that cause or contribute to human diseases

Disease	Disease description	Poly(A) site mutation	Affected gene	Comments	Refs
α-Thalassaemia	Thalassaemias are a group of common human genetic diseases that result from defects in haemoglobin production	AATAAA to AATAAG	*HBA2*		84
β-Thalassaemia		AATAAA to AACAAA	*HBB*	The mutation results in generation of an unstable transcript that is ~900 nt longer	85
Metachromatic leukodystrophy	A neurodegenerative disorder caused by null mutations in *ARSA*	AATAAC to AGTAAC	*ARSA*	Carriers of this mutation show reduced mRNA levels and enzyme activity but do not develop the disease symptoms; this condition is termed 'pseudodeficiency'	86,87
IPEX	A rare multifaceted and fatal disease	AATAAA to AATGAA	*FOXP3*		88
Fabry's disease	A rare and severe X-linked lysosomal storage disease caused by mutations in *GLA*	An AA dinucleotide deletion within the poly(A) site	*GLA*	*GLA* is one of the unusual mammalian genes that lack a 3′UTR and has its PAS within the CDS. The mutation in the poly(A) site results in aberrant 3′ end formation and multiple unstable transcripts	89

ARSA, arylsulfatase A; CDS, coding DNA sequence; *FOXP3*, forkhead box P3; *GLA*, alpha-galactosidase; *HBA2*, haemoglobin, alpha 2 (also known as alpha-2-globin); *HBB*, haemoglobin, beta (also known as beta-globin); IPEX, immune dysfunction, polyendocrinopathy, enteropathy, X linked; PAS, poly(A) signal; UTR, untranslated region. Only selected examples are discussed here. For a more thorough discussion, see REF. 64.

Biological processes that have been linked with broad APA modulation



Manipulation of alternative 3′ UTR expression.