Bi 8 Lecture 3

DNA to RNA: Transcription and Splicing

Katalin Fejes Toth

12 January 2016

Readings for this lecture

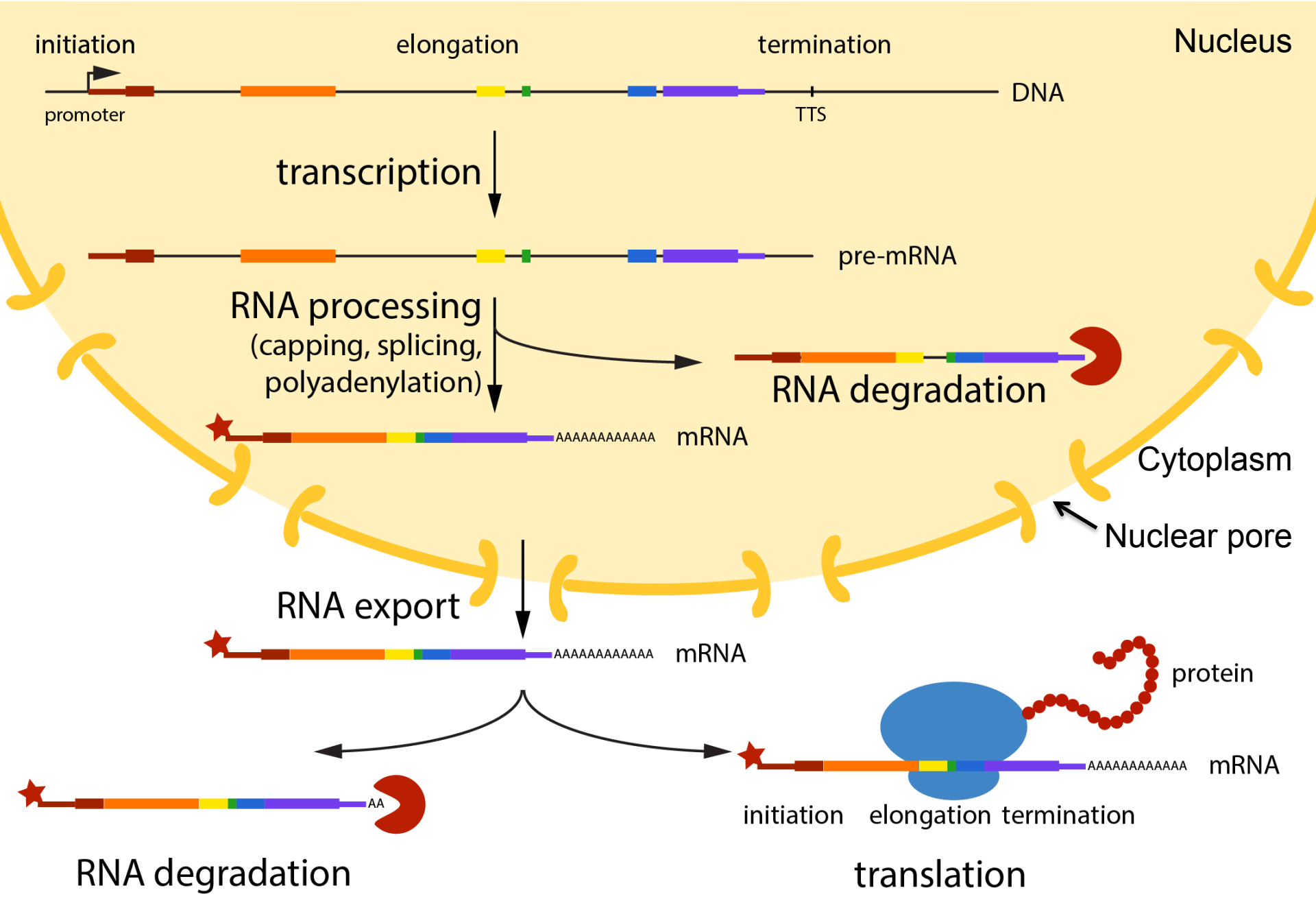
- Alberts 6th edition:
  - Ch. 6: pp. 301-333

- (Alberts 5th edition equivalents:
  - Ch. 6: pp. 329-366)

Suggested seminar

James Manley (Columbia University):
"mRNA Processing and Links to Human Disease"
Wednesday, January 13th, 4pm Kerckhoff 119

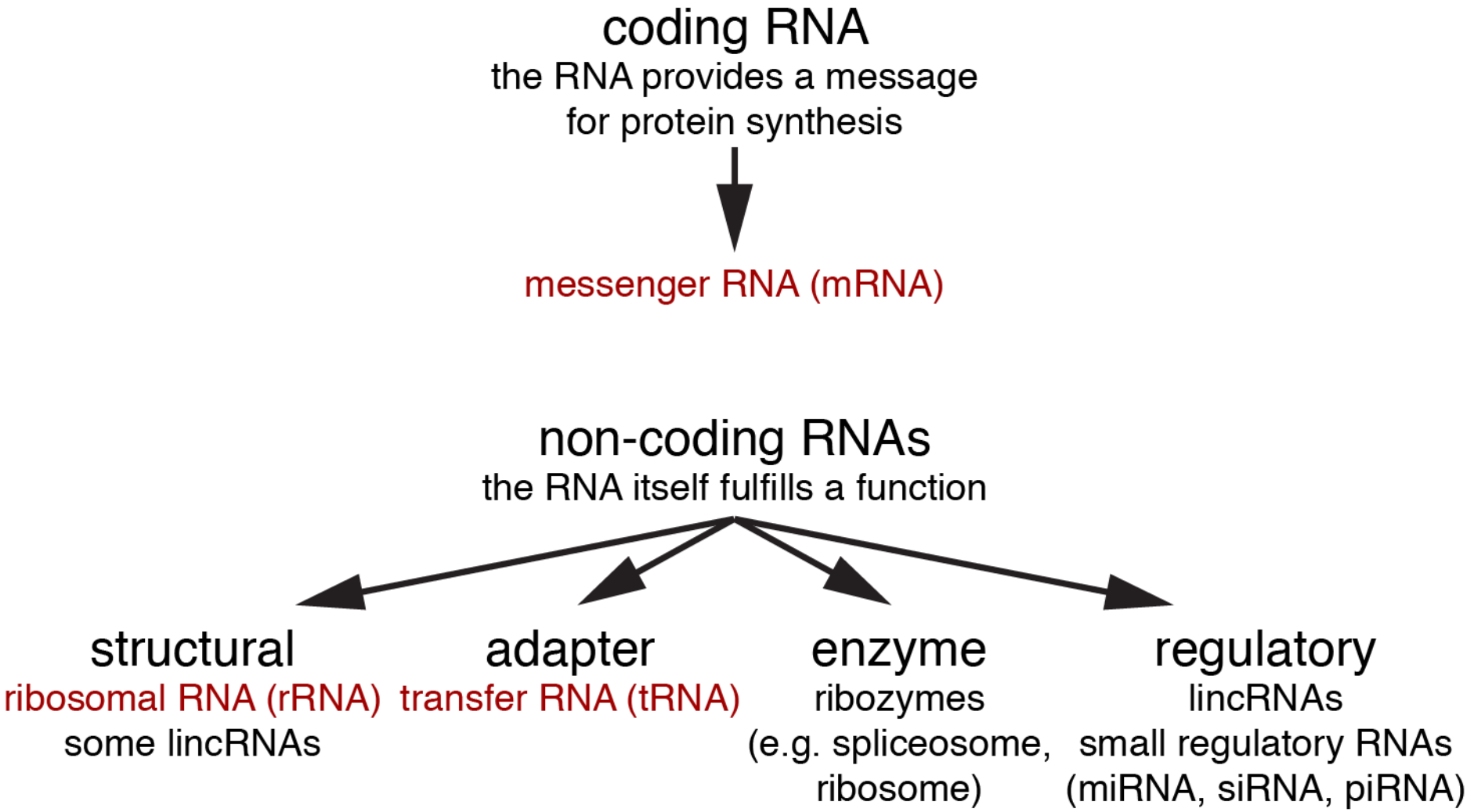
Steps of gene expression



Spatial separation of transcription and translation

- Transcription takes place in the nucleus and mRNAs translocate to the cytoplasm in order to be translated
- Translation is decoupled from transcription (unlike in prokaryotes)
- This allows for
  - transcript processing (e.g. splicing) – enhancing gene function
  - More precise regulation of steps of gene expression
  - Better quality control (barrier of nuclear pore)

Classification of RNAs



Function of RNA is determined by primary or secondary structure

- For some RNAs the base sequence is crucial for function
  - mRNAs
  - Small regulatory RNAs

- Some RNAs fold to form structures that are important for their function (scaffold, catalytic site etc.)
  - tNRAs
  - rRNAs
  - Some lincRNAs (Xist)

The three major kinds of RNA

- rRNA = ribosomal RNA
  - Structural and enzymatic components of protein synthesis
  - ~80% of total RNA in cell, stable, highly folded
- tRNA = transfer RNA
  - Small adaptors that "read" genetic code for protein synthesis
  - ~15% of total RNA in cell, stable, highly folded
- mRNA = messenger RNA
  - Less than 5% of total RNA in cell in aggregate despite ~10^4 different sequences expressed per cell
  - Codes for proteins
  - Special modifications: 3' poly(A) and 5' "cap"
  - Highly dynamic and regulated via synthesis and stability

In eukaryotic cells different RNAs are transcribed by different RNA polymerases

TABLE 7-2 Classes of RNA Transcribed by the Three Eukaryotic Nuclear RNA Polymerases and Their Functions

POLYMERASE	RNA TRANSCRIBED	RNA FUNCTION
RNA polymerase I	Pre r-RNA (28S, 18S, 5.8S rRNAs)	Ribosome components, protein synthesis
RNA polymerase II	mRNA snRNAs miRNAs	Encodes protein RNA Splicing Post-transcriptional gene control
RNA polymerase III	tRNAs 5S rRNA snRNA U6 7S RNA Other stable short RNAs	Protein synthesis Ribosome component, protein synthesis RNA Splicing Signal-recognition particle for insertion of polypeptides into the endoplasmic reticulum Various functions, unknown for many

Table 7-2
Molecular Cell Biology, Sixth Edition

Different polymerases means ability to regulate by different cofactors!

Profound similarities in “business part” of RNA polymerase structures from different organisms

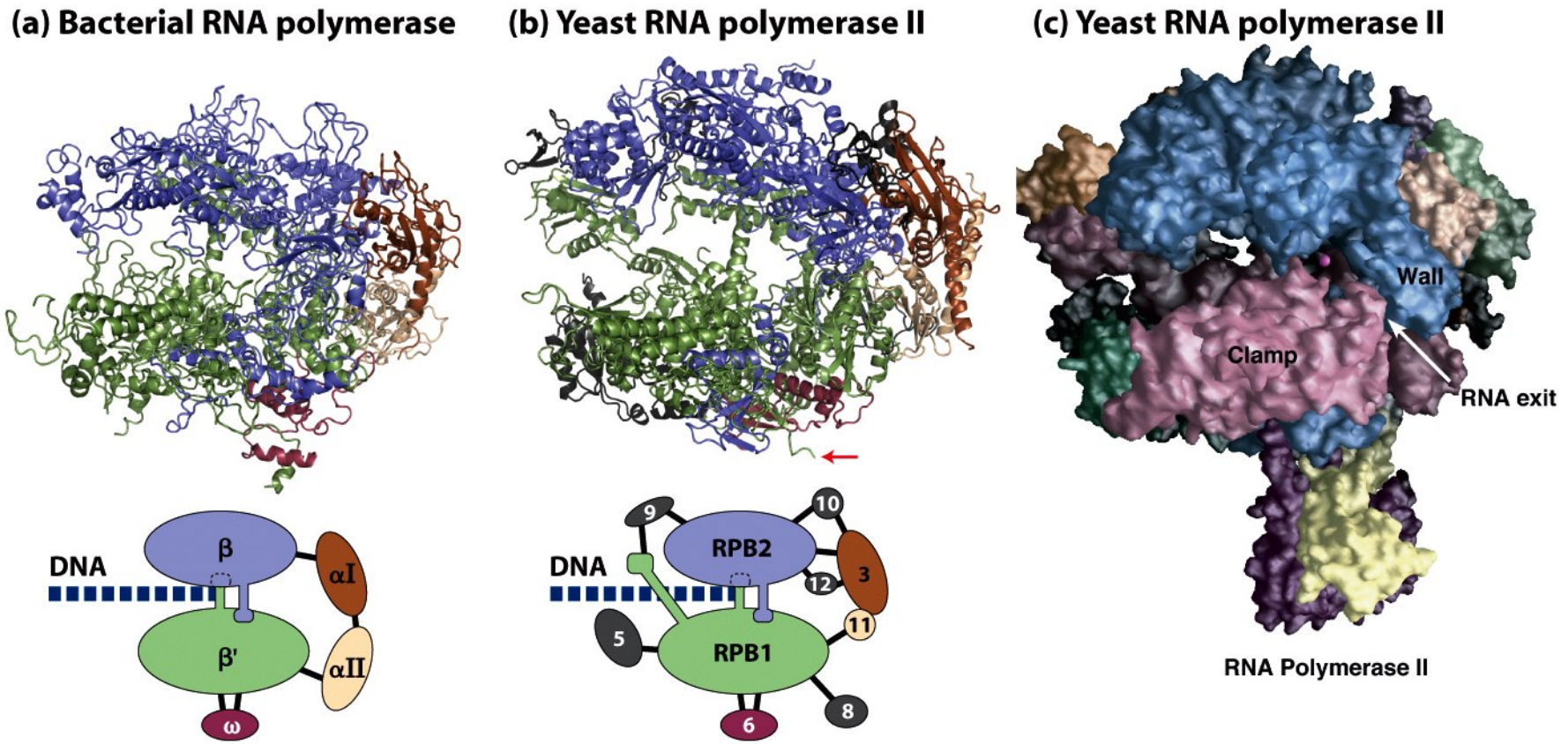


Figure 7-9
Molecular Cell Biology, Sixth Edition

All ***three*** eukaryotic RNA polymerases are structurally related to the ***single*** bacterial RNA polymerase

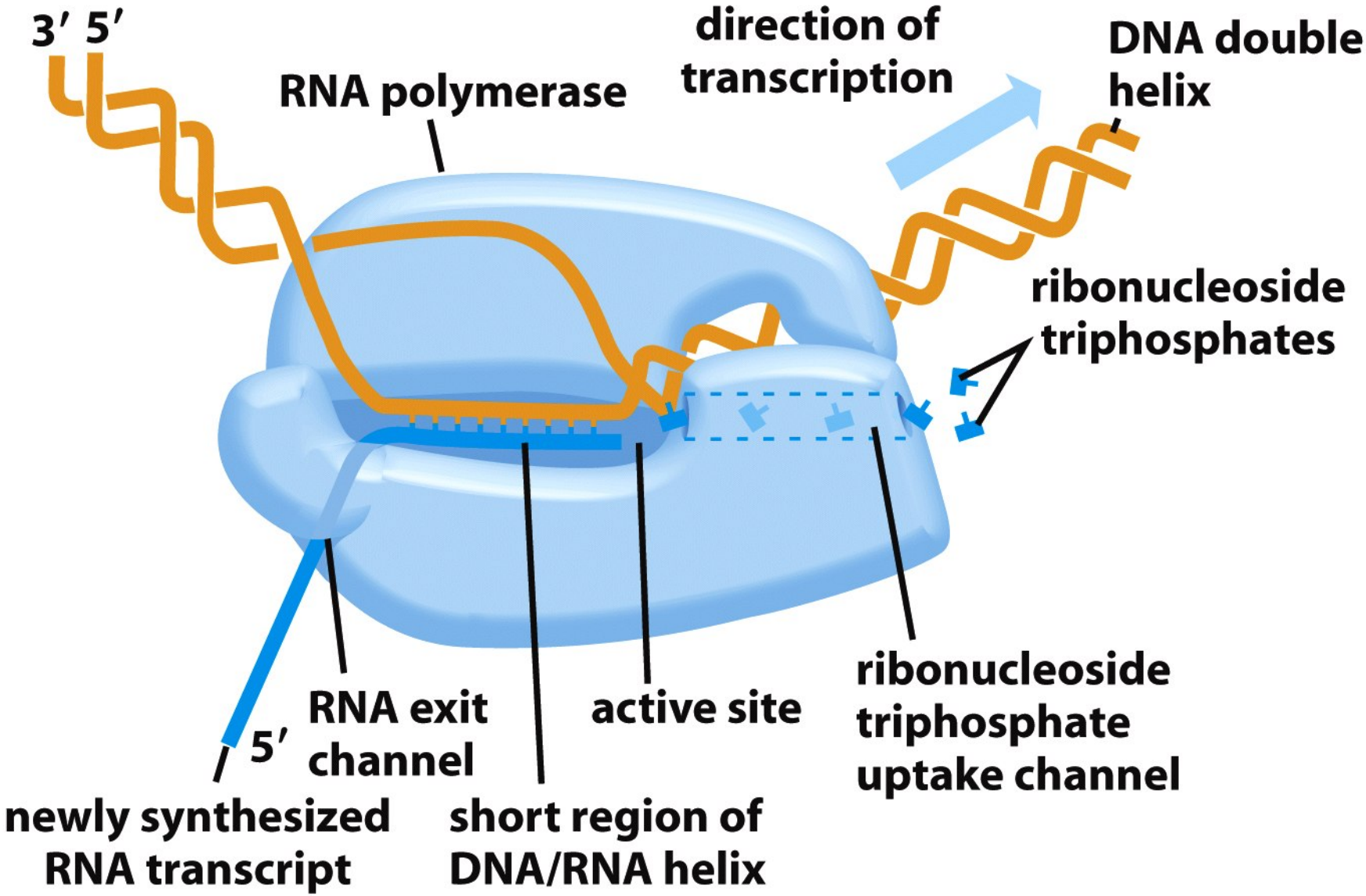
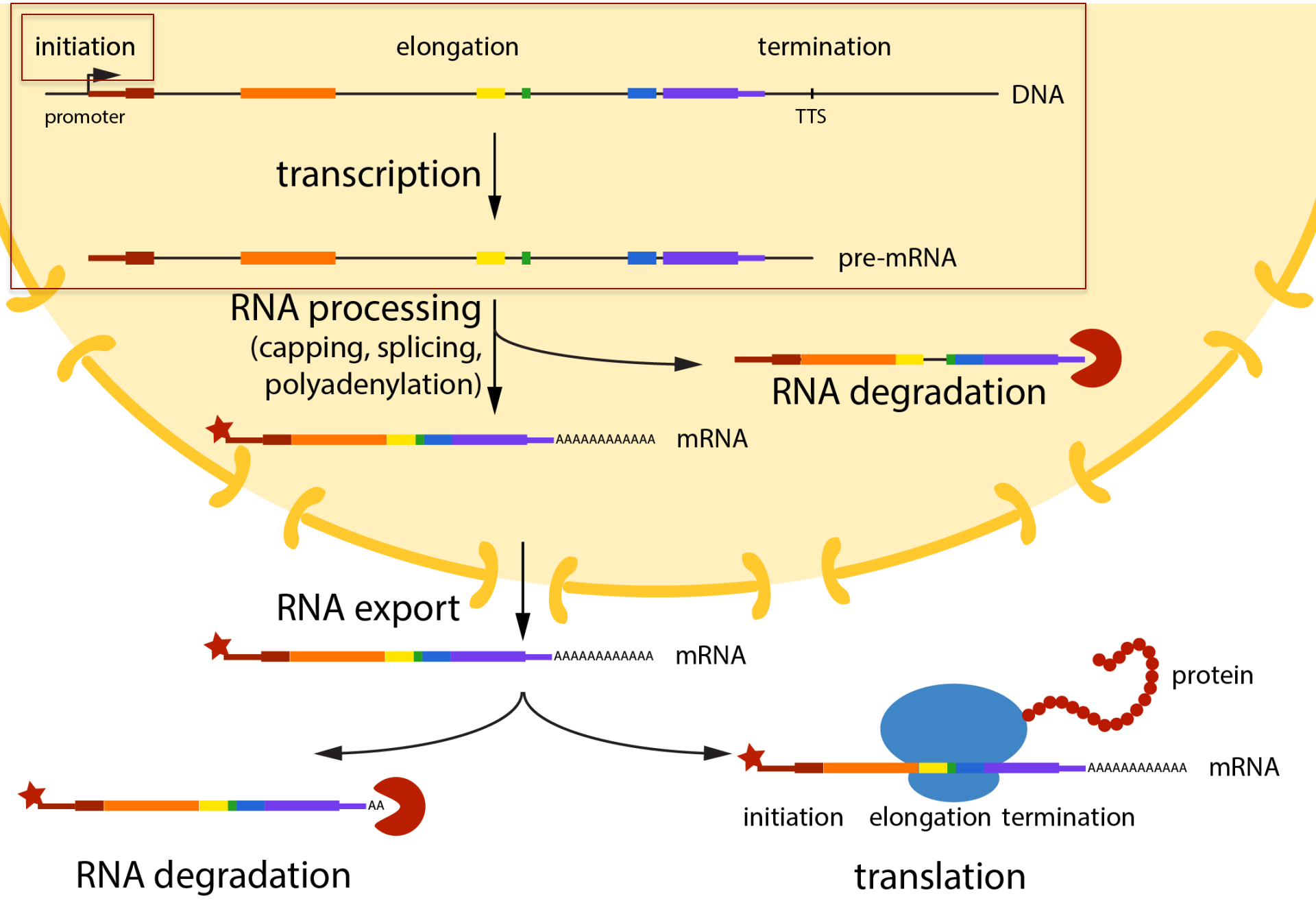


Figure 6-8a *Molecular Biology of the Cell* (© Garland Science 2008)

Two kinds of nucleic acid polymerases work differently: RNA transcription vs. DNA replication

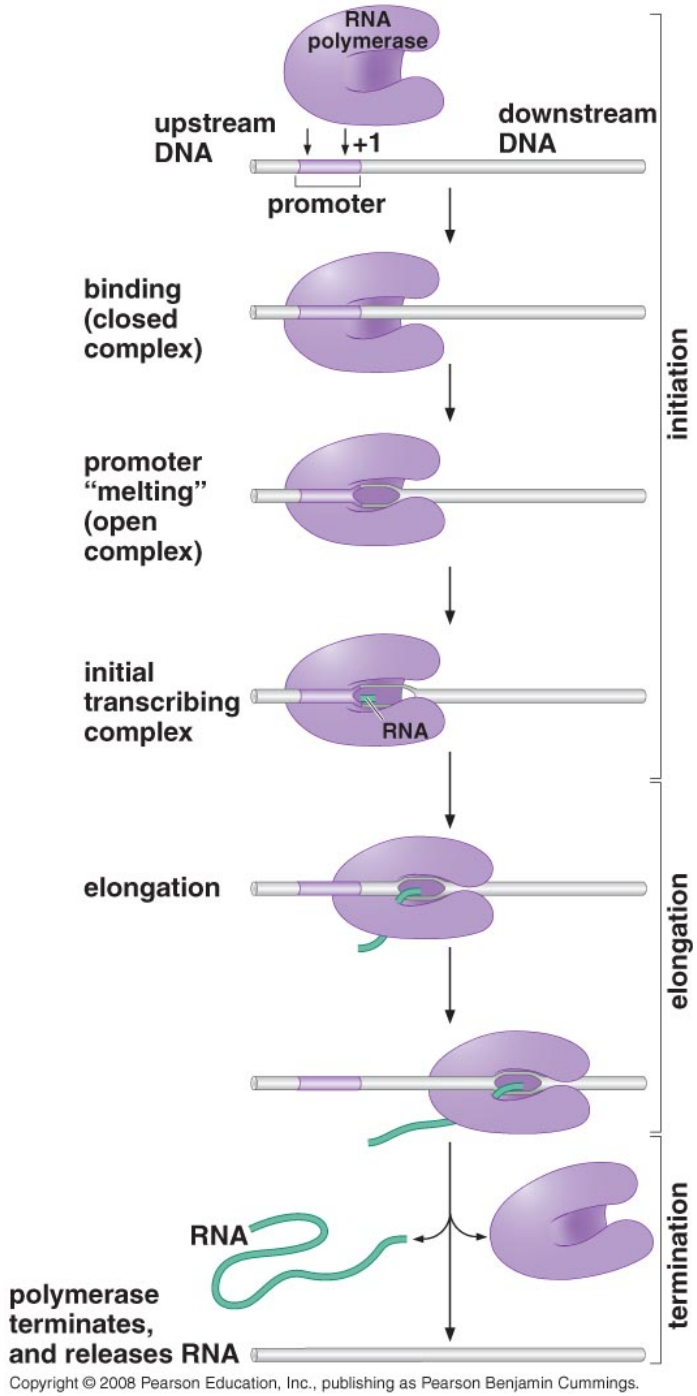
- Both DNA synthesis and RNA synthesis are catalyzed by ***polymerases***
- RNA transcription only copies one strand of a given gene; but both DNA strands must be replicated in exact 1:1 ratio
- RNA polymerase error rates ~10^{-4} ; DNA polymerase error rates ~10^{-7}
- RNA polymerase starts without a primer, wherever it is recruited and released; DNA replication ***always requires*** a primer
- RNA transcription rates are extremely varied between genes; DNA replication must be uniform across genome

Steps of gene expression



Initiating RNA synthesis

Common features for all



Finding a site

Exposing a single DNA strand to serve as template

"Moving the bubble"

Termination

RNA polymerases have a hard job

- Start polymerization of RNA without a primer… *WHERE?*
- Must be able to melt DNA strands to open up a template without a primer to form competing hybrid
- Energetically unfavorable – must break all those H bonds
- Hold first and second rNTPs together with melted DNA before any kind of synthesis can start

Crucial help provided by cofactors

Bacterial σ factor: Recognizes promoter and helps to melt DNA

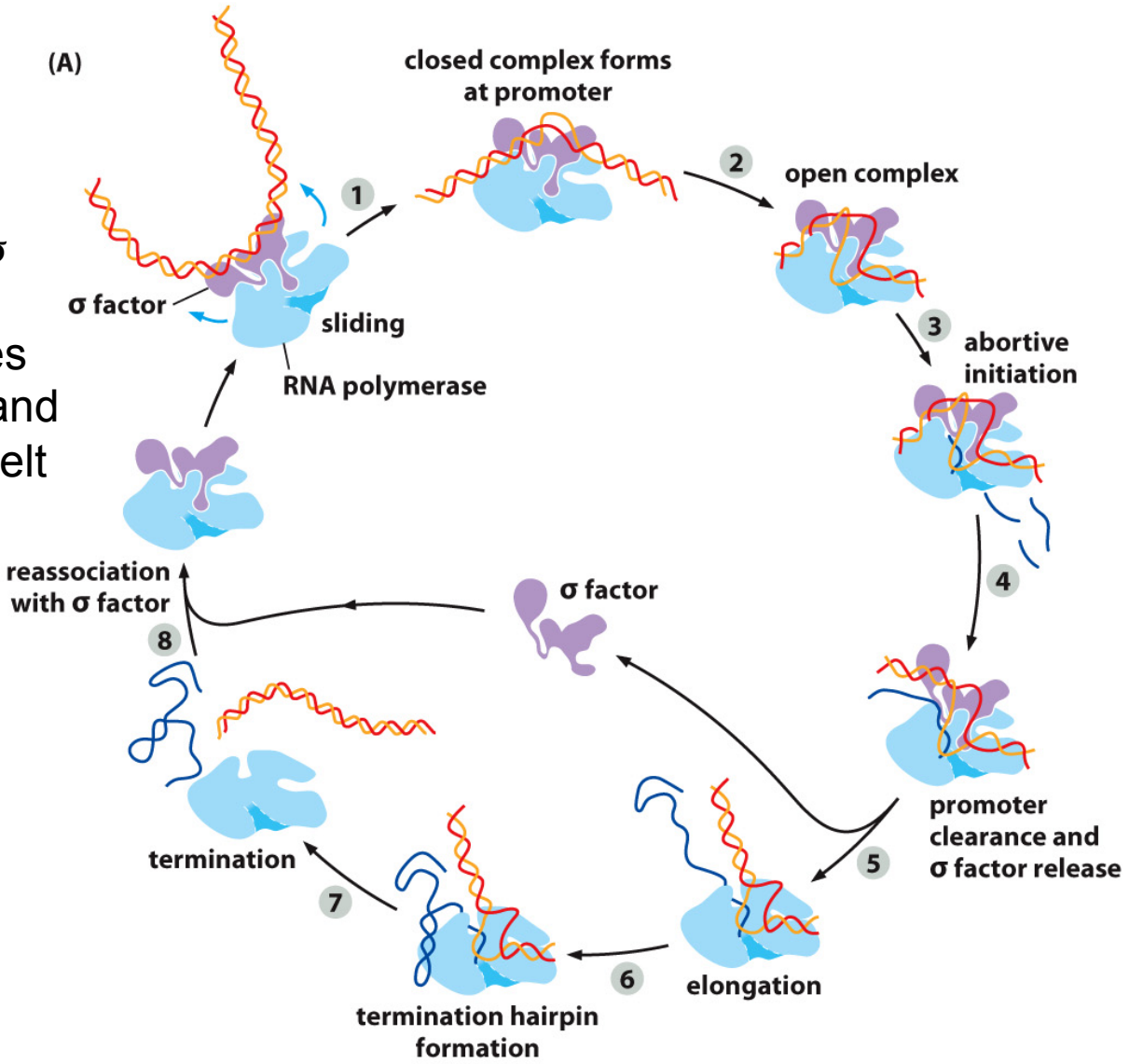
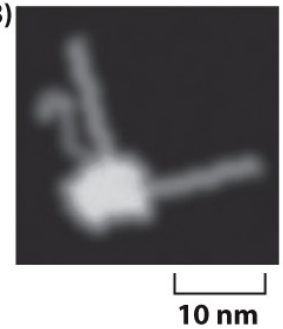
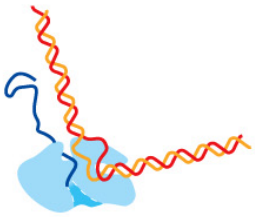


Figure 6-11 Molecular Biology of the Cell 6e (© Garland Science 2015)





Bacterial σ factor: Needs to be released for polymerase to escape the promoter

Figure 6-11 *Molecular Biology of the Cell* (© Garland Science 2008)

Eukaryotic pol II molecules are often helped by “TFII” factors binding to a “TATA box” about 30 bp from the start of transcription

“TAFs” (TBP-associated factors) are also recruited to form huge molecular machine

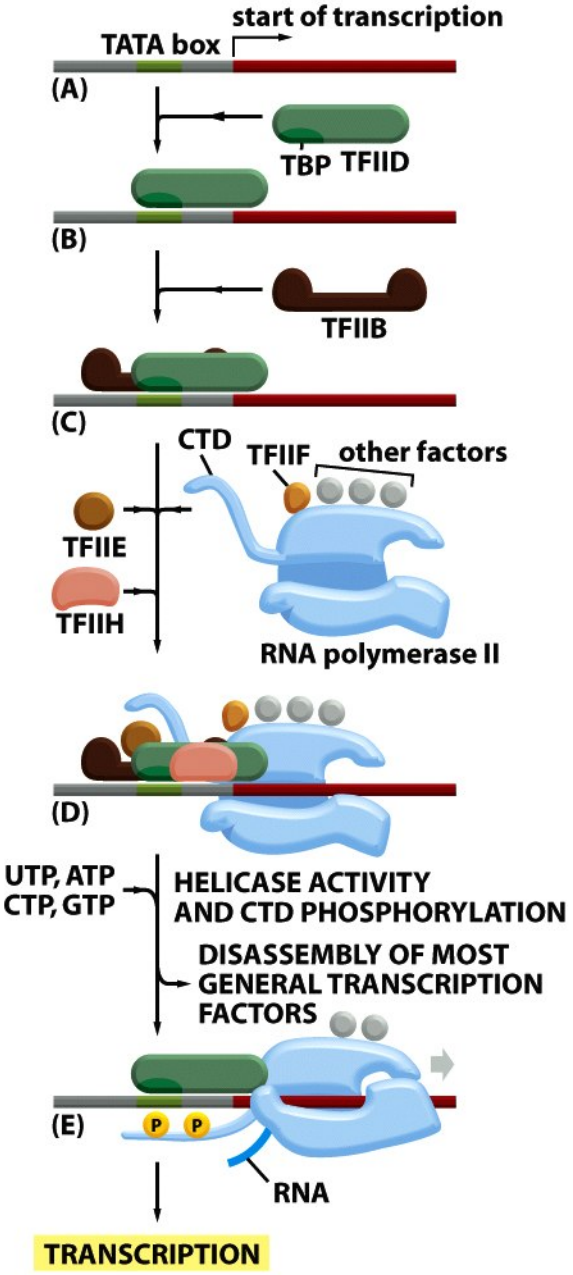


Figure 6-16 *Molecular Biology of the Cell* (© Garland Science 2008)

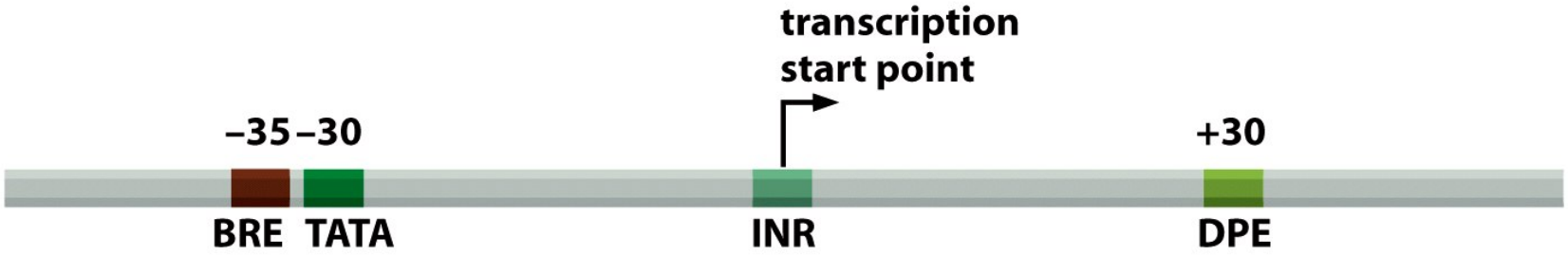
Targeting/recruitment cofactors for pol II include weakly sequence-specific DNA binding factors

TABLE 6–3 The General Transcription Factors Needed for Transcription Initiation by Eukaryotic RNA Polymerase II

Name	Number of subunits	Roles in transition initiation
TFIID TBP subunit TAF subunits	 1 ~11	 Recognizes TATA box Recognizes other DNA sequences near the transcription start point; regulates DNA-binding by TBP
TFIIB	1	Recognizes BRE element in promoters; accurately positions RNA polymerase at the start site of transcription
TFIIF	3	Stabilizes RNA polymerase interaction with TBP and TFIIB; helps attract TFIIE and TFIIH
TFIIE	2	Attracts and regulates TFIIH
TFIIH	9	Unwinds DNA at the transcription start point, phosphorylates Ser5 of the RNA polymerase CTD; releases RNA polymerase from the promoter

TFIID is composed of TBP and ~11 additional subunits called TAFs (TBP-associated factors); CTD, C-terminal domain.

TATA boxes: only the most common among several promoter-associated motifs

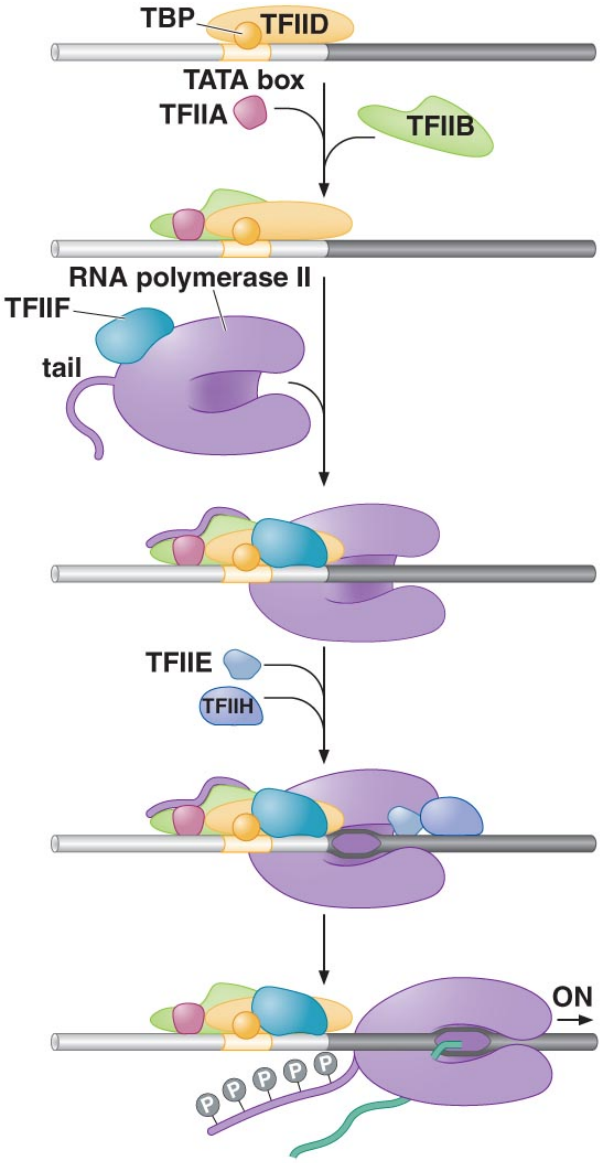


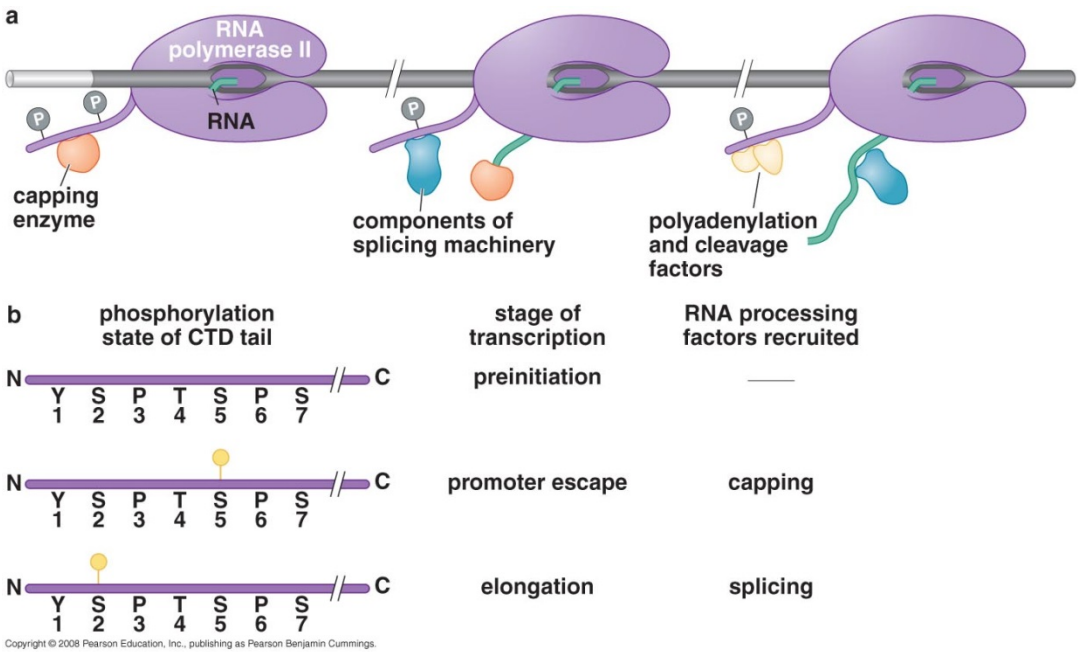
element	consensus sequence	general transcription factor
BRE	G/C G/C G/A C G C C	TFIIB
TATA	T A T A A/T A A/T	TBP
INR	C/T C/T A N T/A C/T C/T	TFIID
DPE	A/G G A/T C G T G	TFIID

But…. Caveat emptor!! These sites are *not* highly specific *nor* conserved in location *nor* present in all cases

Figure 6-17 *Molecular Biology of the Cell* (© Garland Science 2008)

Eukaryotic Pol II: need to phosphorylate C-terminal tail to leave recruitment TF II complex factors behind



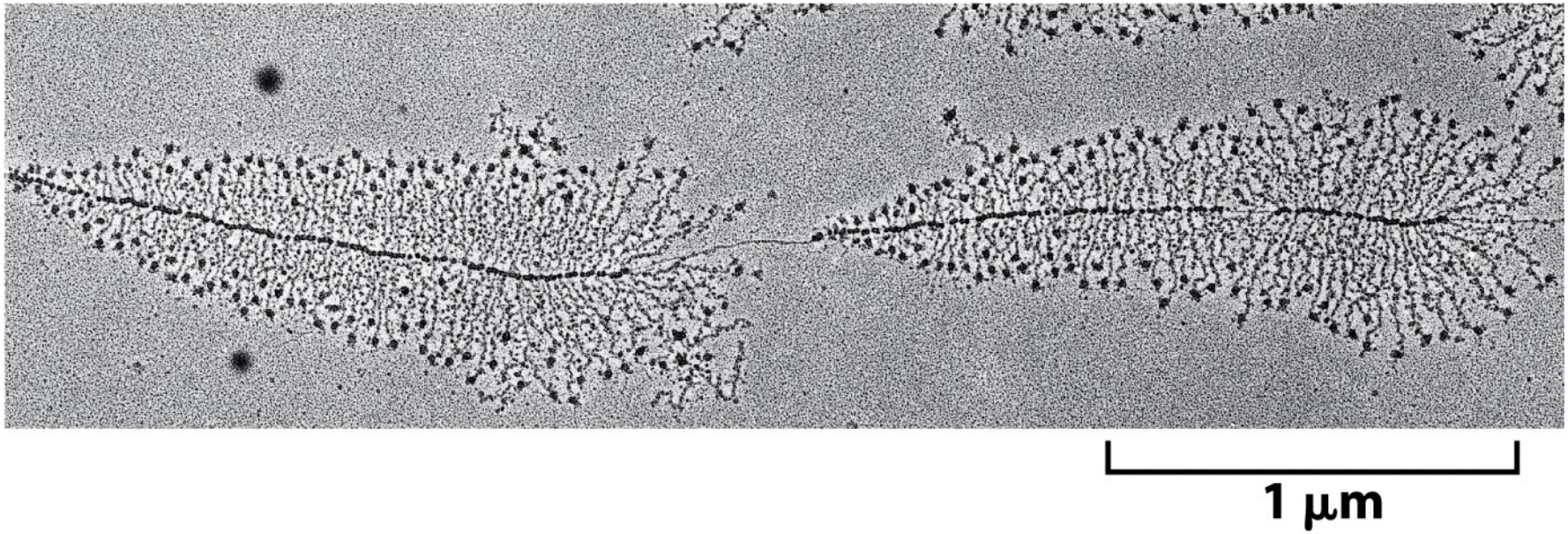


Points for regulation of transcription initiation

- Recruitment of polymerase
  - Depends on an accessible initiation site – TATA or Inr
  - Depends on TFII complex factors (pretty universal)
  - Binding may be facilitated by sequence-specific DNA binding transcription factors
- Licensing for transcriptional initiation
  - Involves large multiprotein complex, "Mediator"
  - Communication between transcription factors and polymerase complex subunits
- Release of polymerase from its initial recruitment site!
  - A newly recognized activity that may need to be triggered by sequence-specific TFs
  - "Paused" or "stalled" polymerase → no useful gene expression

Elongation: rate-limiting

Maximum transcription rate set by elongation rate…
1st polymerase needs to move >100 bp out of the way before next one can load!



Actual number of mRNAs transcribed depends on the elongation rate of the polymerase and how quickly polymerases follow each other.

Figure 6-9 *Molecular Biology of the Cell* (© Garland Science 2008)

Measured rates

- Measured in sea urchin, at 15° C:
  - translocation rate 6-9 nt/sec
  - must take at least 11-17 sec to go 100 nt

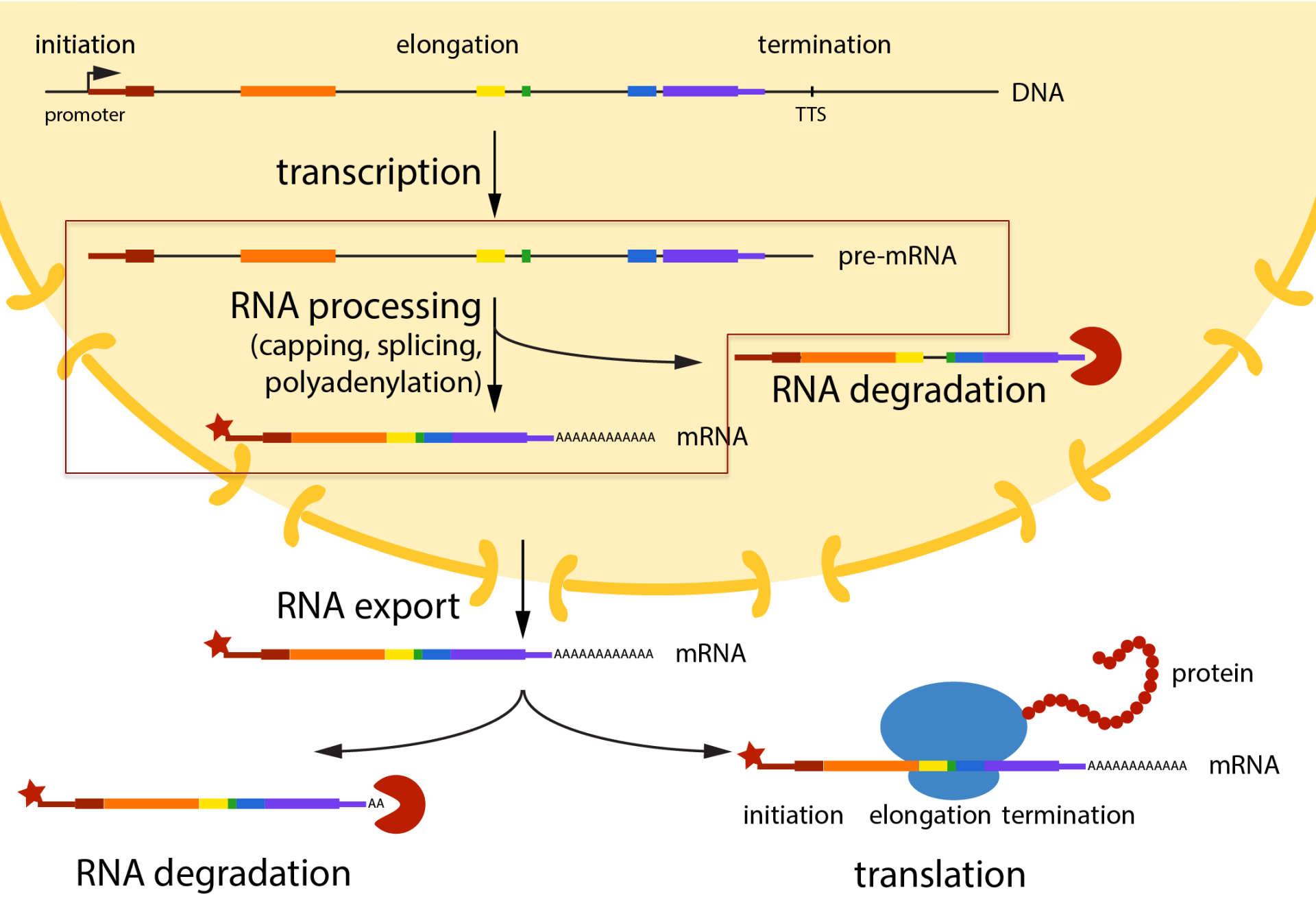
- In mammalian cells at 37° C:
  - translocation rates ~30-50 nt/sec (Q_{10} ~2-2.5)
  - can take at least 2-3.5 sec to go 100 nt

- Elongation is *slow* compared to length of many genes
  - for "gene" 50-100 kb long, can take >30 min after initiation to finish 1 RNA molecule

Determinants of elongation

- Initiation of elongation:
  - NELF (negative elongation factor) holds Pol II 20-50nt from initiation site until released (phosphorylation). Released by pause-release factor (p-TEFb)
  - Certain TF-s regulate elongation in promoter-proximal regions
- Continued reading:
  - Initiation factors released, transcription elongation factors (TEFs) associate with Pol II
- Chromatin (discussed later) – barrier - chromatin factors (chaperons, remodelers, histone/DNA modifying enzymes)

Steps of RNA processing



Coupling transcription and RNA processing

C-terminal tail of RNA pol II acts as a docking site and transport vehicle for RNA processing proteins to keep them close to their substrate as needed during transcription

Phosphate addition to particular amino acids in the polymerase tail controls which factors can ride (and are active) at different stages

High local concentration and dependence on CTD phosphorylation ensures specificity and efficiency of processing (i.e. only Pol II transcripts are capped, polyadenylated, correct splice sites are found etc.)

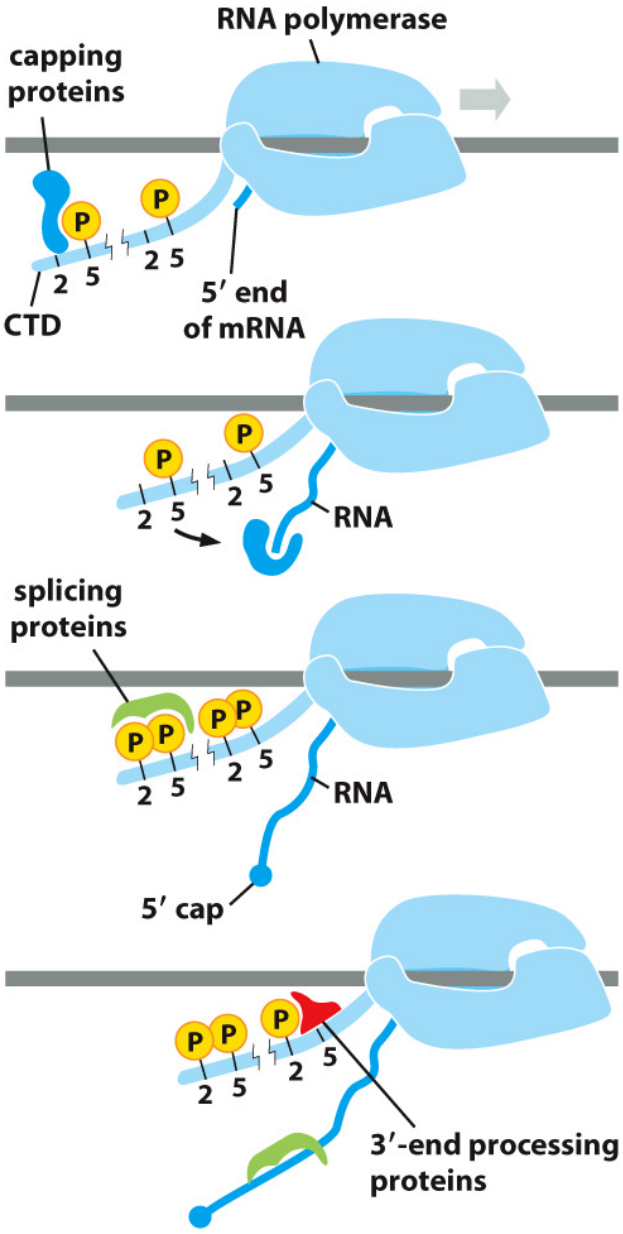
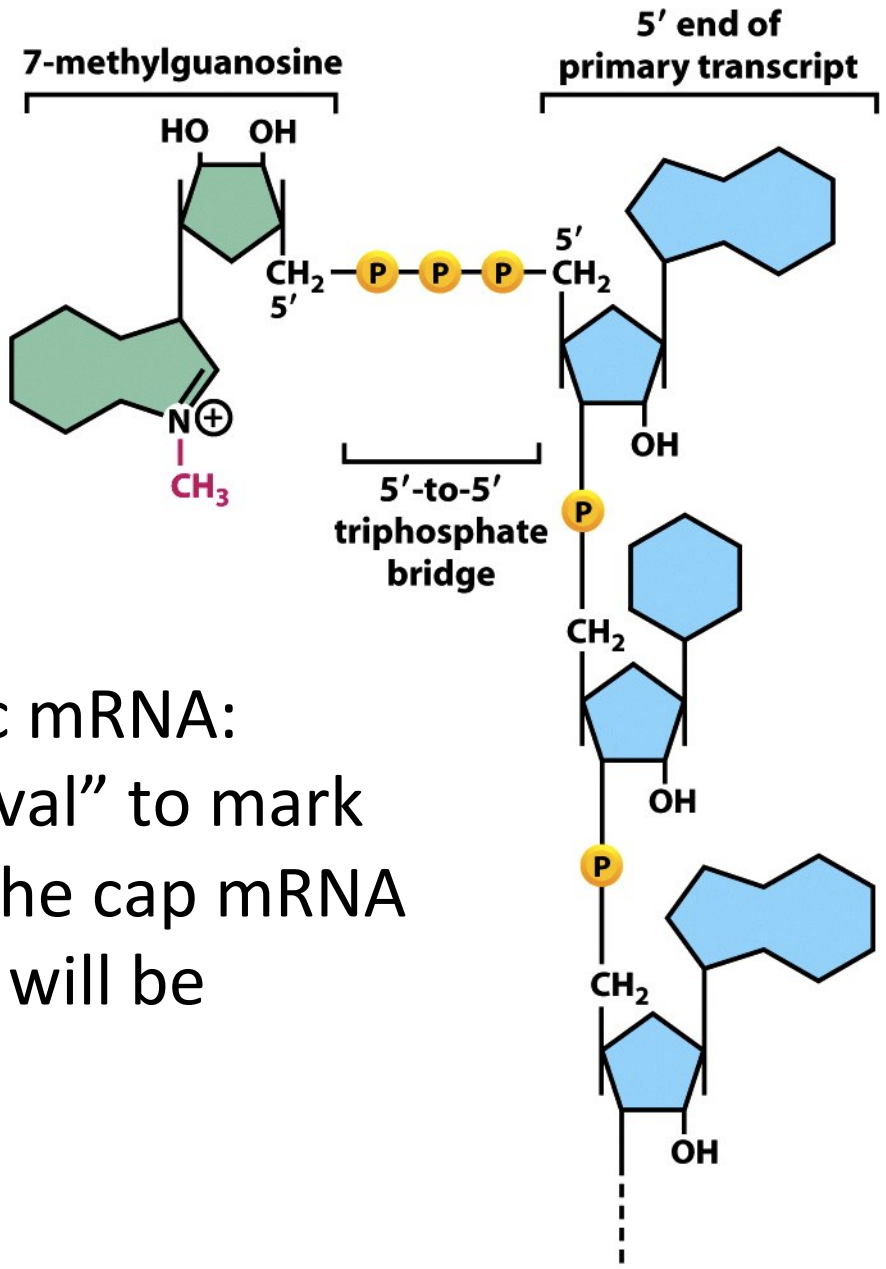


Figure 6-22 Molecular Biology of the Cell 6e (© Garland Science 2015)

Capping of Pol II transcripts

A “backwards”, methylated guanosine residue is used to cap mRNA 5’-triphophate ends shortly after they leave the polymerase tunnel



Quality controls for eukaryotic mRNA: nontemplated “seals of approval” to mark non-degraded ends (without the cap mRNA can not leave the nucleus and will be degraded)

Figure 6-22b *Molecular Biology of the Cell* (© Garland Science 2008)

mRNA splicing

mRNAs typically include multiple exons with introns spliced out at RNA level during synthesis

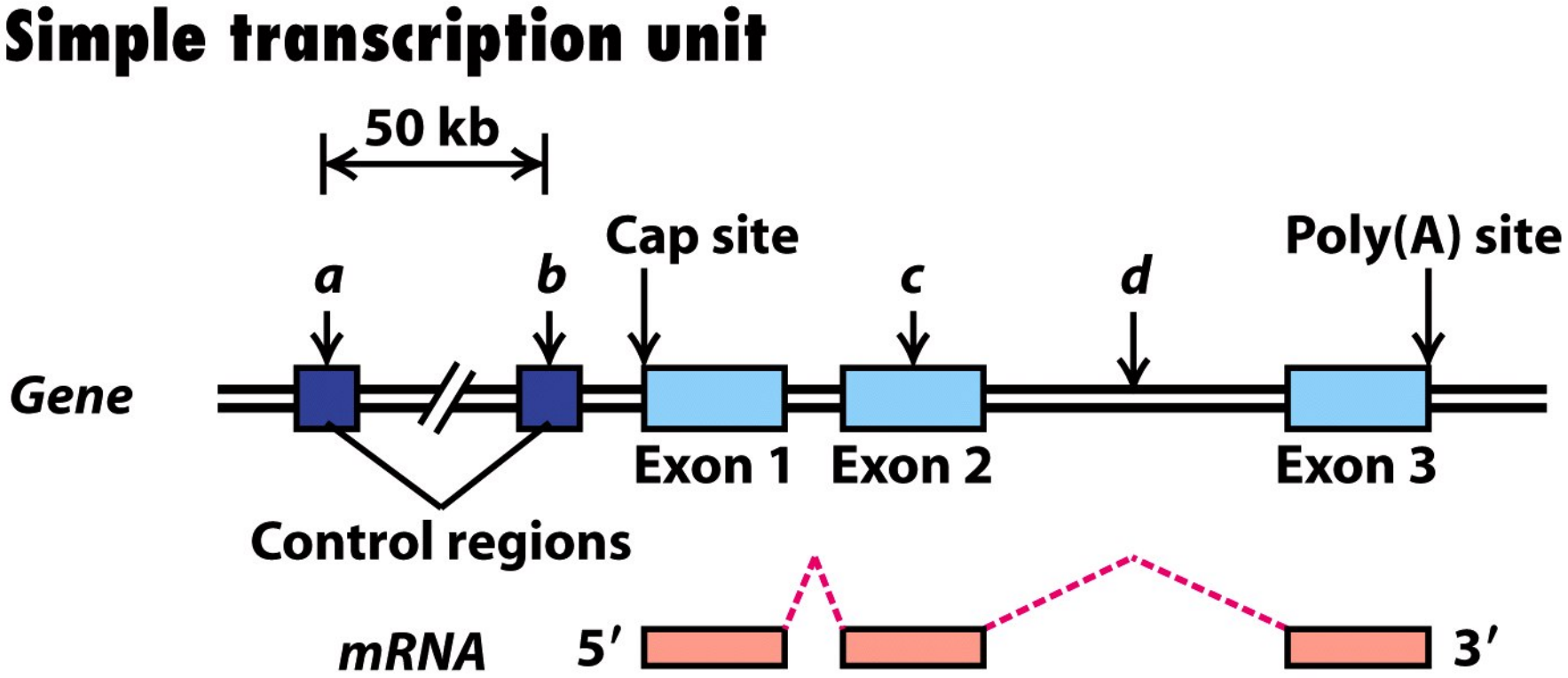


Figure 6-3a
Molecular Cell Biology, Sixth Edition

Splicing: converting genomic primary transcript into mature mRNA – more covalent modification and interpretation of cryptic processing information

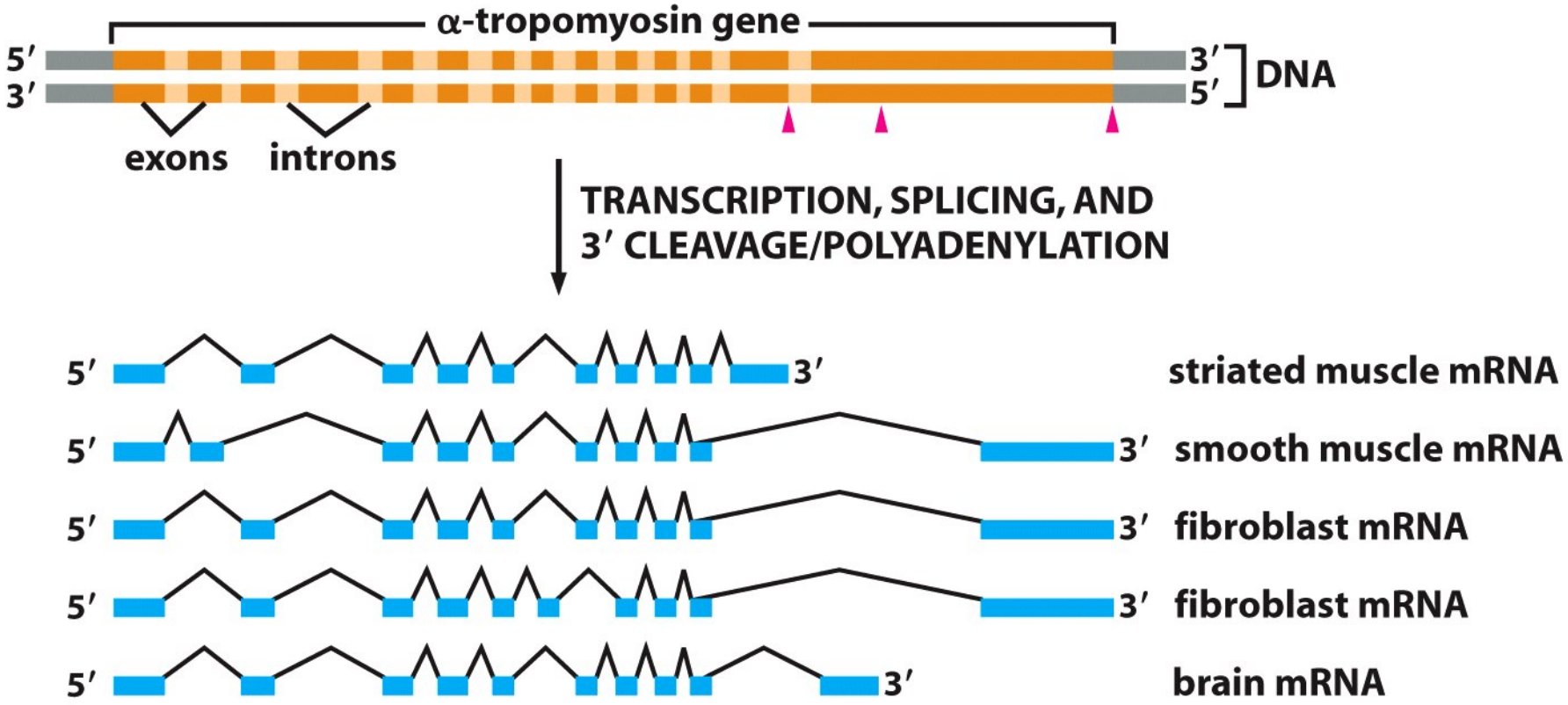
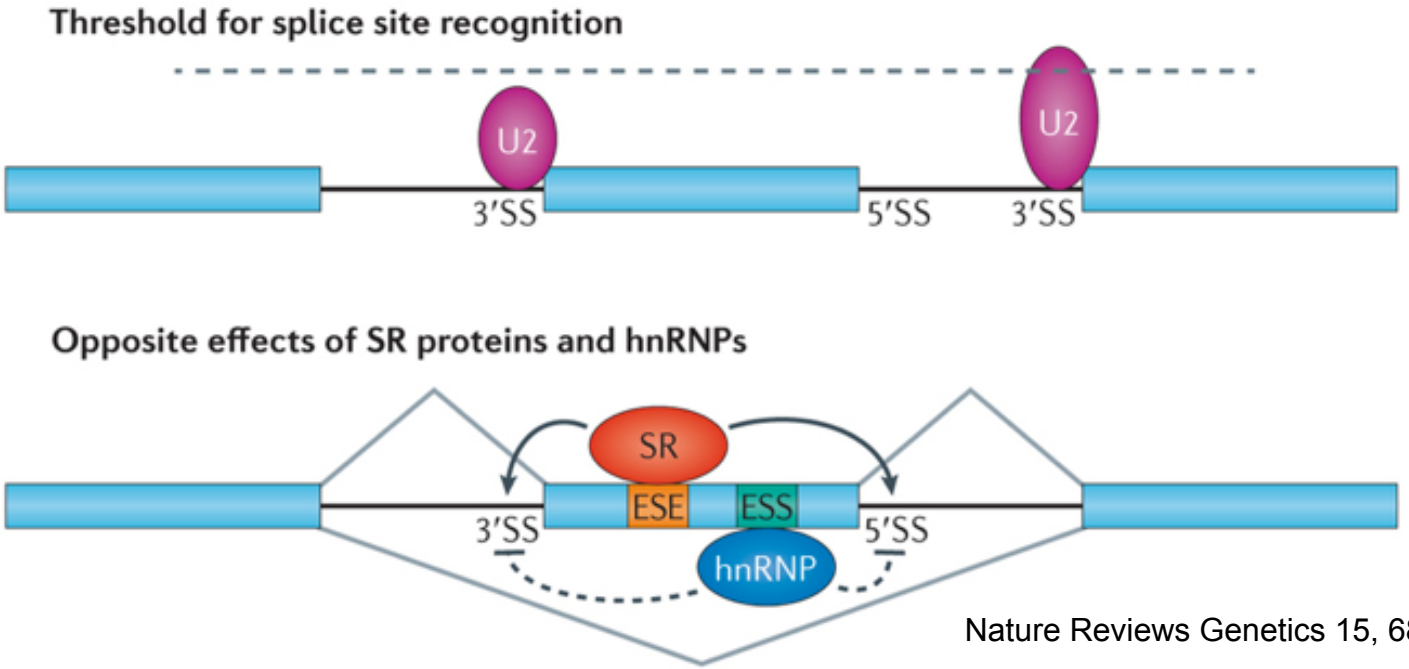


Figure 6-27 *Molecular Biology of the Cell* (© Garland Science 2008)

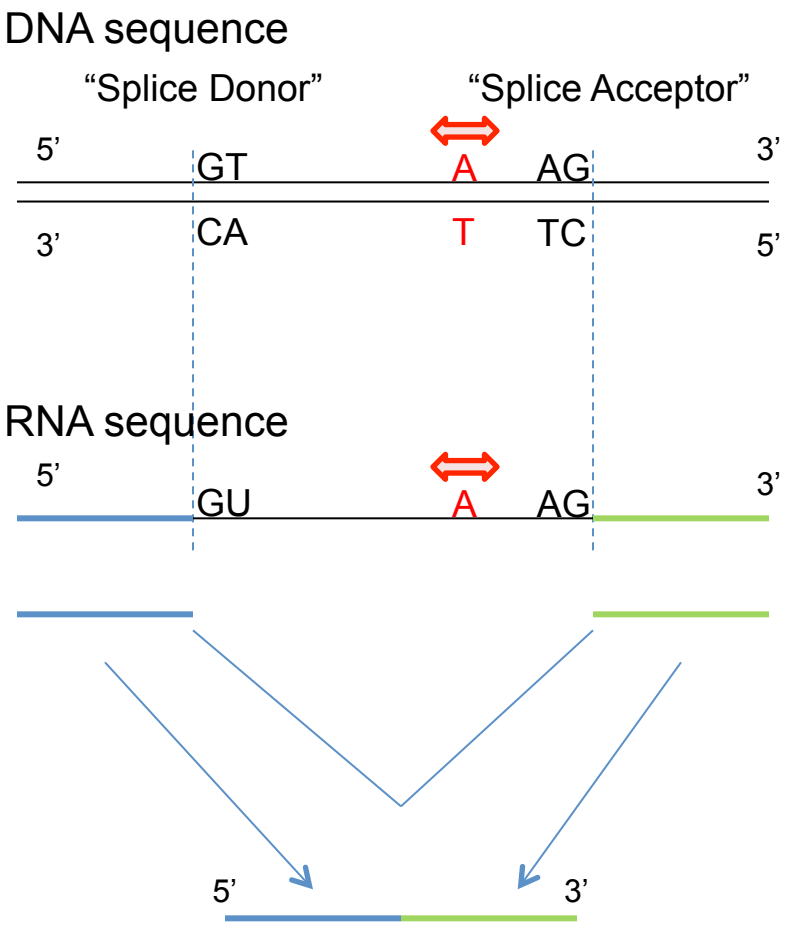
Alternative splicing

- Most genes (~93%) in higher eukaryotes have multiple splice variants
- Alternative splicing results in different proteins from one transcription unit – Complexity!
- Often alternative splicing is controlled by developmental and differentiation signals (through factors that are differentially expressed in the given cell) or by environmental stimuli



Nature Reviews Genetics 15, 689–701 (2014)

Key features of splicing



Bioinformatic challenge: Cells are still better than "we" are at recognizing the right "A" somewhere before an "AG" to define the end of an intron

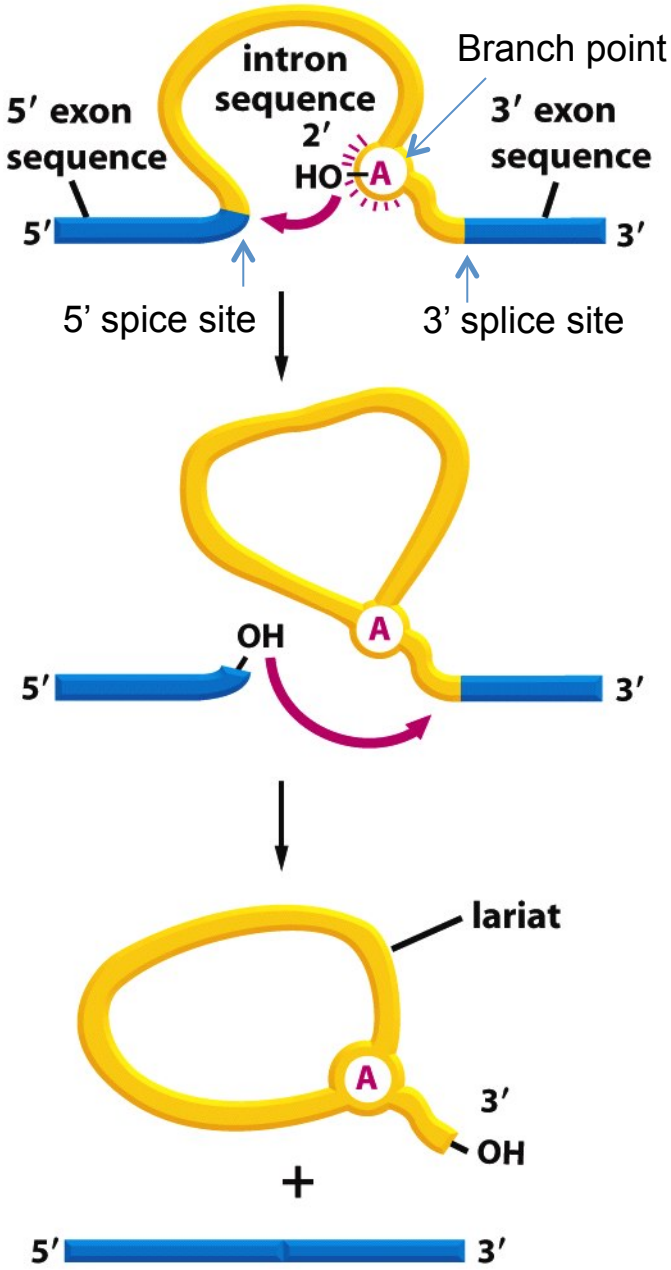
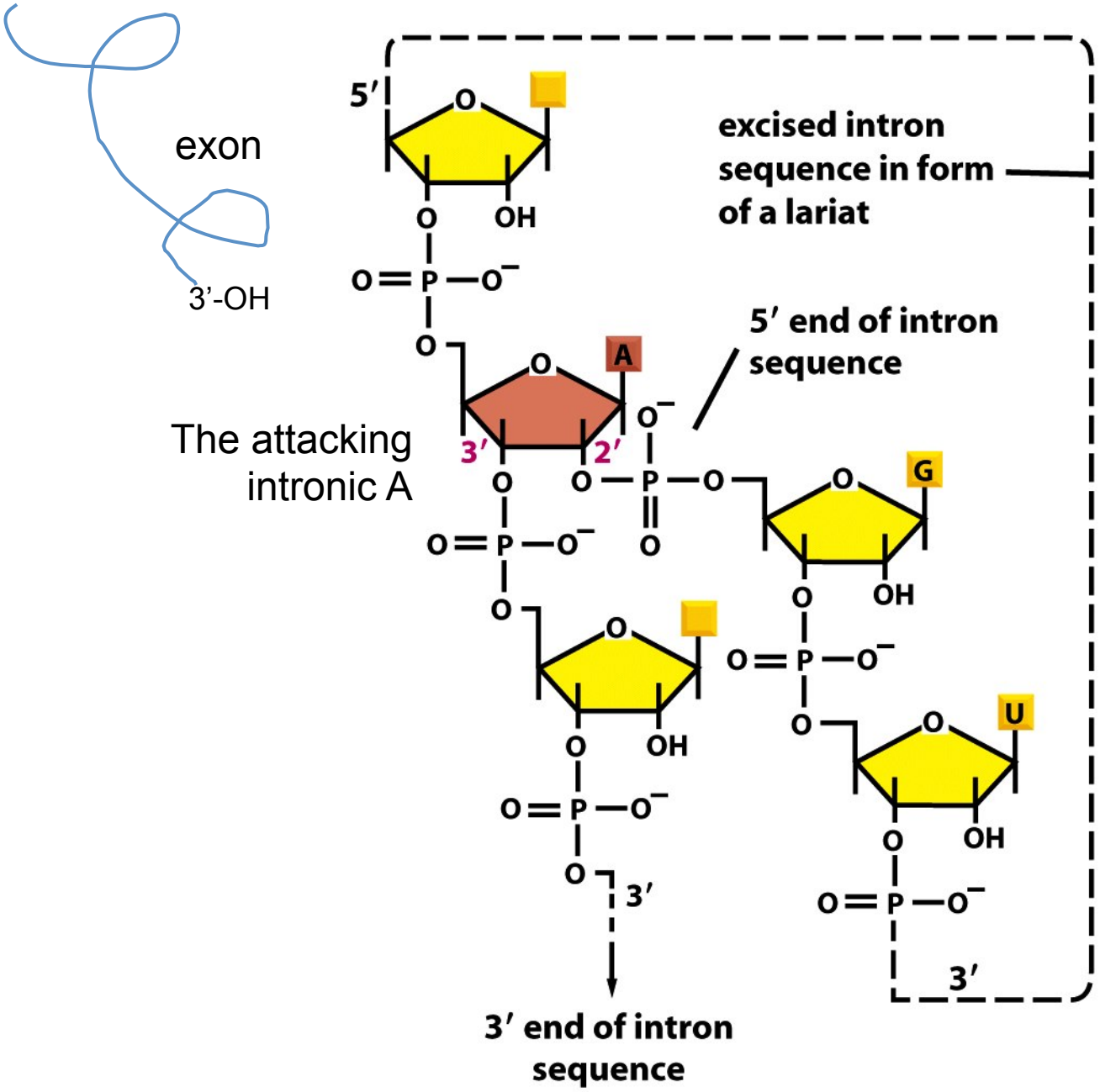


Figure 6-26a *Molecular Biology of the Cell* (© Garland Science 2008)

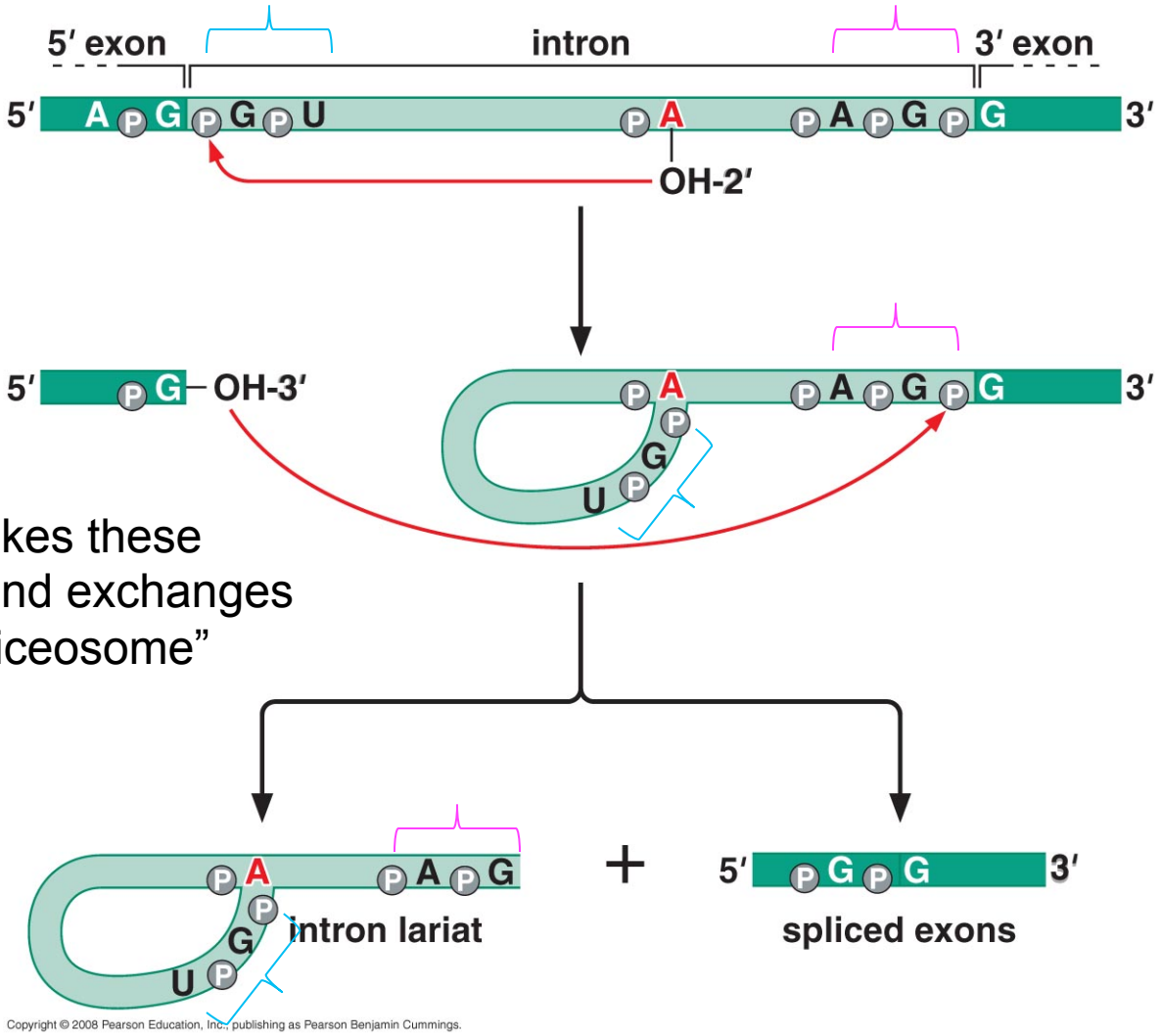


The splicing reaction channels the action of the same kind of 2'-OH attack chemistry that can also cause RNA to self-destruct…

but with a different target

Figure 6-26b *Molecular Biology of the Cell* (© Garland Science 2008)

Highly controlled attack followed by highly controlled "retaliation" causes cleavage and joining at low energy cost



Machinery that makes these phosphodiester bond exchanges specific is the "Spliceosome" (discussed later)

The final product…. Or one form of it

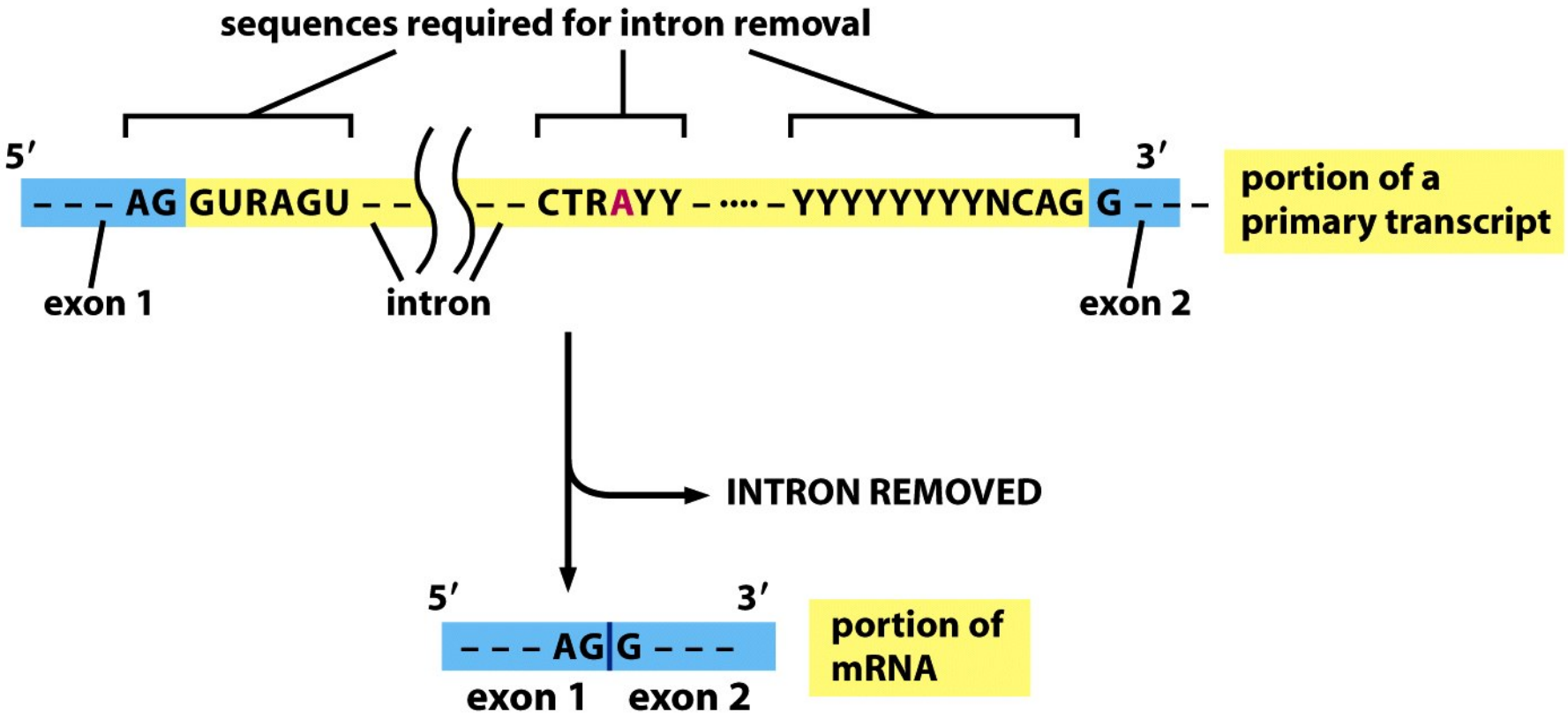
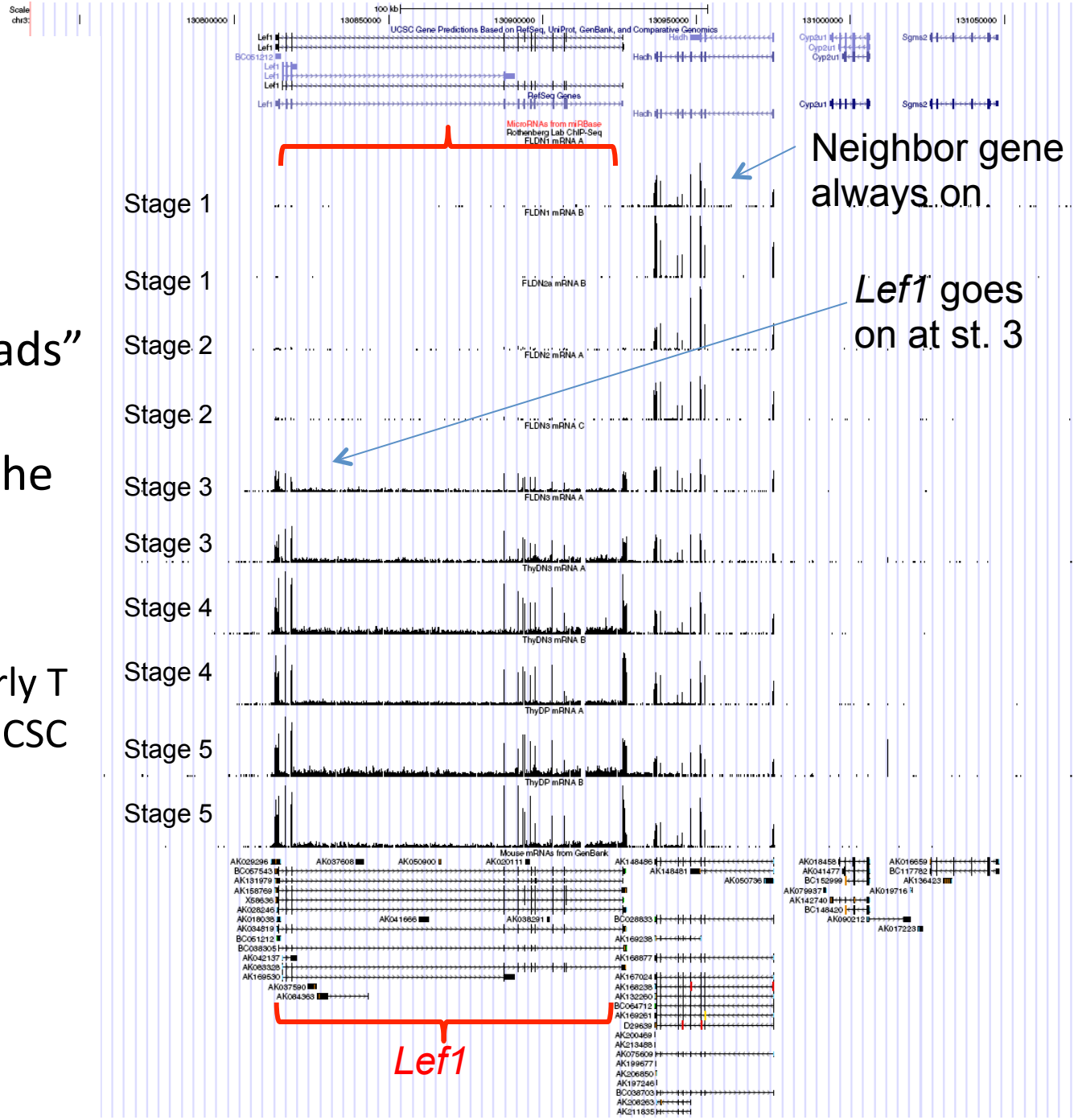


Figure 6-28 *Molecular Biology of the Cell* (© Garland Science 2008)

High-throughput sequencing of bulk cDNA can generate millions of short “reads” that can be aligned computationally to the genome… revealing expressed exons

(activation of *Lef1*, in early T cells: RNA-seq data on UCSC genome browser)

(J. A. Zhang, A. Mortazavi, B. Williams, B.J. Wold, E.V. Rothenberg, 2012, Cell)

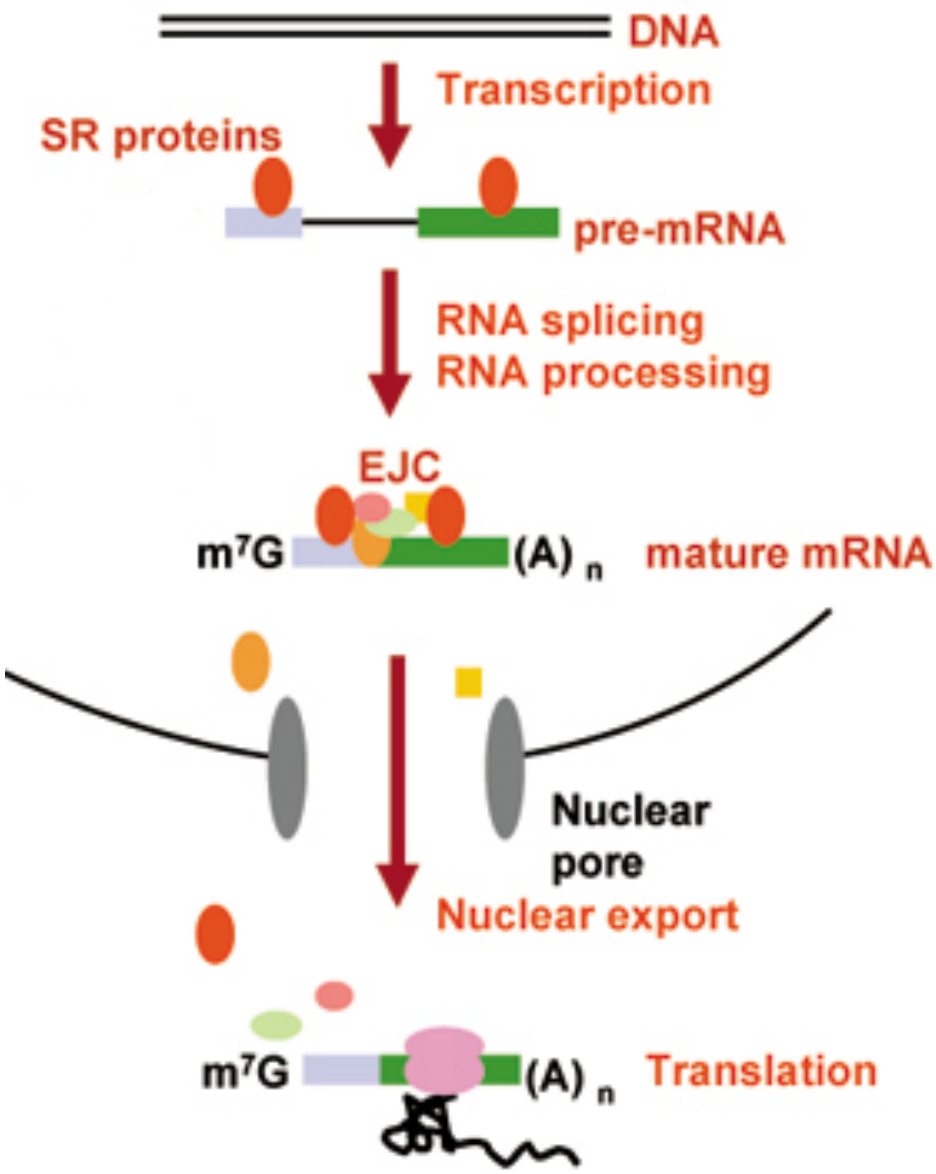


Splicing affects RNA fate

Experiment:
intronless reporters express less efficiently

Reason: Exon Junction Complex (EJC): factors deposited during splicing that stay associated with the transcript until translation.

EJC is required for nuclear export and translation



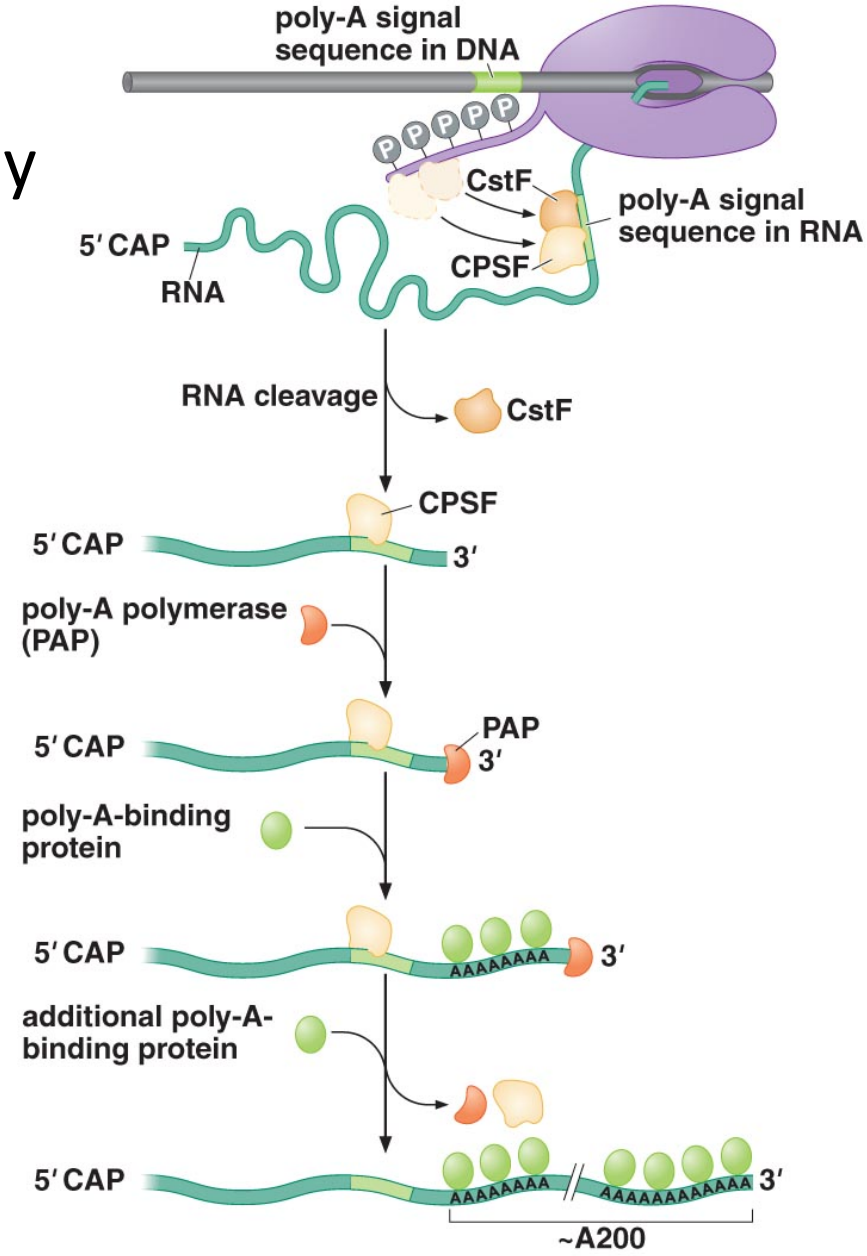
http://www.tmd.ac.jp/mri/mri-end/en/research.html

Polyadenylation of Pol II transcripts

Most mRNAs are "finished off" by addition of a NONTEMPLATED stretch of poly(A) at 3' ends

Capped, poly(A)+ RNA = "Good" RNA
(histone genes are not polyadenylated)

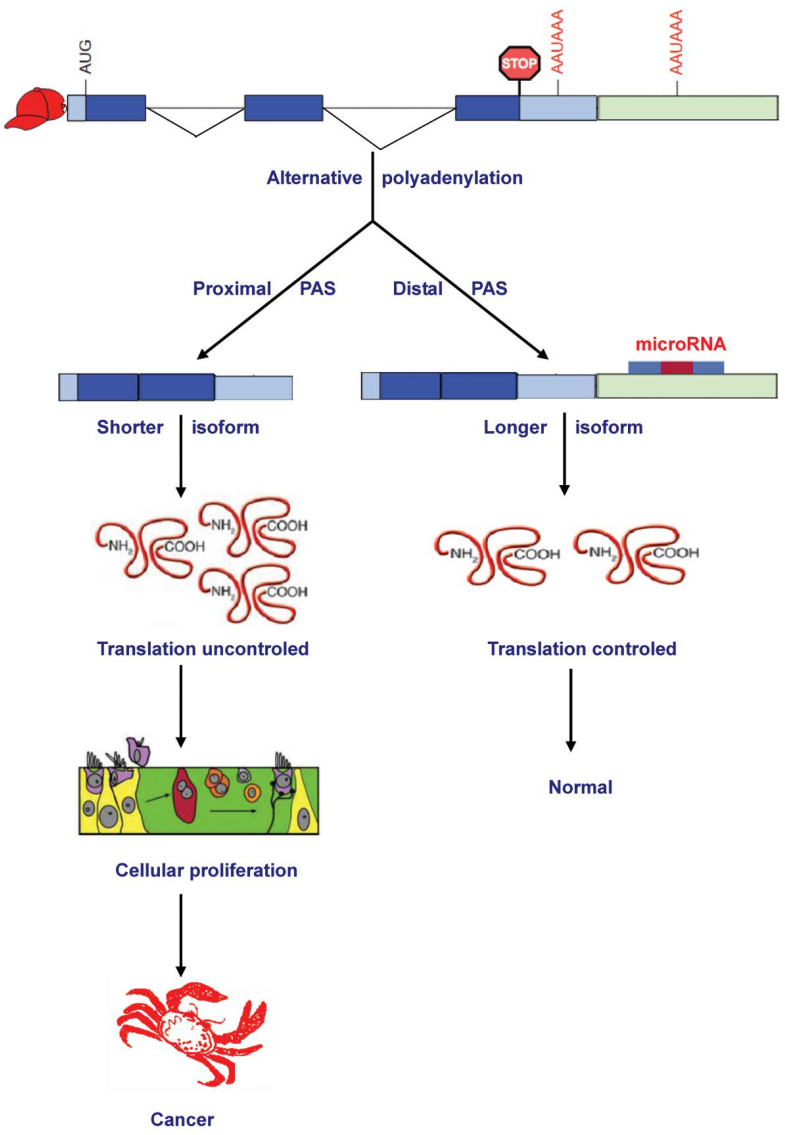
This will be CRUCIAL in control of protein synthesis!



Alternative polyadenylation

- most genes have numerous alternative polyA sites
- regulated during differentiation: undifferentiated and cancer cells use proximal termination sites

Lack of regulatory signals (e.g. miRNA binding sites) present in longer isoforms

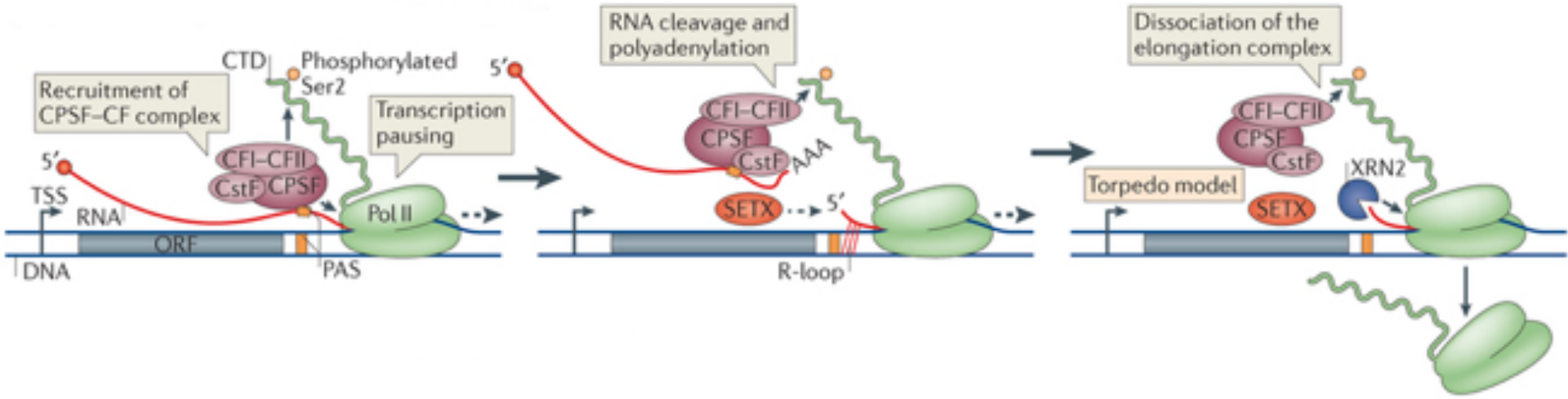


(www.spandidos-publications.com/ijo/43/4/995)

Transcription termination

RNA polymerase does not stop transcribing immediately after reaching the polyA site.

Torpedo model of transcriptional termination



Nature Reviews Molecular Cell Biology 16, 190–202 (2015)

Sequence-based detection of RNA expression

- Net concentration of sequences from a given gene in a given cell population can be measured by hybridization
  - On “Northern” blots or microarrays
  - To PCR primers, for measurement by amplification in reverse transcriptase-PCR (conventional or real-time)

- Newest techniques: massive high-throughput sequencing of cDNA from an RNA sample to determine all transcript levels at once (RNA-seq)
- Especially valuable with complete genomic sequence assembly
- Advantages:
  - See everything that is present in RNA, known or not
  - All molecules are on same “natural scale”

Common methods of measuring RNA expression

Oligonucleotides each detects one part of a target RNA

Once mRNA sequences are known, probes for mRNAs of interest can be designed and synthesized… (commercially available)

use hybridization to measure levels of multiple RNA species in biological samples at once

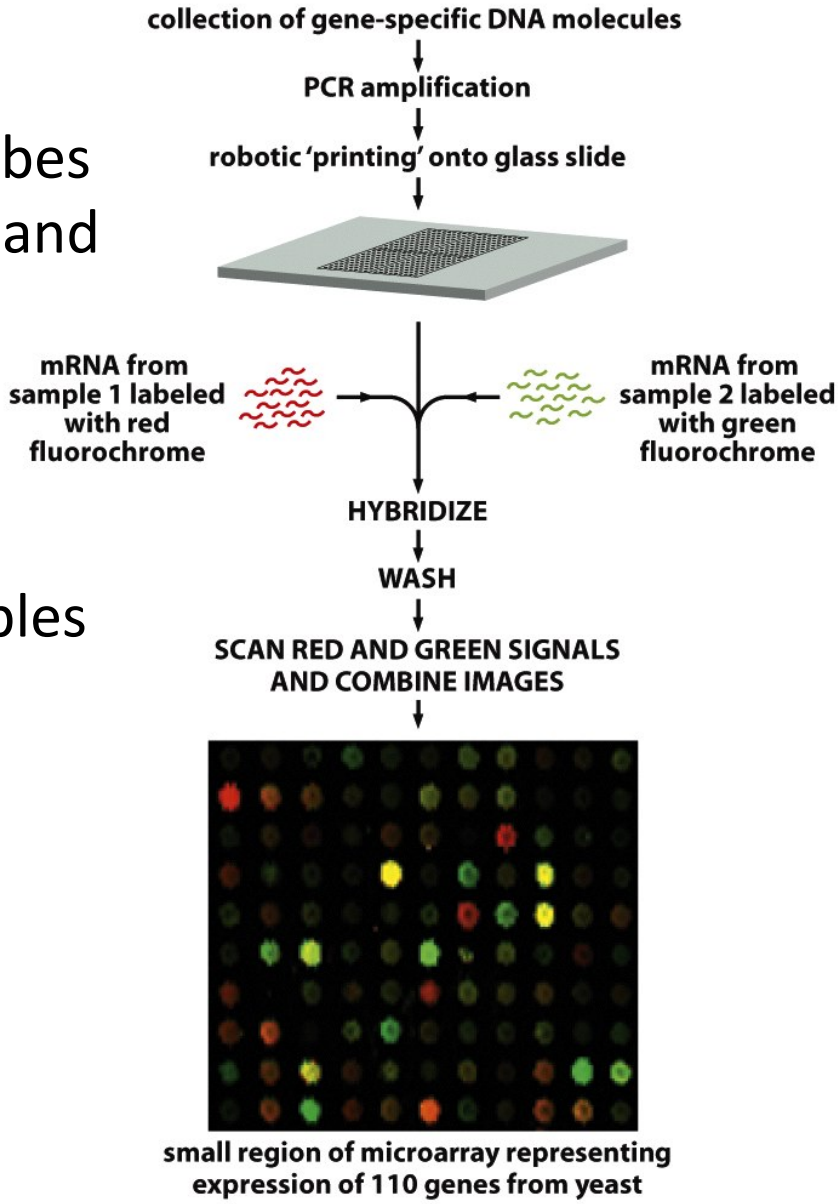


Figure 8-73 *Molecular Biology of the Cell* (© Garland Science 2008)

But…RNA-seq and microarray hybridization may not tell you which isoforms are made… or even if they exist

Structures come from cDNA clones

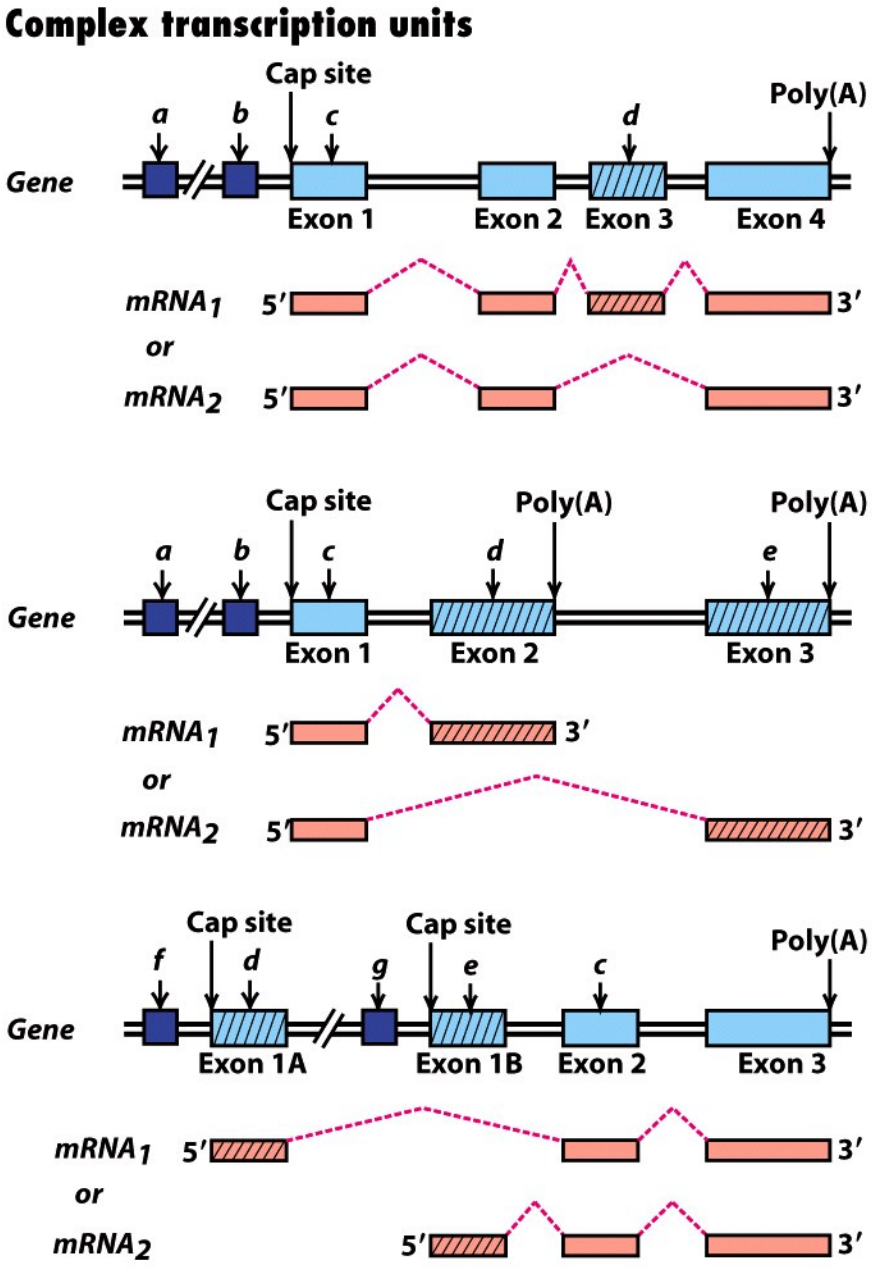


Figure 6-3b
Molecular Cell Biology, Sixth Edition

Analyzing expression of RNAs of different structures

- Cloned, long probe bridging feature to assay
  - ssDNA form
  - RNA form can also be made
- Hybridize with target RNA
- Measure not only *amounts* but also *length(s)* of hybridized probe fragments
- Can use position of label in probe to get more info…
  - End labeling
  - Uniform labeling

Distinguishing single-stranded from double stranded

- Crudest: present unlabeled form stuck to a solid support (filter, tissue section, etc.) and then ask whether labeled probe stays bound after hybridization

- If you just care about duplex: single-strand specific nucleases degrade away the unhybridized part
  - S1 degrades ss cDNA probes
  - RNase degrades ss RNA probes

- There are also ways to recover both duplex and single-stranded forms of a DNA probe:
  - chromatography on hydroxyapatite (low vs. hi phosphate buffer)

Classic intron mapping with a single-stranded DNA probe

Shows presence of sequences from both exons in RNA and lengths of sequence expressed from both

This hybridization to labeled genomic DNA does not directly show that these sequences are linked in RNA

But hybridization of cloned cDNA to genomic DNA – the reverse experiment -- will show that they were linked in RNA! (Consider also how you could use PCR…)

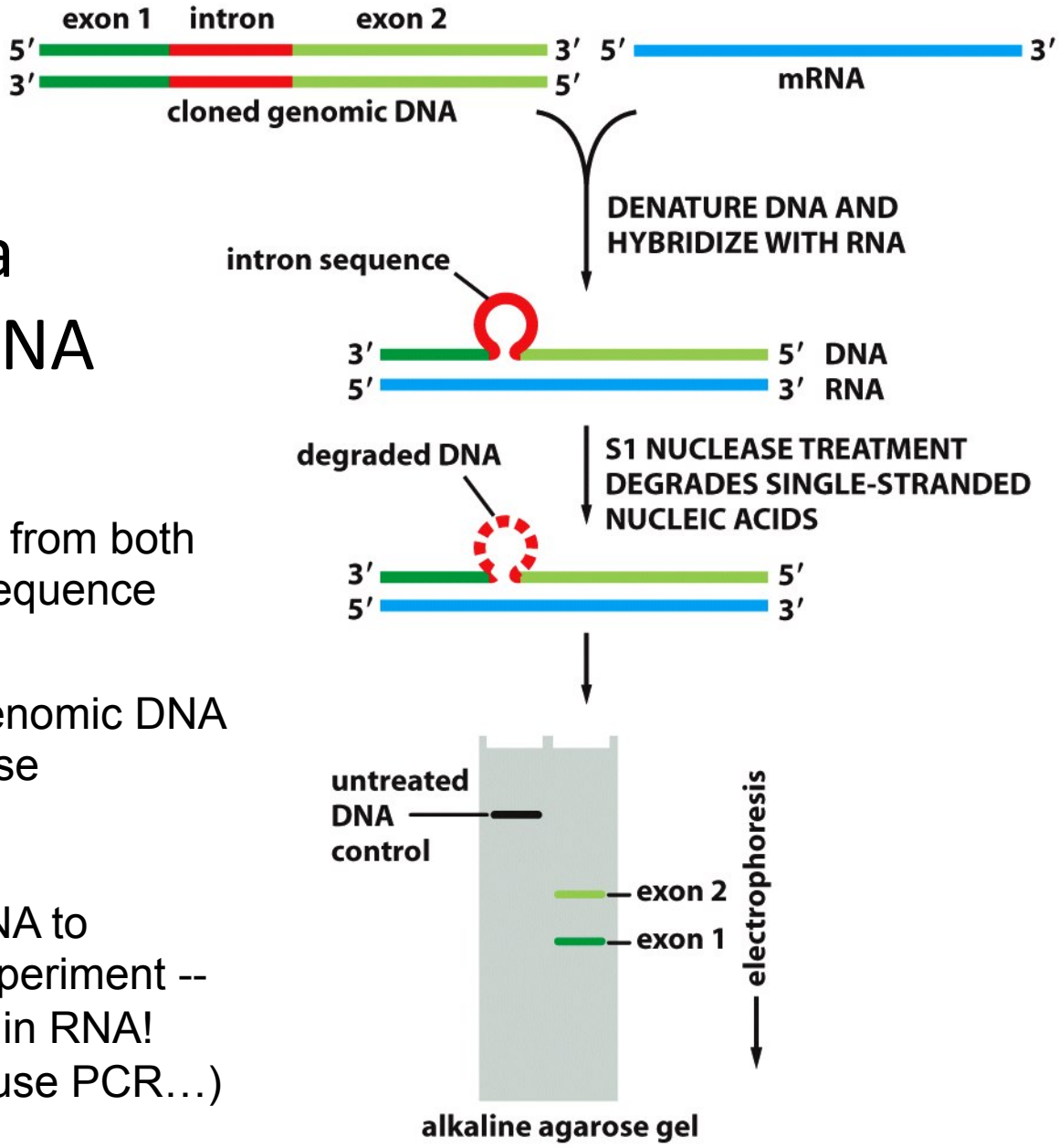


Figure 8-37 *Molecular Biology of the Cell* (© Garland Science 2008)