Liam.Keegan@ceitec.muni.cz

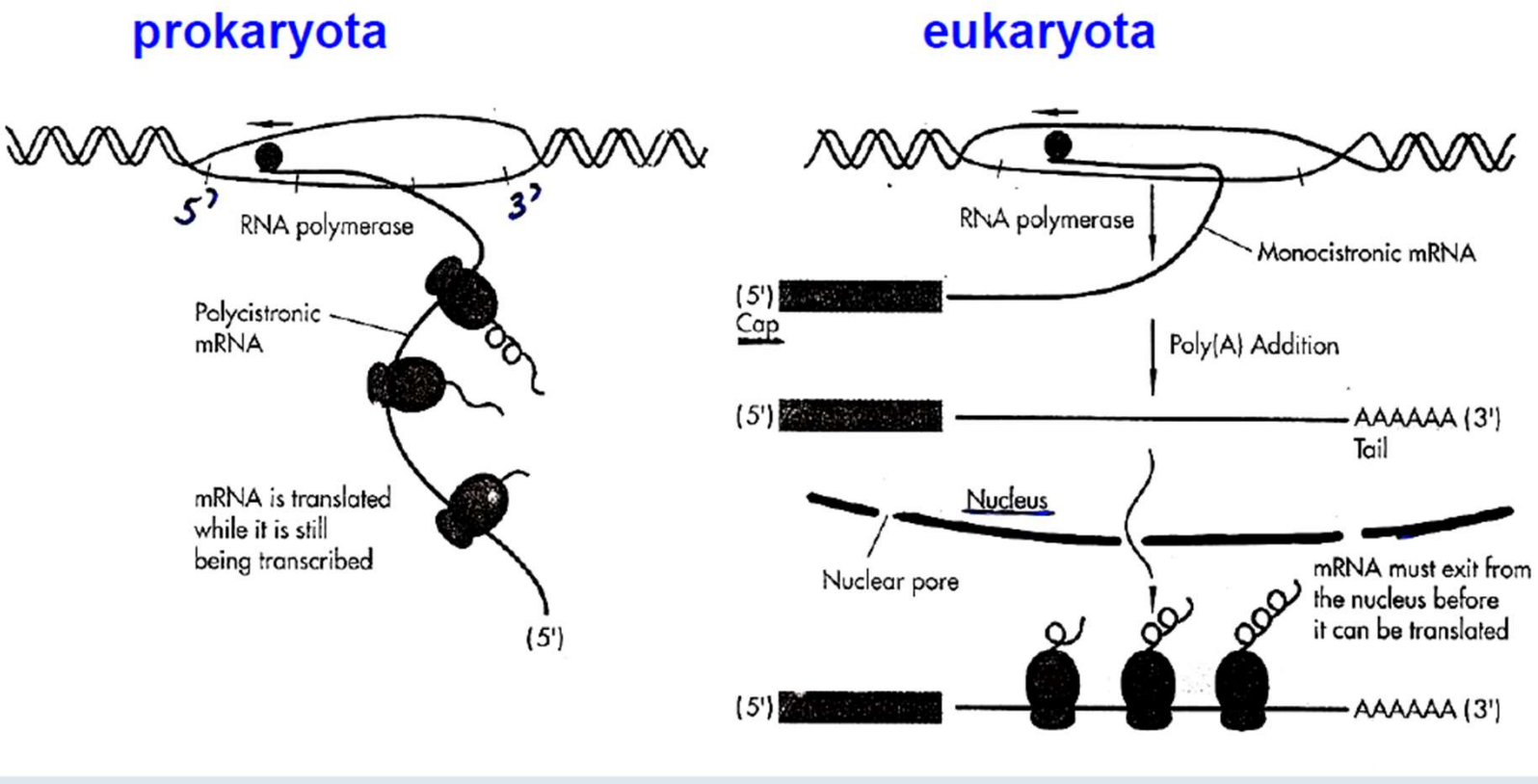
Lewin, GENES XI

Chapter 19, 26
Bacterial transcription, Operons, prokaryotic gene regulation and eukaryotic transcription

Lewin's GENES XII (PDF)
https://pdfroom.com/books/lewins-genes-xii/Wx5aDYKl2BJ

biotecher.ir
https://biotecher.ir/wp-content/uploads/2018/06/bio...

Transcription and translation occur together and simultaneously in bacteria but separately in eukaryotes



26.2 Operons are Structural Gene Clusters that Are Coordinately Controlled

- Genes coding for proteins that function in the same pathway may be located adjacent to one another and controlled as a single unit that is transcribed into a **polycistronic mRNA**.

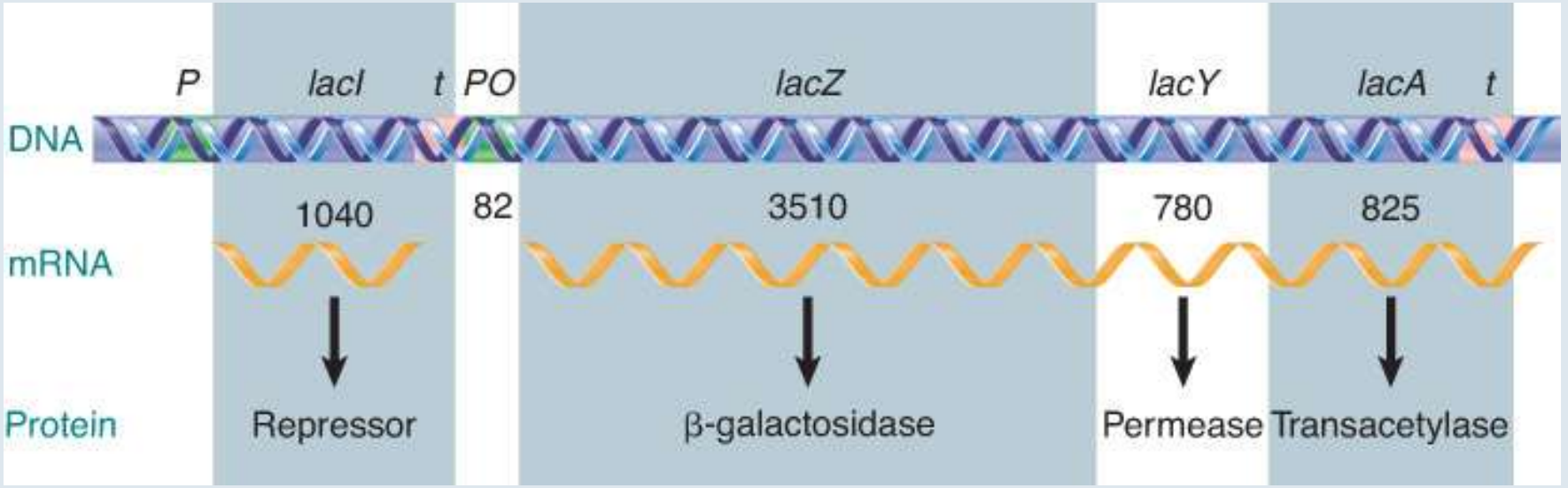
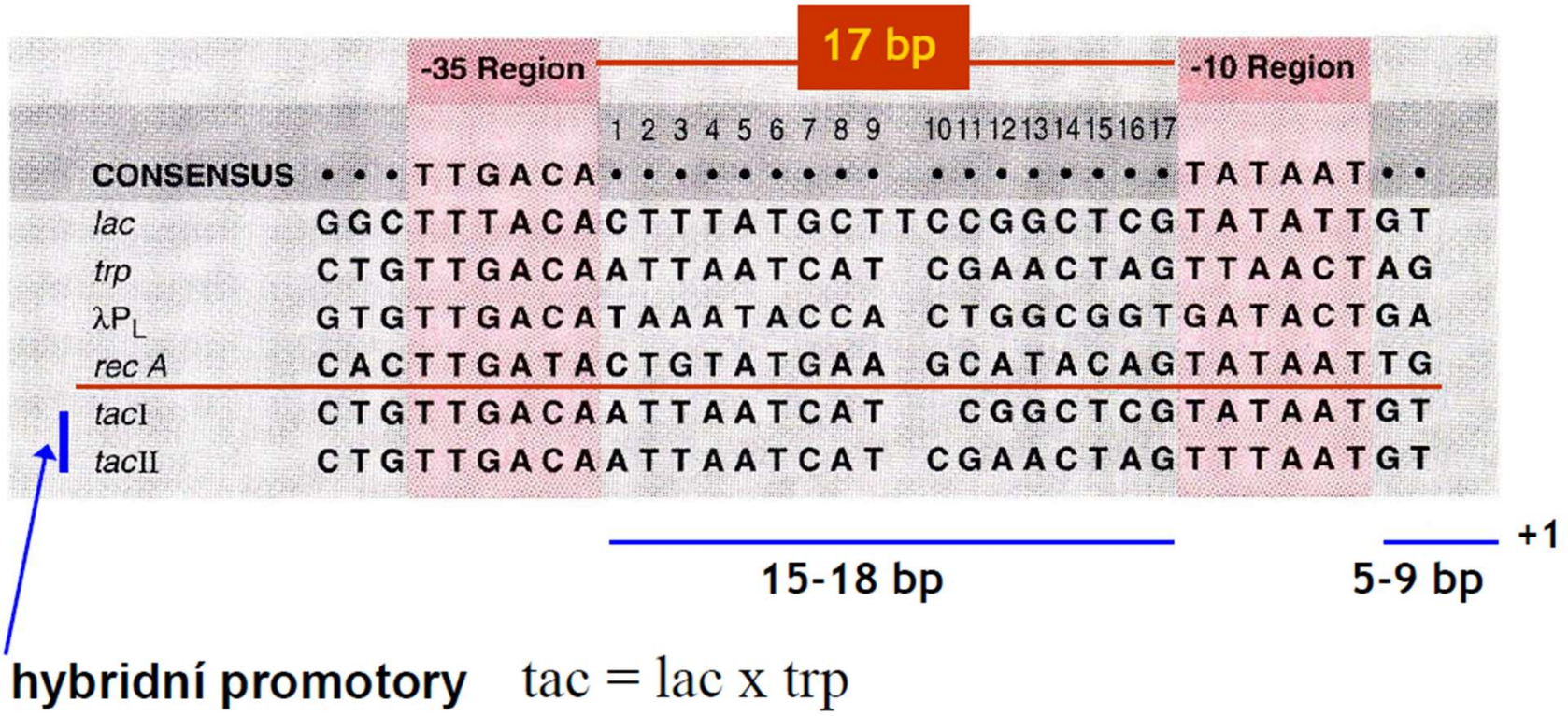


Figure 26.05: The lac operon occupies ~6000 bp of DNA.

E. coli promoters have -35 TTGACA and -10 TATAAT consensus sequences



hybridní promotory tac = lac x trp

RNA polymerase melts DNA in a bubble and synthesizes RNA 5’ to 3’ along the template DNA strand

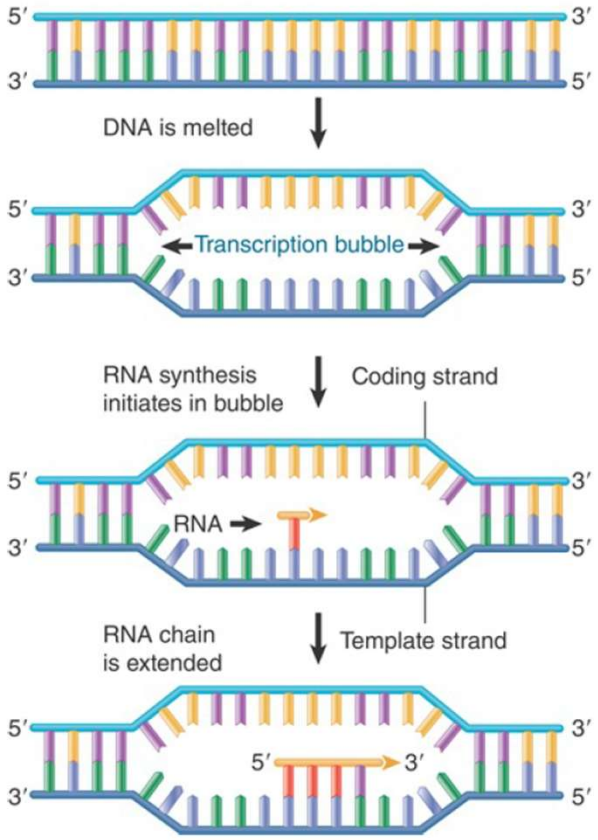


FIGURE 17.3 DNA strands separate to form a transcription bubble. RNA is synthesized by complementary base pairing with one of the DNA strands.

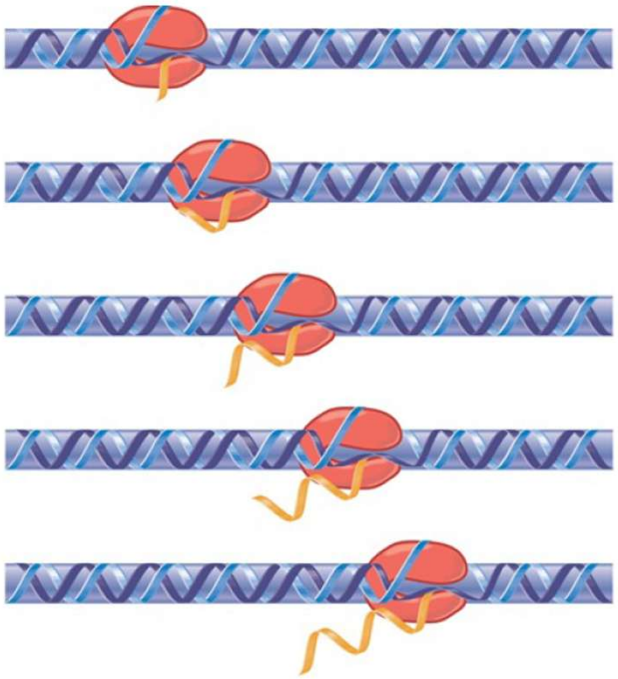


FIGURE 17.4 Transcription takes place in a bubble, in which RNA is synthesized by base pairing with one strand of DNA in the transiently unwound region. As the bubble progresses, the DNA duplex reforms behind it, displacing the RNA in the form of a single polynucleotide chain.

The transcription cycle: RNA polymerase initiates, elongates and terminates RNA transcripts

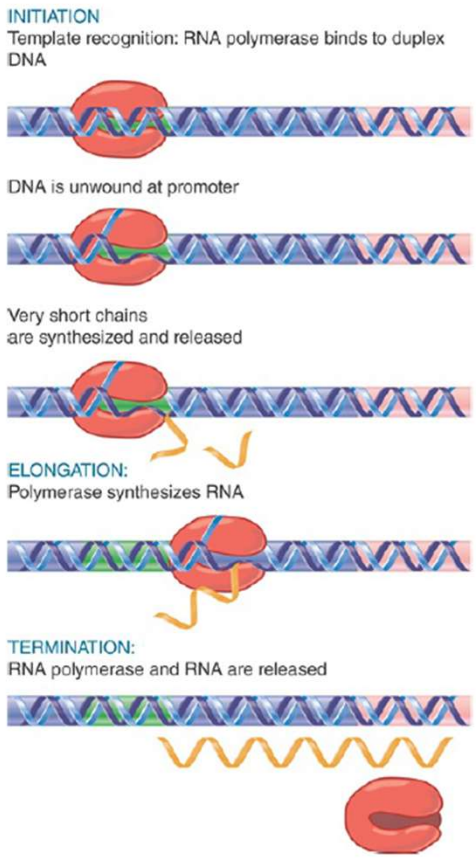


FIGURE 17.6 Transcription has three stages: The enzyme binds to the promoter and melts DNA and remains stationary during initiation; moves along the template during elongation; and dissociates at termination.

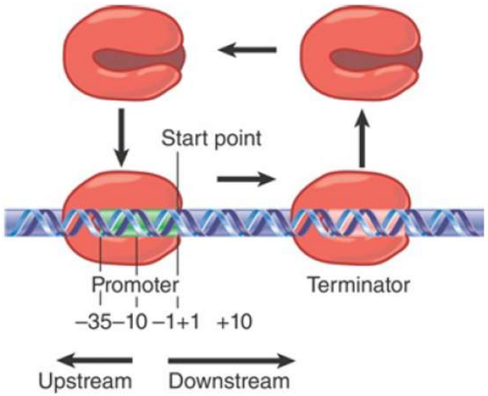
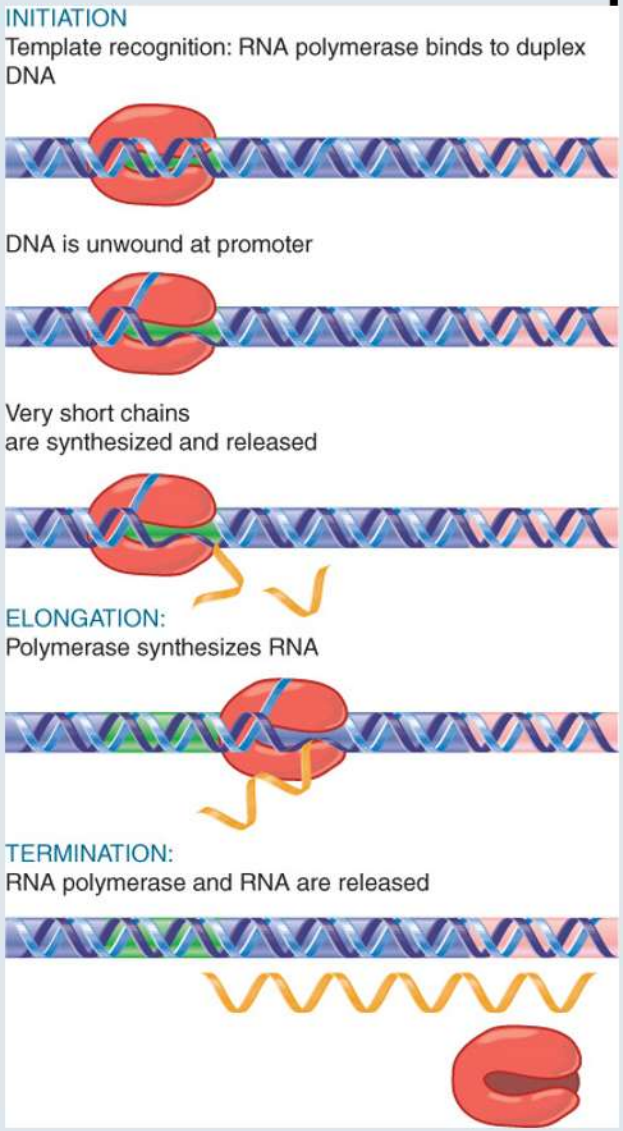


FIGURE 17.2 A transcription unit is a sequence of DNA transcribed into a single RNA, starting at the promoter and ending at the terminator.

19.3 The Transcription Reaction Has Three Stages



- RNA polymerase binds to a promoter site on DNA to form a **closed complex**.
- RNA polymerase initiates transcription (**initiation**) after opening the DNA duplex to form a transcription bubble (the **open complex**).

Figure 19.06: Transcription has three stages.

E. coli RNA polymerase

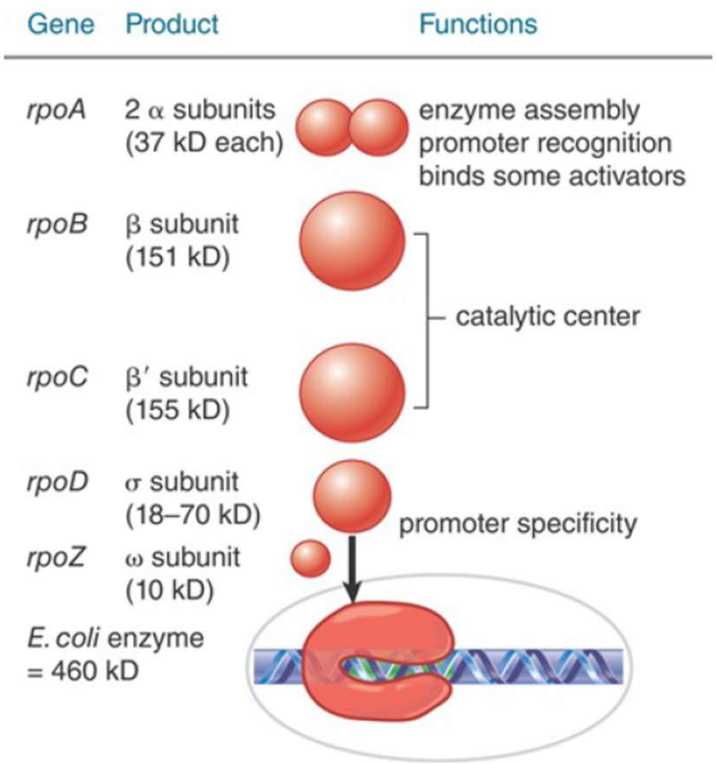


FIGURE 17.7 Eubacterial RNA polymerases have five types of subunits: α, β, β′, and ω have rather constant sizes in different bacterial species, but σ varies more widely.

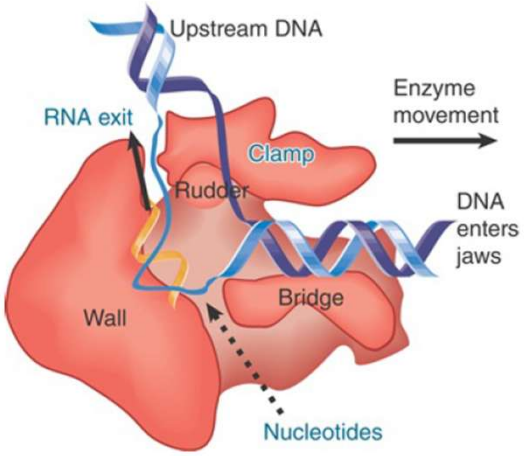


FIGURE 17.25 DNA is forced to make a turn at the active site by a wall of protein. Nucleotides may enter the active site through a pore in the protein.

19.4 Bacterial RNA Polymerase Consists of Multiple Subunits

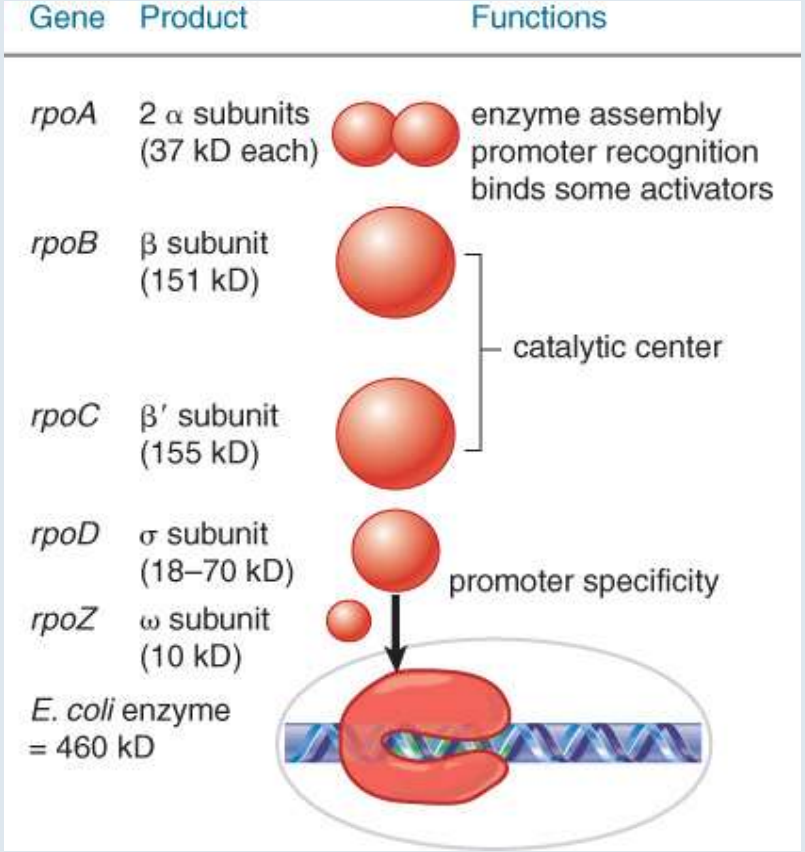


Figure 19.07: Eubacterial RNA polymerases have five types of subunits.

- **holoenzyme –** The RNA polymerase form that is competent to initiate transcription. It consists of the five subunits of the **core enzyme** and **σ factor**.
- Bacterial RNA core polymerases are ~400 kD multisubunit complexes with the general structure $\alpha_2\beta\beta'\omega$.

19.4 Bacterial RNA Polymerase Consists of Multiple Subunits

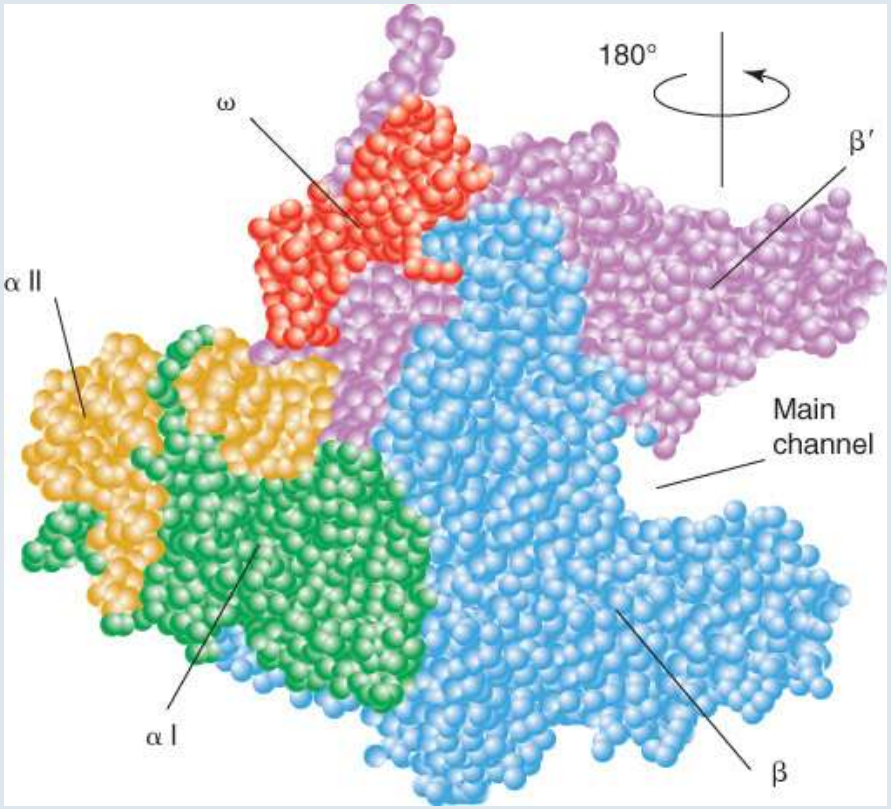


Figure 19.08: The upstream face of the core RNA polymerase, illustrating the 'crabclaw' shape of the enzyme.

Adapted from K. M. Geszvain and R. Landick (ed. N. P. Higgins). The Bacterial Chromosome. American Society for Microbiology, 2004.

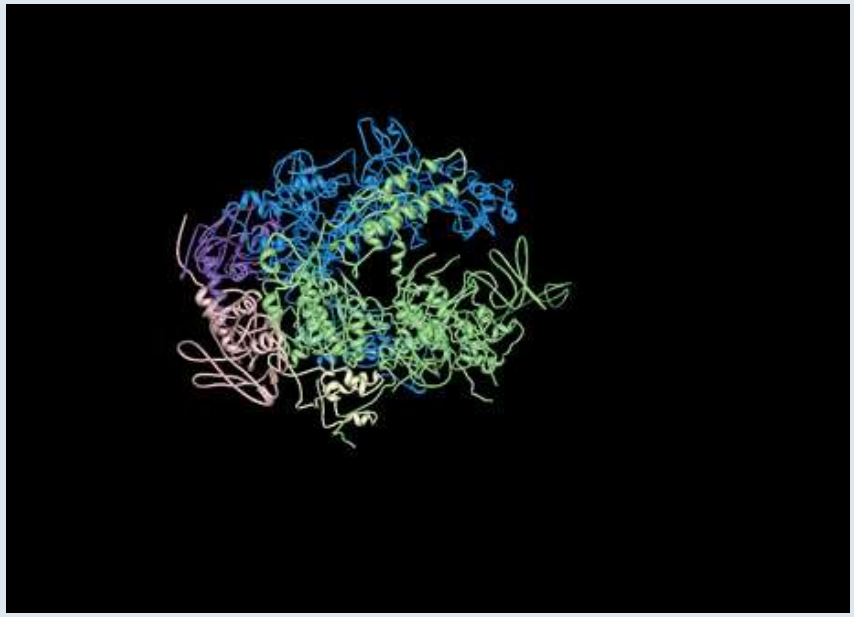


Figure 19.09: The structure of the RNA polymerase core enzyme for the bacterium Thermus aquaticus.

Structure from Protein Data Bank 1HQM. L. Minakhin, et al., Proc. Natl. Acad. Sci. USA 98 (2001): 892-897.

19.4 Bacterial RNA Polymerase Consists of Multiple Subunits

- Catalysis derives from the β and β′ subunits.
- **CTD (C-terminal domain) –** The domain of RNA polymerase that is involved in stimulating transcription by contact with regulatory proteins.

19.5 RNA Polymerase Holoenzyme Consists of the Core Enzyme and Sigma Factor

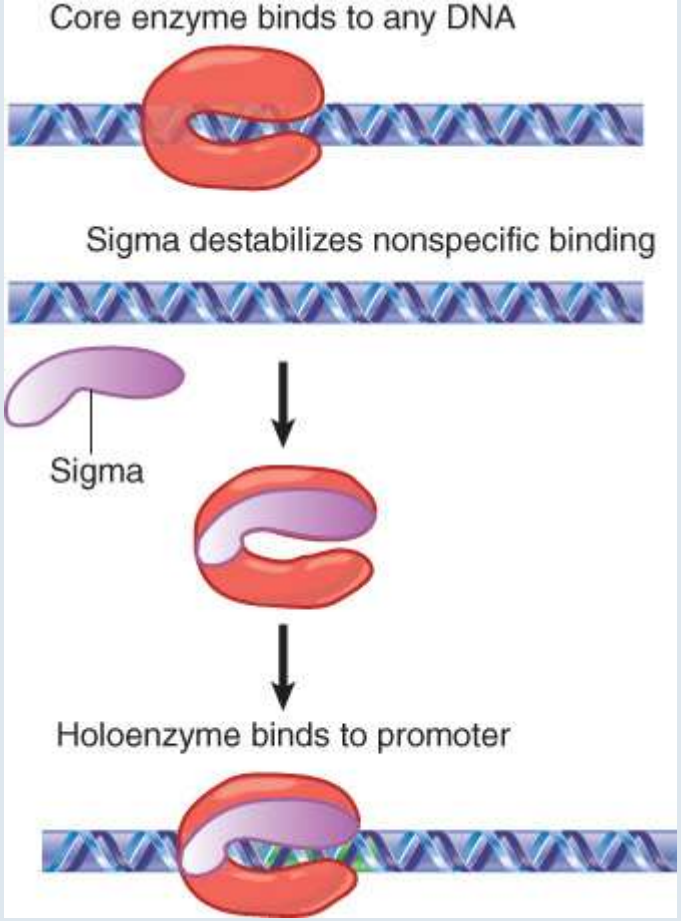
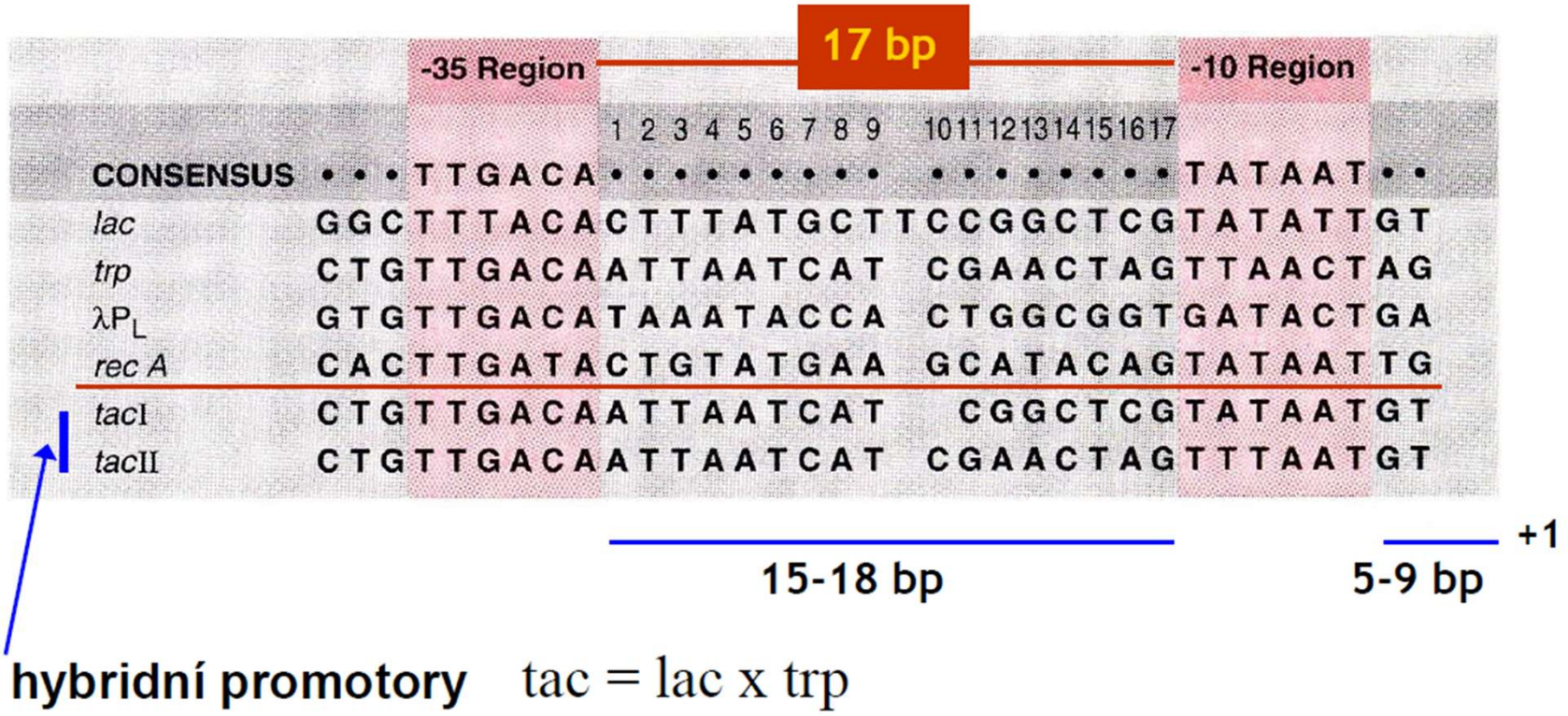


Figure 19.10: Core enzyme binds indiscriminately to any DNA.

- Bacterial RNA polymerase can be divided into the $\alpha_2\beta\beta'\omega$ core enzyme that catalyzes transcription and the σ subunit that is required only for initiation.
- Sigma factor changes the DNA-binding properties of RNA polymerase so that its affinity for general DNA is reduced and its affinity for promoters is increased.

E. coli promoters have -35 TTGACA and -10 TATAAT consensus sequences



hybridní promotory tac = lac x trp

19.7 The Holoenzyme Goes Through Transitions in the Process of Recognizing and Escaping from Promoters

- When RNA polymerase binds to a promoter, it separates the DNA strands to form a transcription bubble and incorporates nucleotides into RNA.

19.7 The Holoenzyme Goes Through Transitions in the Process of Recognizing and Escaping from Promoters

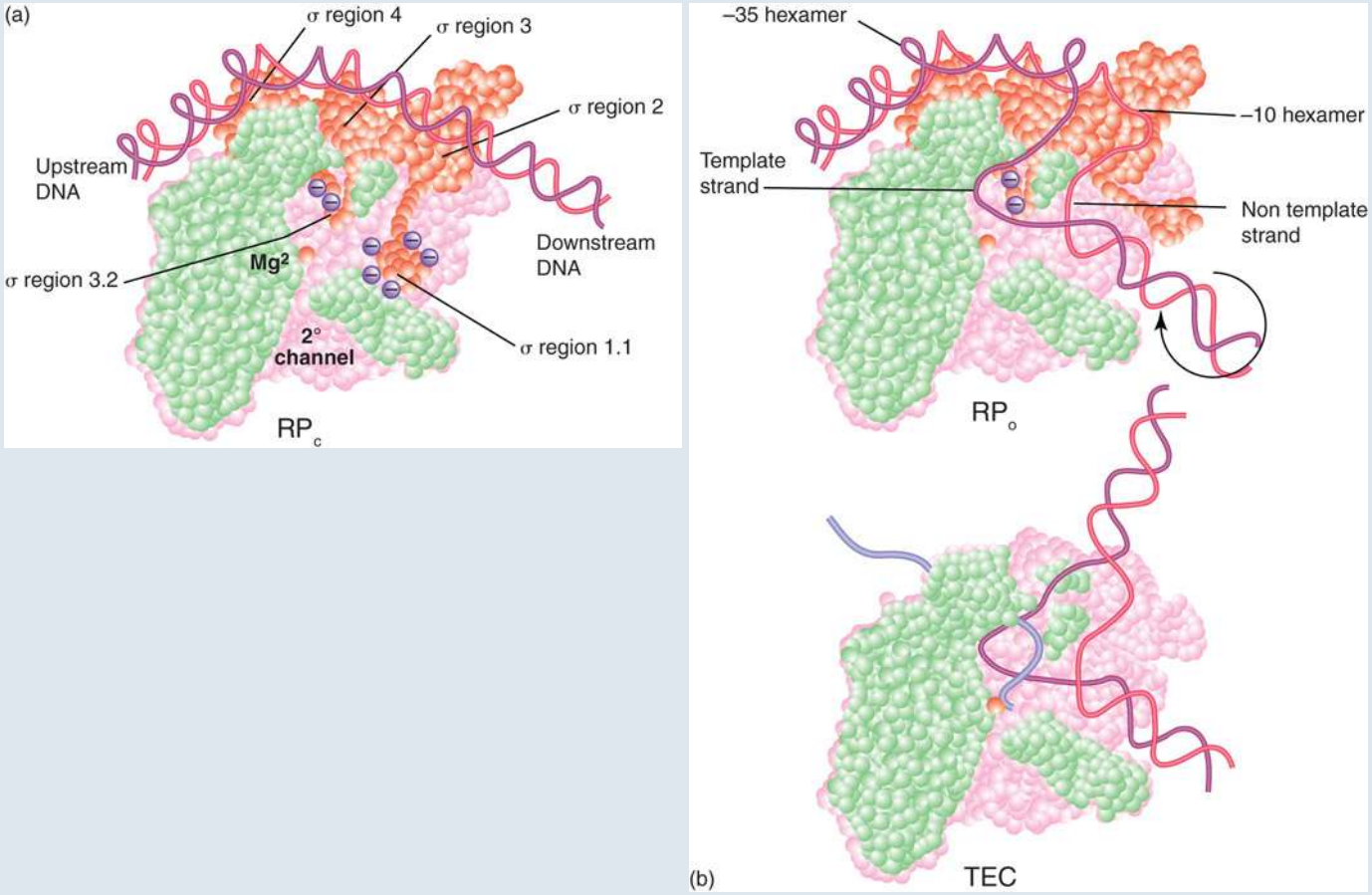


Figure 19.12: RNA polymerase passes through several steps prior to elongation.

Adapted from S. P. Haugen, W. Ross, and R. L. Gourse, Nat. Rev. Microbiol. 6 (2008): 507-519.

19.7 The Holoenzyme Goes Through Transitions in the Process of Recognizing and Escaping from Promoters

- **ternary complex –** The complex in initiation of transcription that consists of RNA polymerase and DNA as well as a dinucleotide that represents the first two bases in the RNA product.
- There may be a cycle of **abortive initiations** before the enzyme moves to the next phase.
- Sigma factor is usually released from RNA polymerase when the nascent RNA chain reaches ~10 bases in length.

19.7 The Holoenzyme Goes Through Transitions in the Process of Recognizing and Escaping from Promoters

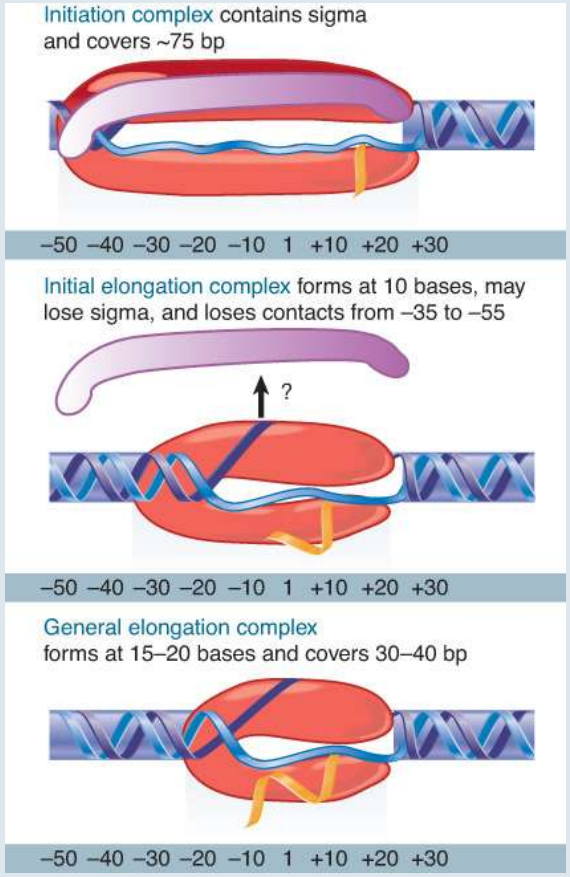


Figure 19.13: RNA polymerase initially contacts the region from –55 to +20.

19.8 Sigma Factor Controls Binding to DNA by Recognizing Specific Sequences in Promoters

- **conserved sequence –** Sequences in which many examples of a particular nucleic acid or protein are compared and the same individual bases or amino acids are always found at particular locations.
- A promoter is defined by the presence of short **consensus sequences** at specific locations.

19.8 Sigma Factor Controls Binding to DNA by Recognizing Specific Sequences in Promoters

- The promoter consensus sequences usually consist of a purine at the start point, a hexamer with a sequence close to TATAAT centered at ~ –10 **(–10 element** or **TATA box)**, and another hexamer with a sequence similar to TTGACA centered at ~–35 **(–35 element)**.
- Individual promoters usually differ from the consensus at one or more positions.

19.8 Sigma Factor Controls Binding to DNA by Recognizing Specific Sequences in Promoters

- Promoter efficiency can be affected by additional elements as well.
- **UP element –** A sequence in bacteria adjacent to the promoter, upstream of the –35 element, that enhances transcription.

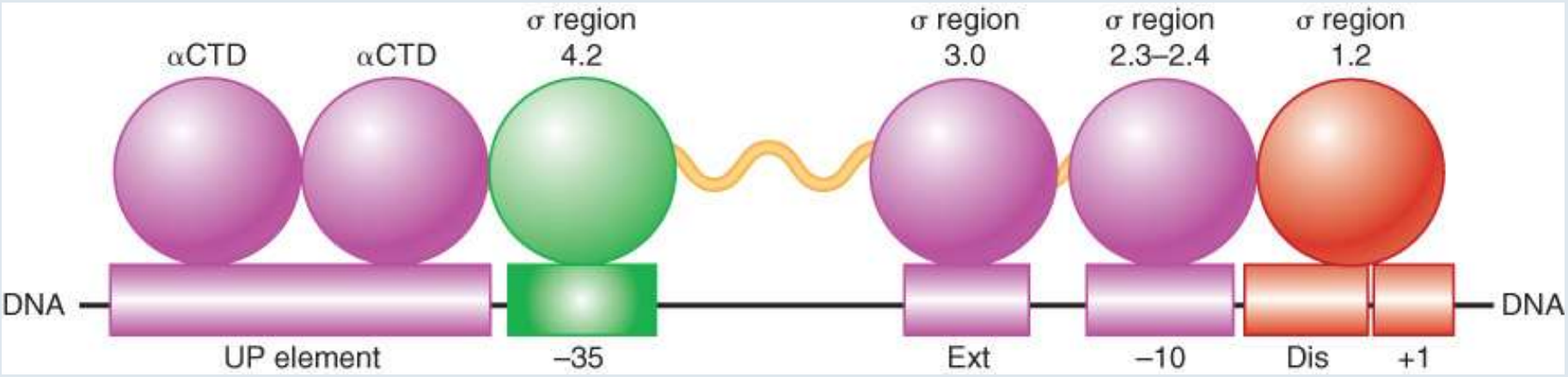


Figure 19.14: DNA elements and RNA polymerase modules that contribute to promoter recognition by sigma factor.

Adapted from S. P. Haugen, W. Ross, and R. L. Gourse, Nat. Rev. Microbiol. 6 (2008): 507-519.

19.10 Multiple Regions in RNA Polymerase Directly Contact Promoter DNA

- The structure of σ^{70} changes when it associates with core enzyme, allowing its DNA-binding regions to interact with the promoter.

19.10 Multiple Regions in RNA Polymerase Directly Contact Promoter DNA

- Multiple regions in σ^{70} interact with the promoter.
- The α subunit also contributes to promoter recognition.

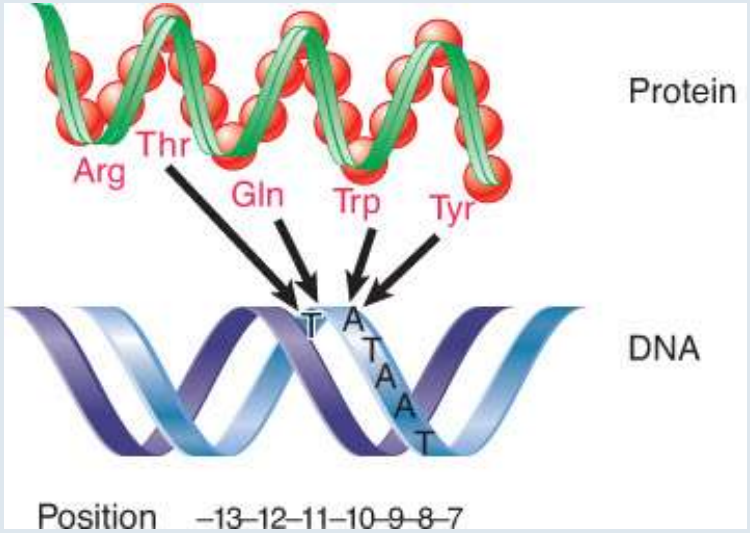


Figure 19.17: Amino acids in the 2.4 -helix of 70 contact specific bases in the coding strand of the –10 promoter sequence.

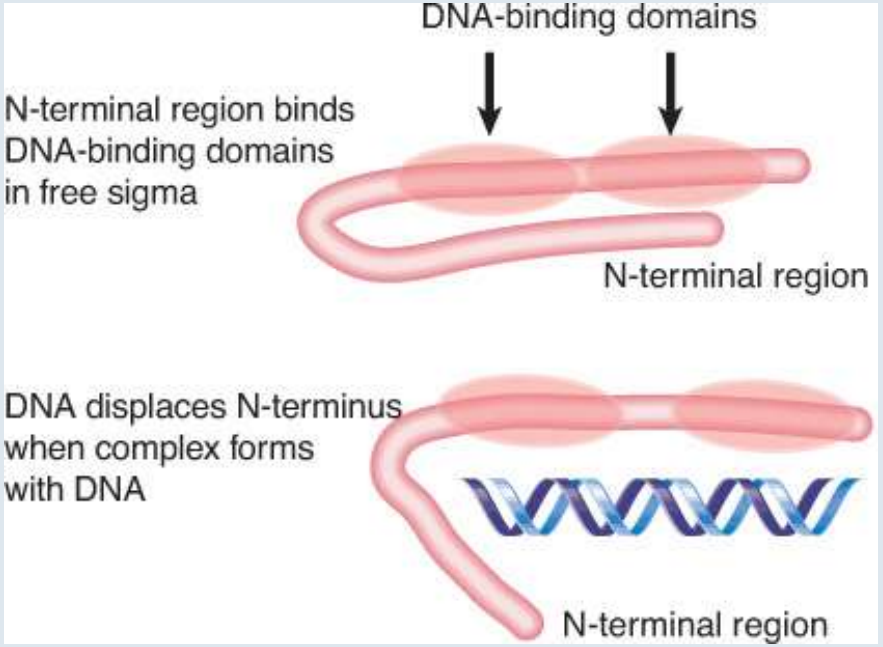
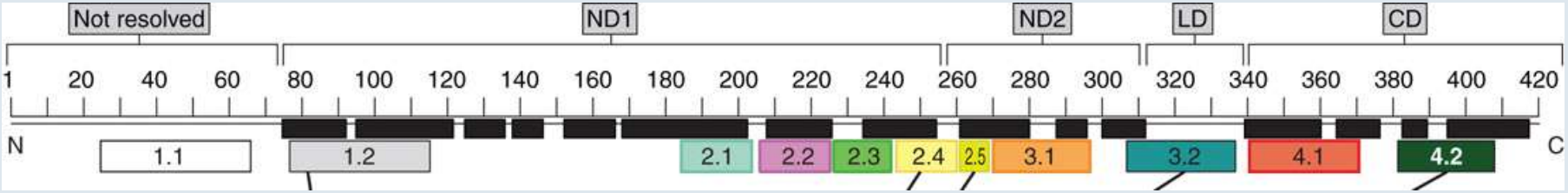
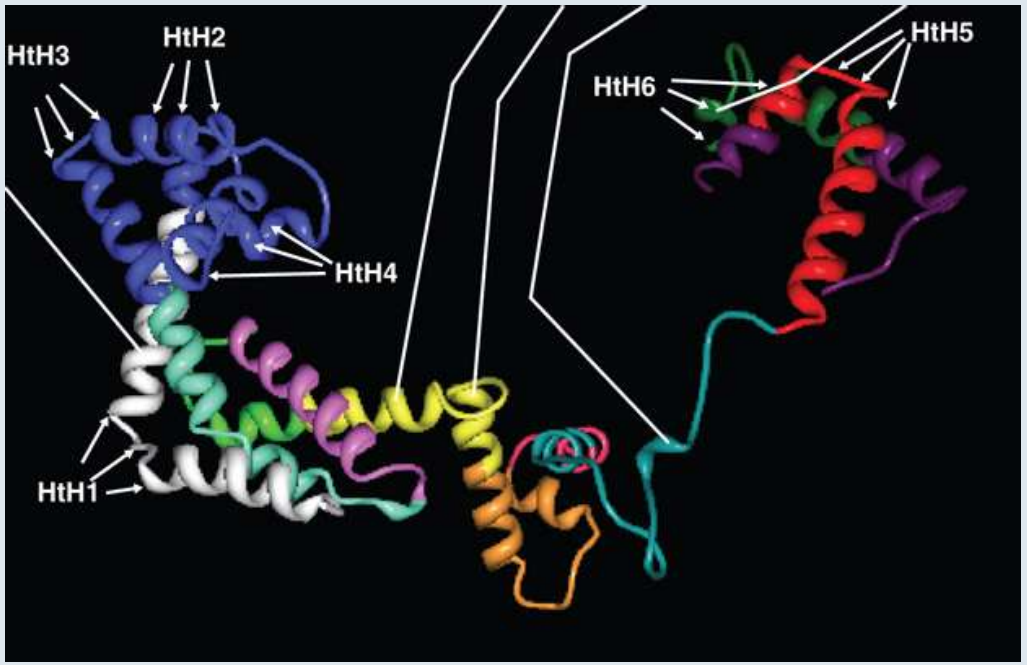


Figure 19.18: The N-terminus of sigma blocks the DNA-binding regions from binding to DNA.

19.10 Multiple Regions in RNA Polymerase Directly Contact Promoter DNA



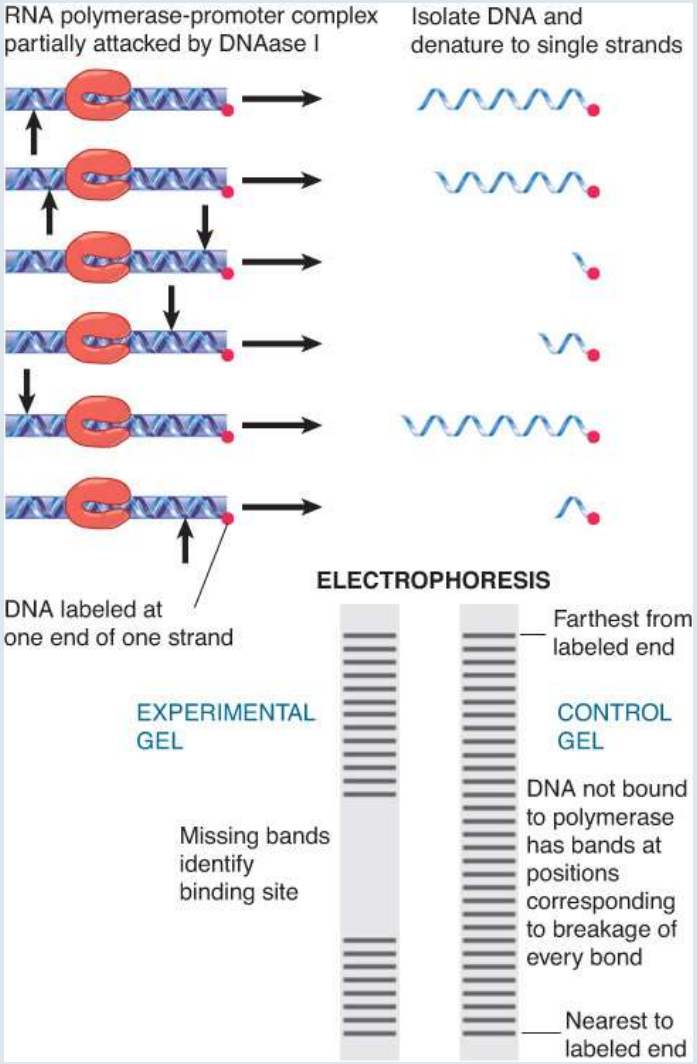


Structure from Protein Data Bank 1IW7. D. G. Vassylyev, et al., Nature 417 (2002): 712-719.

Figure 19.16: The structure of sigma factor in the context of the holoenzyme:-10 and -35 interactions.

Illustration adapted from D. G. Vassylyev, et al., Nature 417 (2002): 712-719.

19.11 RNA Polymerase–Promoter and DNA–Protein Interactions Are the Same for Promoter Recognition and DNA Melting



- **footprinting –** A technique for identifying the site on DNA bound by some protein by virtue of the protection of bonds in this region against attack by nucleases.

Figure 19.21: Footprinting identifies DNA-binding sites for proteins by their protection against nicking.

19.11 RNA Polymerase–Promoter and DNA–Protein Interactions Are the Same for Promoter Recognition and DNA Melting

- The consensus sequences at –35 and –10 provide most of the contact points for RNA polymerase in the promoter.
- The points of contact lie primarily on one face of the DNA.
- Melting the double helix begins with base flipping within the promoter.

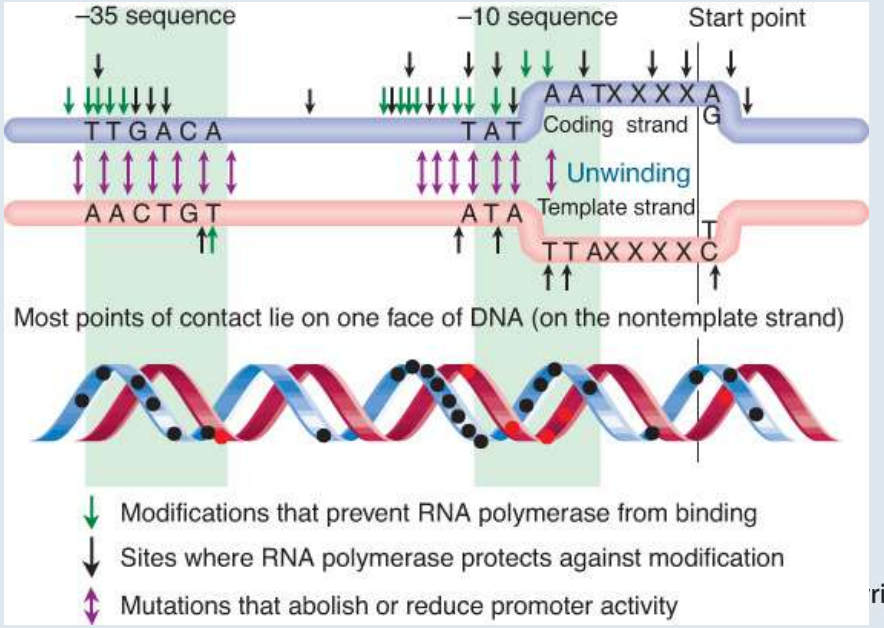


Figure 19.22: One face of the promoter contains the contact points for RNA.

19.12 Interactions Between Sigma Factor and Core RNA Polymerase Change During Promoter Escape

- A domain in sigma occupies the RNA exit channel and must be displaced to accommodate RNA synthesis.
- Abortive initiations usually occur before the enzyme forms a true elongation complex.
- Sigma factor is usually released from RNA polymerase by the time the nascent RNA chain reaches ~10 nt in length.

19.12 Interactions Between Sigma Factor and Core RNA Polymerase Change During Promoter Escape

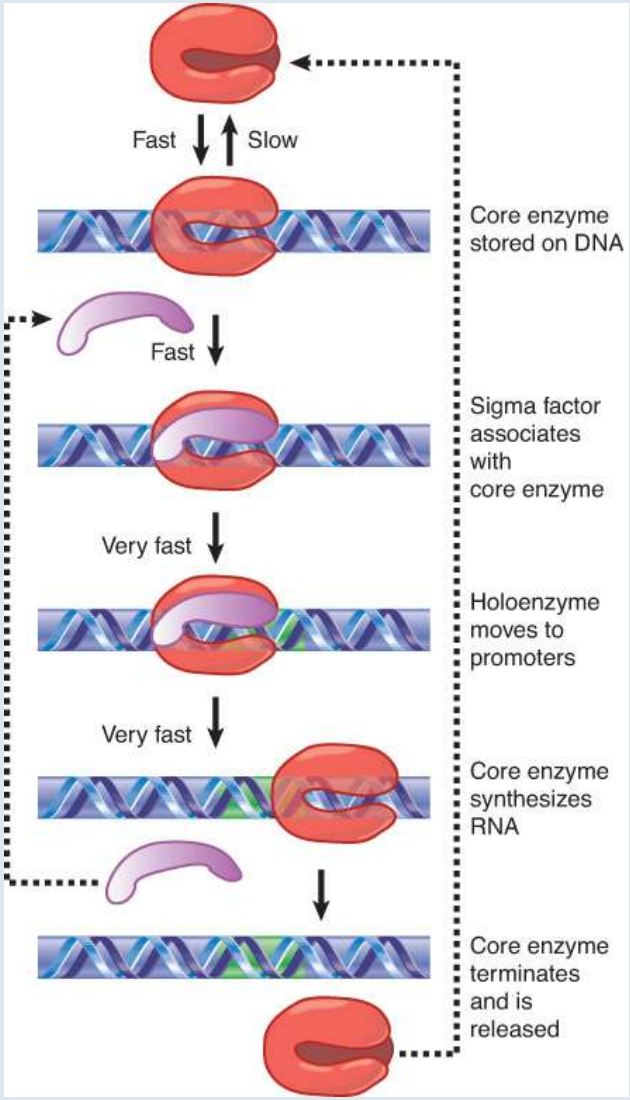


Figure 19.24: Sigma factor and core enzyme recycle at different points in transcription.

RNA polymerase sigma 70 keeps a firm grip on promoter -35 and -10 sequences during DNA melting and transcription initiation

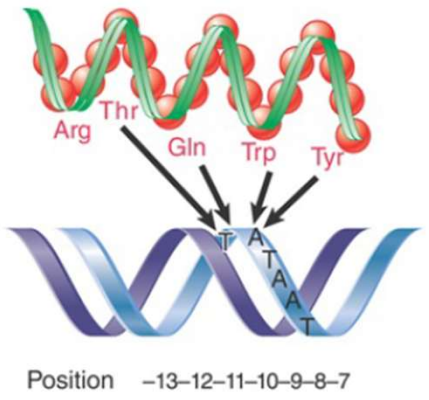


FIGURE 17.16 Amino acids in the 2.4 α-helix of β^{70} contact specific bases in the coding strand of the −10 promoter sequence.

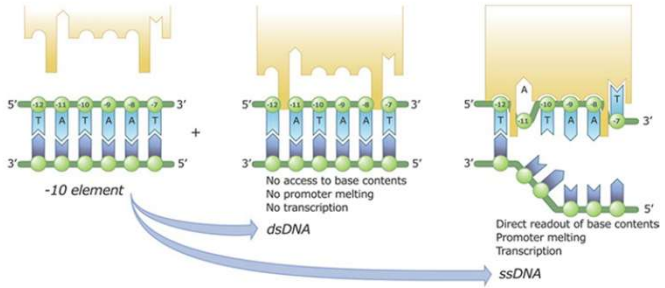


FIGURE 17.22 Sequence-specific recognition of the −10 element by region 2 of σ. The DNA backbone is represented by green circles, bases of the nontemplate strand by dark blue polygons, and bases of the template strand by light blue polygons. The sequence of the nontemplate strand corresponds to the consensus of the −10 element. Region 2 of σ is shown as an orange polygon.

Data from X. Liu, et al., *Cell* 147 (2011): 1218–1219.

Major groove recognition of -35 promoter sequence and unusual minor groove recognition at -10 by RNA polymerase.

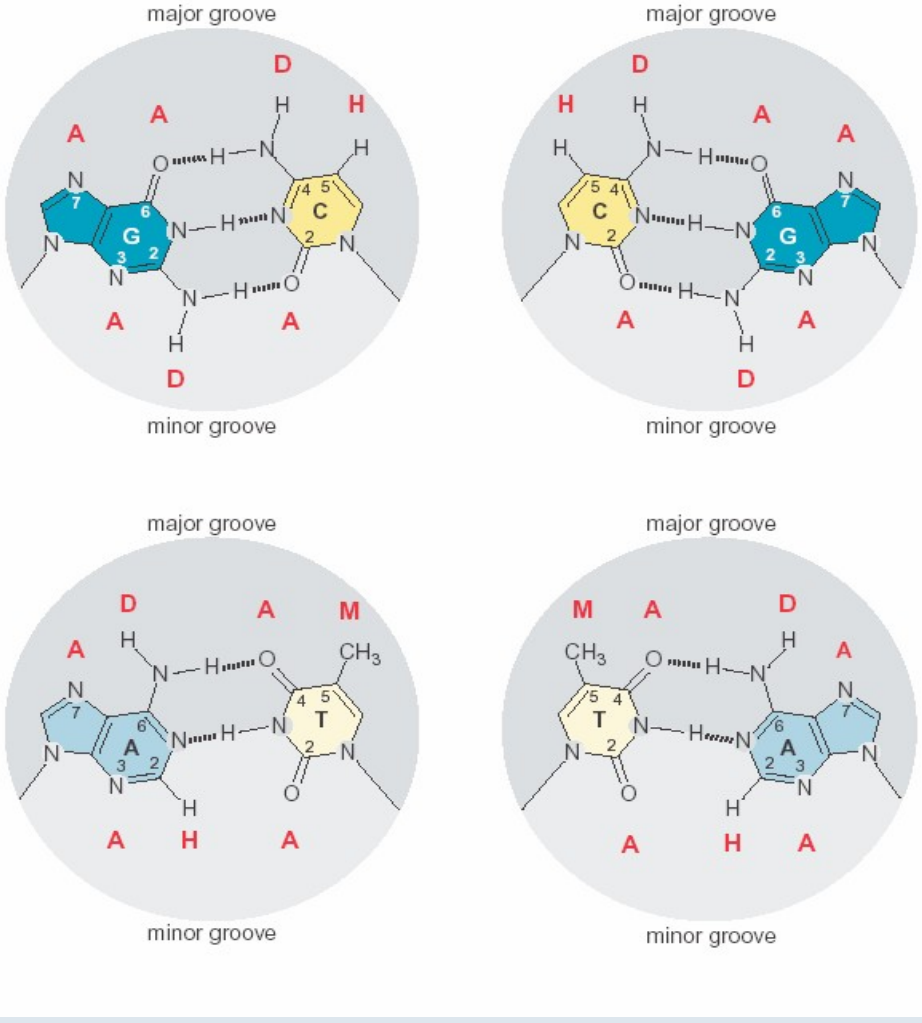
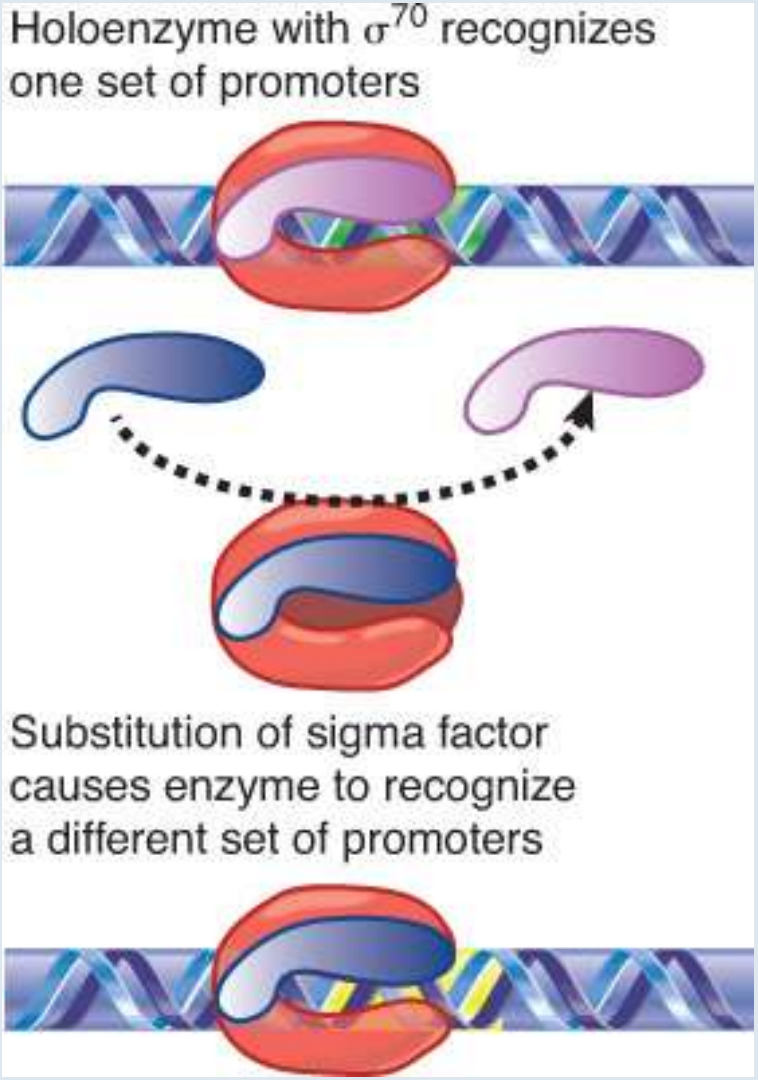


FIGURE 6-10 Chemical groups exposed in the major and minor grooves from the edges of the base pairs. The letters in red identify hydrogen bond acceptors (**A**), hydrogen bond donors (**D**), nonpolar hydrogens (**H**), and methyl groups (**M**).

19.19 Competition for Sigma Factors Can Regulate Initiation



- *E. coli* has seven sigma factors, each of which causes RNA polymerase to initiate at a set of promoters defined by specific –35 and –10 sequences.

Figure 19.36: The sigma factor associated with core enzyme determines the set of promoters at which transcription is initiated.

19.19 Competition for Sigma Factors Can Regulate Initiation

- The activities of the different sigma factors are regulated by different mechanisms.
- **anti-sigma factor –** A protein that binds to a sigma factor to inhibit its ability to utilize specific promoters.

Gene	Factor	Use
rpoD	σ^{70}	most required functions
rpoS	σ^{S}	stationary phase/some stress responses
rpoH	σ^{32}	heat shock
rpoE	σ^{E}	periplasmic/extracellular proteins
rpoN	σ^{54}	nitrogen assimilation
rpoF	σ^{F}	flagellar synthesis/chemotaxis
fecI	σ^{fecI}	iron metabolism/transport

Figure 19.37: In addition to 70, E. coli has several sigma factors that are induced by particular environmental conditions.

19.19 Competition for Sigma Factors Can Regulate Initiation

- **Heat-shock response –** A set of loci that is activated in response to an increase in temperature that causes proteins to denature (and other abuses to the cell).
  - All organisms have this response.
  - The gene products usually include chaperones that act on denatured proteins.

E. coli has other sigma factors in addition to sigma 70

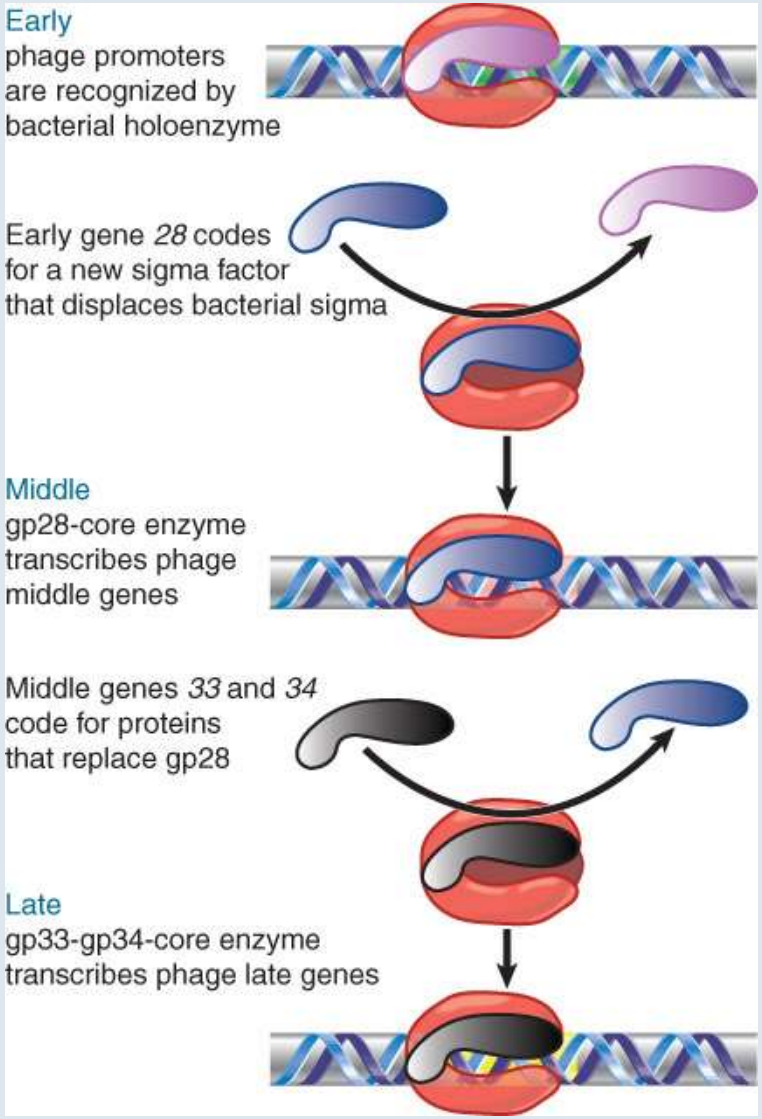
TABLE 17.2 In addition to σ^{70}, *E. coli* has several sigma factors that are induced by particular environmental conditions. (A number in the name of a factor indicates its mass.)

Gene	Factor	Use
rpoD	σ^{70}	Most required functions
rpoS	σ^{S}	Stationary phase/some stress responses
rpoH	σ^{32}	Heat shock
rpoE	σ^{E}	Periplasmic/extracellular proteins
rpoN	σ^{54}	Nitrogen assimilation
rpoF	σ^{F}	Flagellar synthesis/chemotaxis
fecI	σ^{fecI}	Iron metabolism/transport

TABLE 17.1 *E. coli* sigma factors recognize promoters with different consensus sequences.

Subunit (Gene)	Size (Number of Amino Acids)	Approximate Number of Promoters	Promoter Sequence Recognized
Sigma 70 (*rpoD*)	613	1,000	TTGACA–16 to 18 bp–TATAAT
Sigma 54 (*rpoN*)	477	5	CTGGNA–6 to 18 bp–TATAAT
Sigma S (*rpoS*)	330	100	TTGACA–16 to 18 bp–TATAAT
Sigma 32 (*rpoH*)	284	30	CCCTTGAA–13 to 15 bp–CCCGATNT
Sigma F (*rpoF*)	239	40	CTAAA–15 bp–GCCGATAA
Sigma E (*rpoE*)	202	20	GAA–16 bp–YCTGA
Sigma FecI (*fecI*)	173	1–2	?

19.20 Sigma Factors May Be Organized into Cascades



- A **cascade** of sigma factors is created when one sigma factor is required to transcribe the gene coding for the next sigma factor.
- In Bacillus subtilis the **early genes** of phage SPO1 are transcribed by host RNA polymerase.

Figure 19.39: Transcription of phage SPO1 genes is controlled by two successive substitutions of the sigma factor that change the initiation specificity.

19.15 Bacterial RNA Polymerase Terminates at Discrete Sites

- There are two classes of terminators: Those recognized solely by RNA polymerase itself without the requirement for any cellular factors are usually referred to as "**intrinsic terminators**."
  - Others require a cellular protein called **rho** and are referred to as "**rho-dependent terminators**."

19.15 Bacterial RNA Polymerase Terminates at Discrete Sites

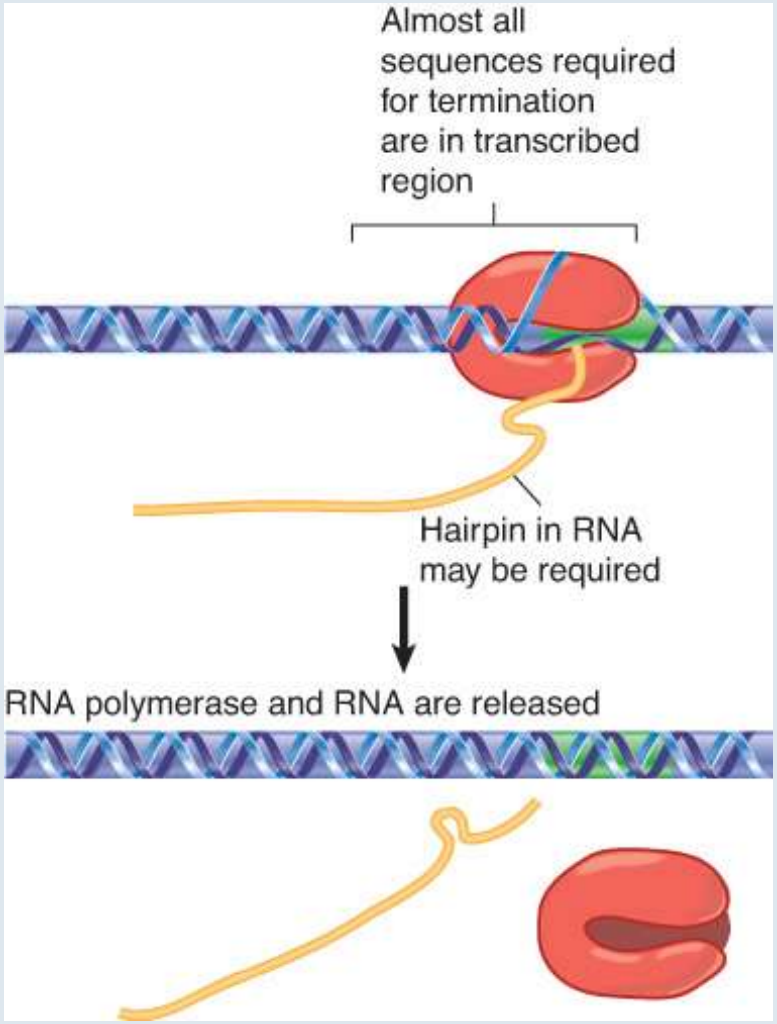


Figure 19.28: The DNA sequences required for termination are located upstream of the terminator sequence.

19.15 Bacterial RNA Polymerase Terminates at Discrete Sites

- Intrinsic termination requires recognition of a terminator sequence in DNA that codes for a **hairpin** structure in the RNA product.
- The signals for termination lie mostly within *sequences already transcribed* by RNA polymerase, and thus termination relies on scrutiny of the template and/or the RNA product that the polymerase is transcribing.

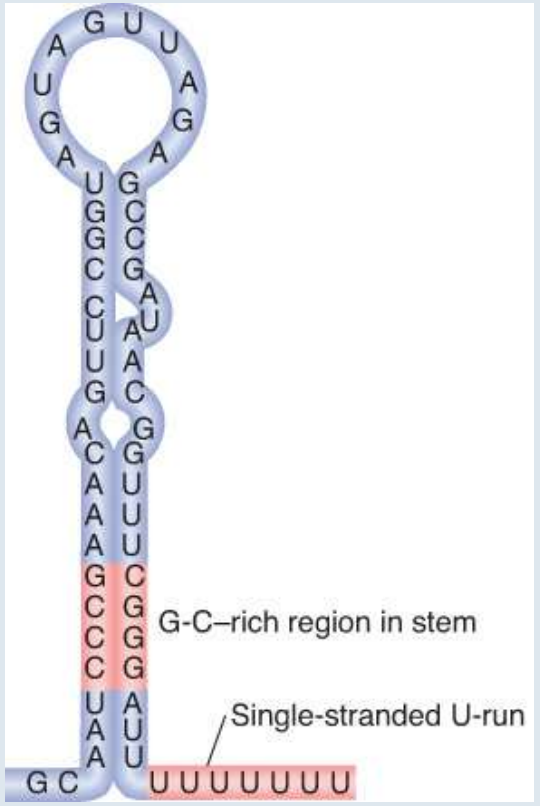


Figure 19.29: Intrinsic terminators include palindromic regions that form hairpins varying in length from 7 to 20 bp.

19.15 Bacterial RNA Polymerase Terminates at Discrete Sites

- **readthrough –** It occurs at transcription or translation when RNA polymerase or the ribosome, respectively, ignores a termination signal because of a mutation of the template or the behavior of an accessory factor.
- **antitermination –** A mechanism of transcriptional control in which termination is prevented at a specific terminator site, allowing RNA polymerase to read into the genes beyond it.

19.16 How Does Rho Factor Work?

- Rho factor is a protein that binds to nascent RNA and tracks along the RNA to interact with RNA polymerase and release it from the elongation complex.
- ***rut* –** An acronym for rho utilization site, the sequence of RNA that is recognized by the rho termination factor.

19.16 How Does Rho Factor Work?

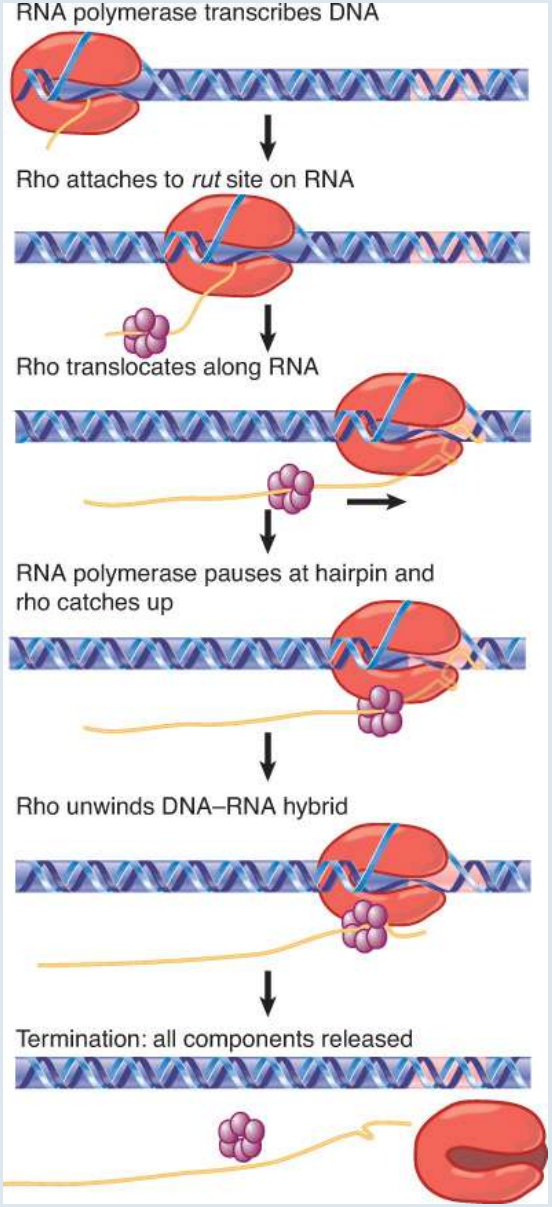


Figure 19.30: Rho factor binds to RNA at a rut site and translocates along RNA until it reaches the RNA–DNA hybrid in RNA polymerase.

19.16 How Does Rho Factor Work?

- **polarity –** The effect of a mutation in one gene in influencing the expression (at transcription or translation) of subsequent genes in the same transcription unit.
- **antitermination complex –** Proteins that allow RNA polymerase to transcribe through certain terminator sites.

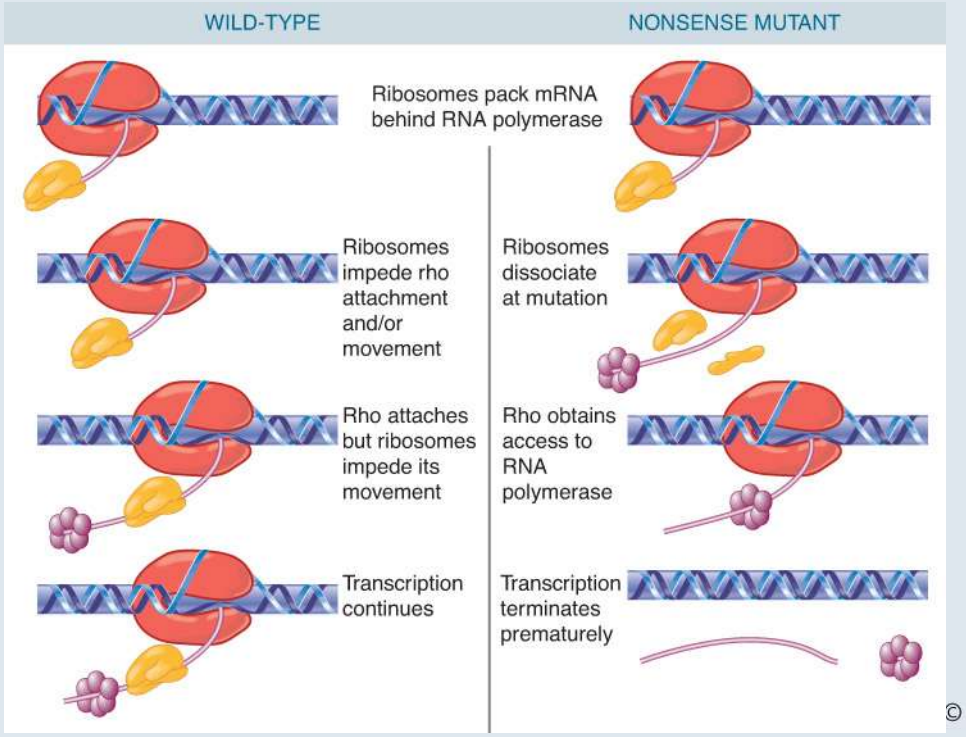


Figure 19.33: The action of rho factor may create a link between transcription and translation.

19.22 Antitermination Can Be a Regulatory Event

- An antitermination complex allows RNA polymerase to read through terminators.

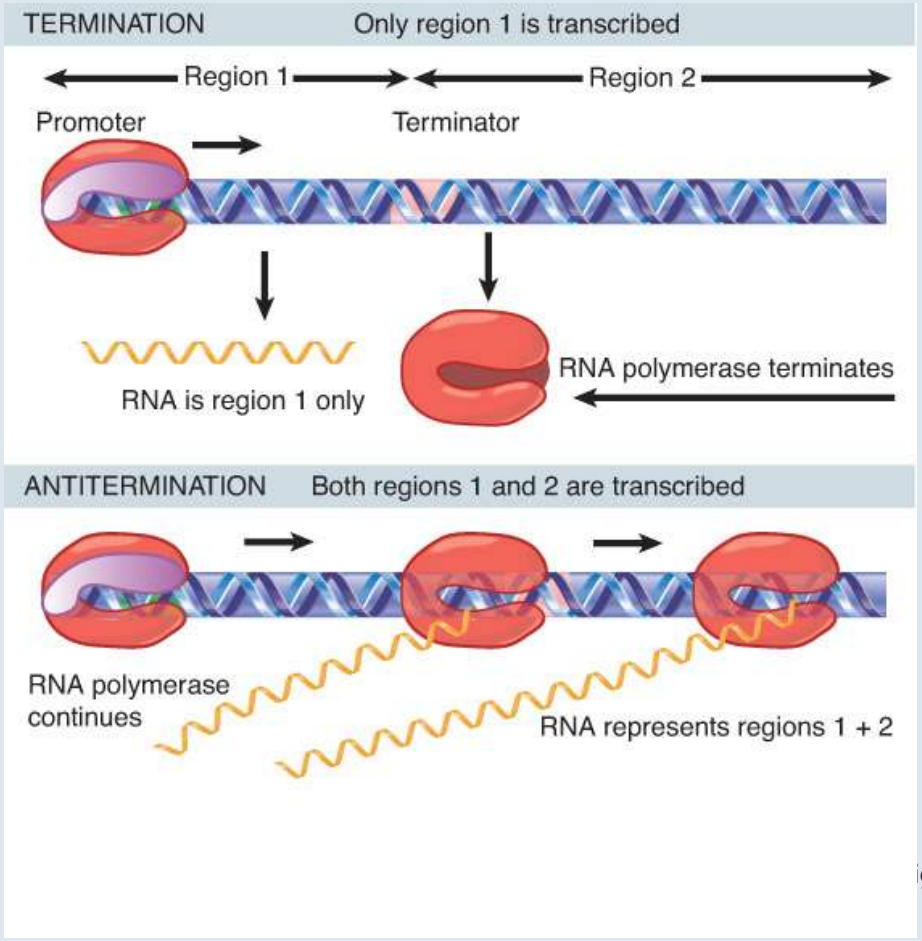
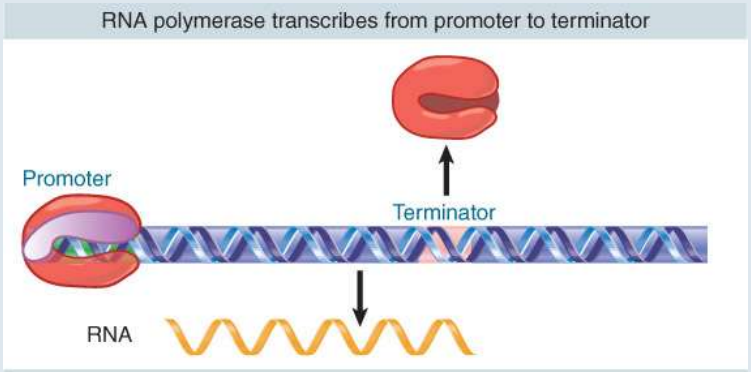


Figure 19.44: Antitermination can control transcription by determining whether RNA polymerase terminates or reads through a particular terminator.

19.22 Antitermination Can Be a Regulatory Event



Antitermination proteins act on specific terminators

Transcription unit	Promoter	Terminator	Antitermination Protein
Immediate early	P_L	t_L	pN
Immediate early	P_{R1}	t_{R1}	pN
Late	$P_{R'}$	$t_{R'}$	pQ

Figure 19.45: An antitermination protein can act on RNA polymerase to enable it to read through a specific terminator.

- ***nut*** — An acronym for N utilization site, the sequence of DNA that is recognized by the N antitermination factor.
- Phage lambda uses antitermination systems for regulation of both its early and late transcripts, but the two systems work by completely different mechanisms.

19.22 Antitermination Can Be a Regulatory Event

- Binding of factors to the nascent RNA links the antitermination proteins to the terminator site through an RNA loop.
- Antitermination of transcription also occurs in rRNA operons.

Operons and *E. coli* gene regulation

26.1 Introduction

- In **negative regulation**, a repressor protein binds to an **operator** to prevent a gene from being expressed.
- In **positive regulation**, a transcription factor is required to bind at the promoter in order to enable RNA polymerase to initiate transcription.

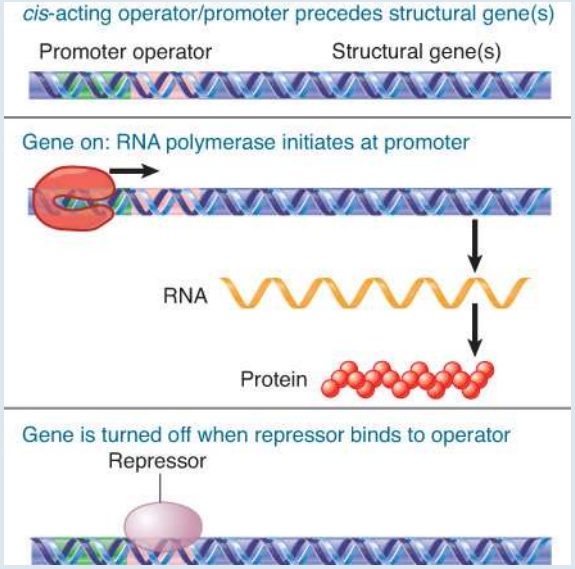


Figure 26.02: In negative control, a trans-acting repressor binds to the cis-acting operator to turn off transcription.

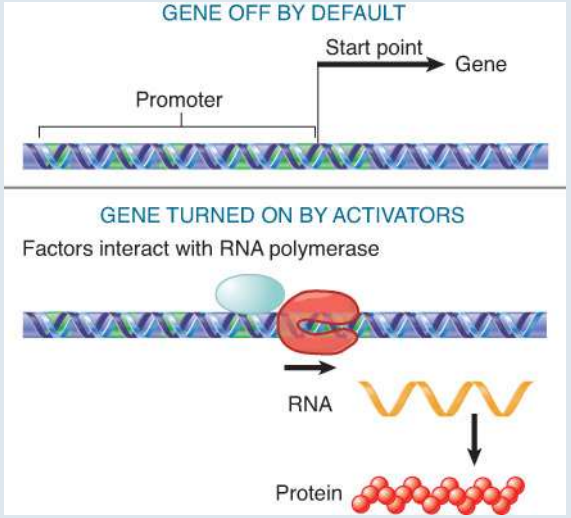


Figure 26.03: In positive control, a trans-acting factor must bind to cis-acting site in order for RNA polymerase to initiate transcription at the promoter.

26.1 Introduction

- In **inducible regulation**, the gene is regulated by the presence of its substrate (the **inducer**). This is good for catabolic pathways for adapting to new food sources, for instance to use lactose or galactose sugars to grow if there is no glucose.

- In **repressible regulation**, the gene is regulated by the product of its enzyme pathway (the **corepressor**). Genes encoding enzymes in anabolic pathways to make amino acids or nucleotides or other needed products can reduce expression when there is enough product.

26.3 Famous inducible genes and repressors in *E. coli*

The *lac* Operon Is under Negative control by lac repressor

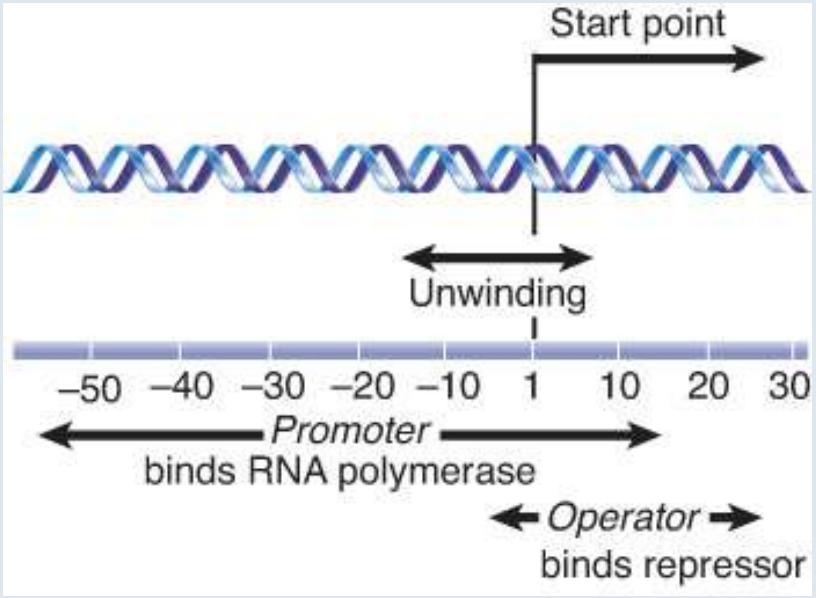


Figure 26.06: lac repressor and RNA polymerase bind at sites that overlap around the transcription startpoint of the lac operon.

- Transcription of the *lacZYA* operon is controlled by a repressor protein (the ***lac* repressor**) that binds to an operator that overlaps the promoter at the start of the cluster.
- In the absence of β-galactosides, the *lac* operon is expressed only at a very low (basal) level.

26.3 The *lac* Operon Is under negative transcriptional control by repressor and is inducible by galactosides

- The repressor protein is a tetramer of identical subunits coded by the *lacI* gene.
- β-galactoside sugars, the substrates of the *lac* operon, are its inducer.
- Addition of specific β-galactosides induces transcription of all three genes of the *lac* operon.
- The *lac* mRNA is extremely unstable; as a result, induction can be rapidly reversed.

26.4 *lac* Repressor Is Controlled by a Small-Molecule Inducer

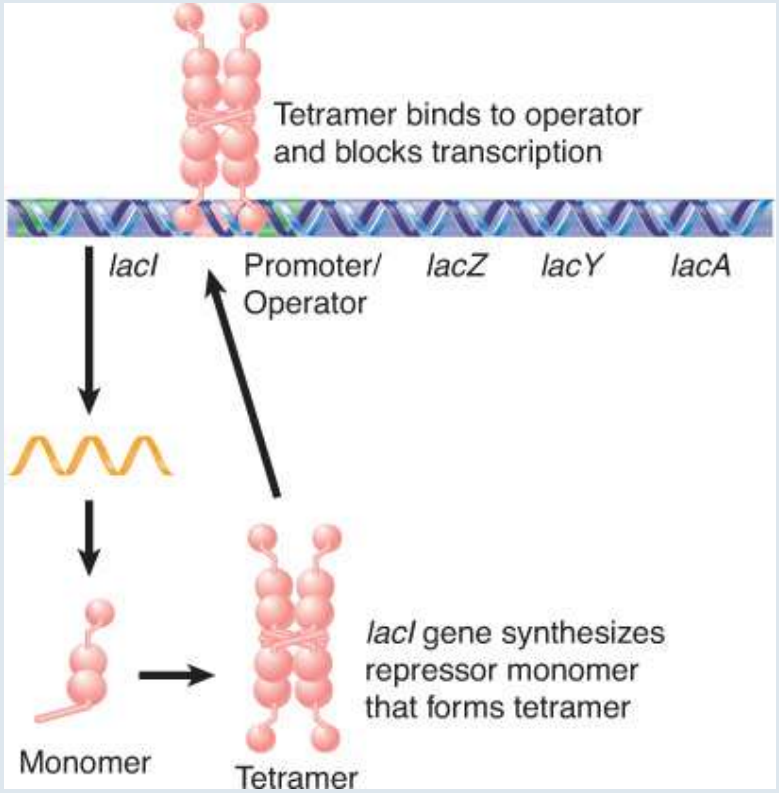


Figure 26.08: lac repressor maintains the lac operon in the inactive condition by binding to the operator.

- An inducer functions by converting the repressor protein into a form with lower operator affinity.
- Repressor has two binding sites, one for the operator DNA and another for the inducer.
- **gratuitous inducer –** Inducers that resemble authentic inducers of transcription, but are not substrates for the induced enzymes.(IPTG for lac repressor)

26.4 *lac* Repressor Is Controlled by a Small-Molecule Inducer

- Repressor is inactivated by an allosteric interaction in which binding of inducer at its site changes the properties of the DNA-binding site (**allosteric control**).
- The true inducer is **allolactose**, not the actual substrate of β-galactosidase. In experiments we induce with artificial IPTG, not metabolized, a gratuitous inducer.

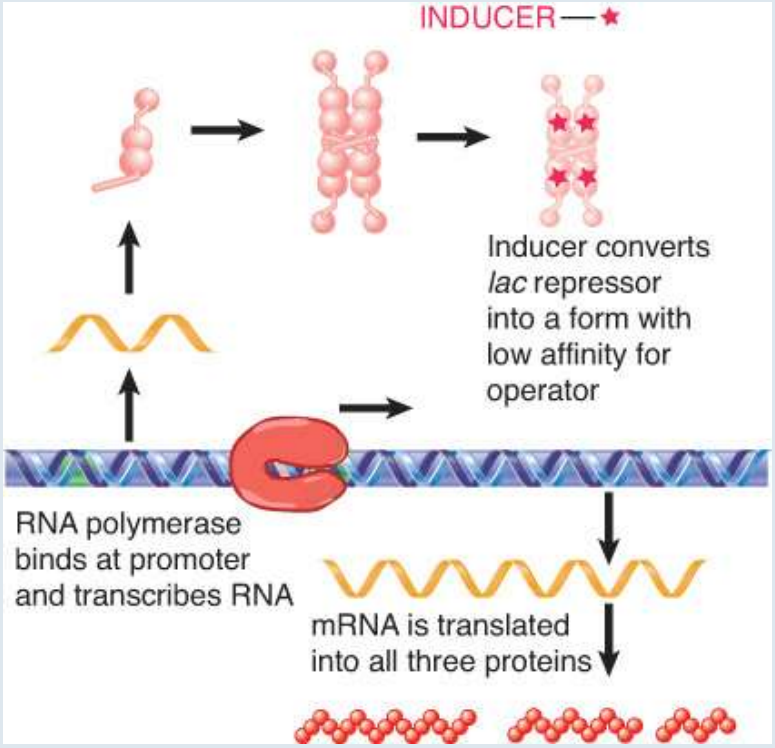


Figure 26.09: Addition of inducer converts repressor to a form with low affinity for the operator. This allows RNA polymerase to initiate transcription.

26.7 *lac* Repressor Is a Tetramer Made of Two Dimers

- A single repressor subunit can be divided into the N-terminal DNA-binding domain, a hinge, and the core of the protein.
- The DNA-binding domain contains two short α-helical regions that bind the major groove of DNA.
- The inducer-binding site and the regions responsible for multimerization are located in the core.

26.7 *lac* Repressor Is a Tetramer Made of Two Dimers

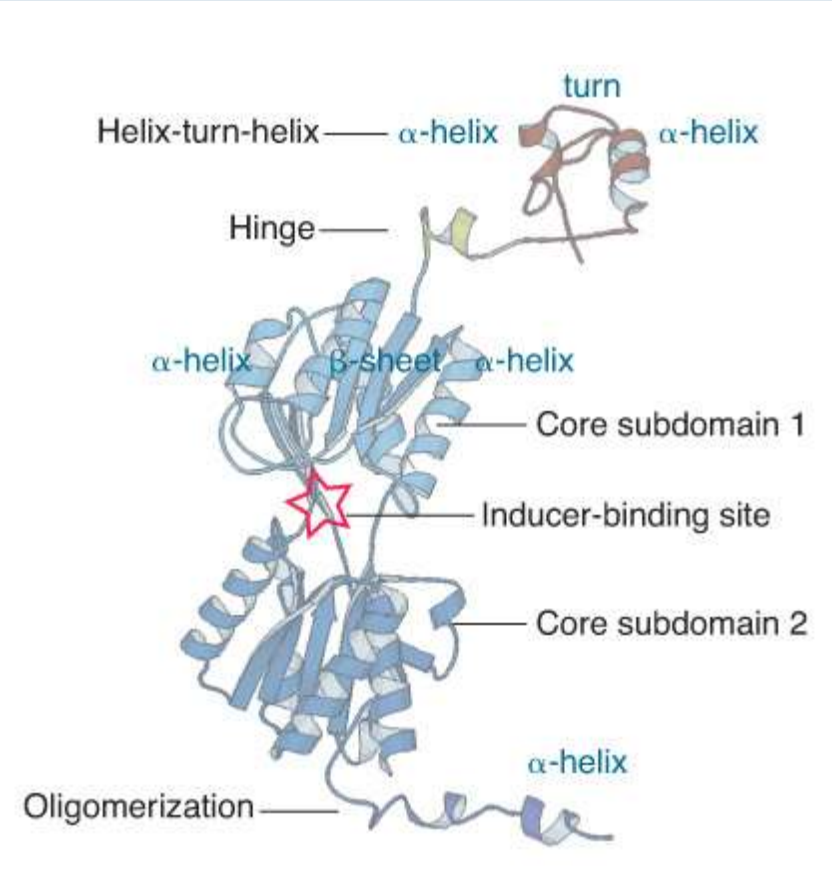


Figure 26.13: The structure of a monomer of Lac repressor identifies several independent domains.

Structure from Protein Data Bank 1LBG. M. Lewis, et al., Science 271 (1996): 1247-1254. Photo courtesy of Hongli Zhan and Kathleen S. Matthews, Rice University.

26.7 *lac* Repressor Is a Tetramer Made of Two Dimers

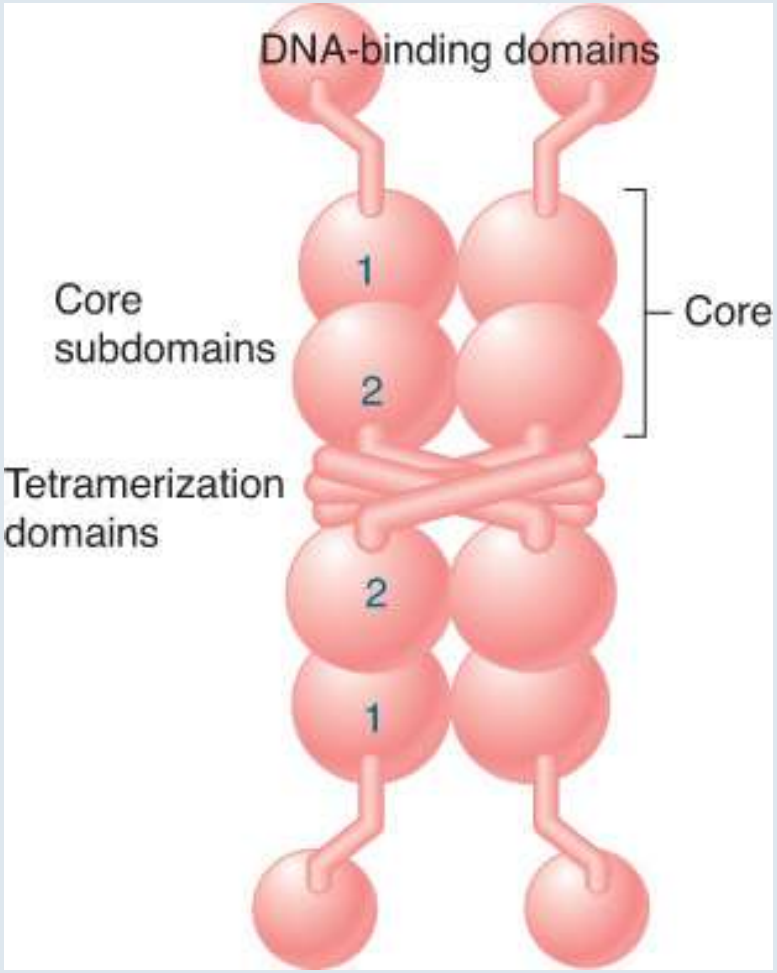


Figure 26.15: The repressor tetramer consists of two dimers.

- Monomers form a dimer by making contacts between core subdomains 1 and 2.
- Dimers form a tetramer by interactions between the tetramerization helices.

26.7 *lac* Repressor Is a Tetramer Made of Two Dimers

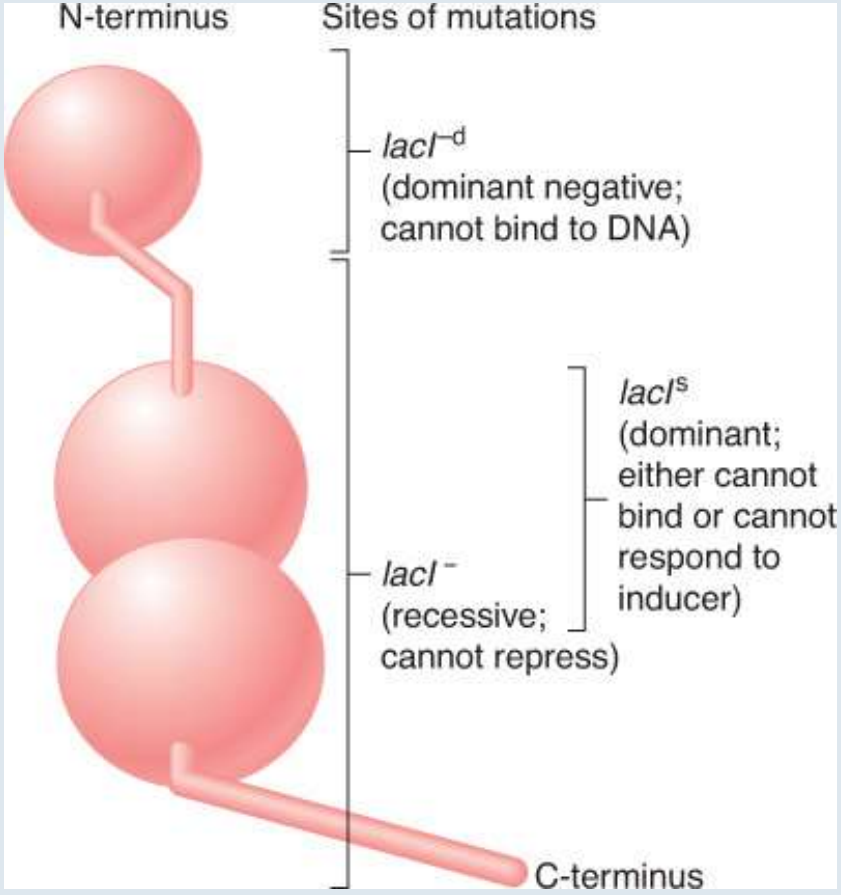


Figure 26.16: The locations of three type of mutations in lactose repressor are mapped on the domain structure of the protein.

- Different types of mutations occur in different domains of the repressor protein.

26.8 *lac* Repressor Binding to the Operator Is Regulated by an Allosteric Change in Conformation

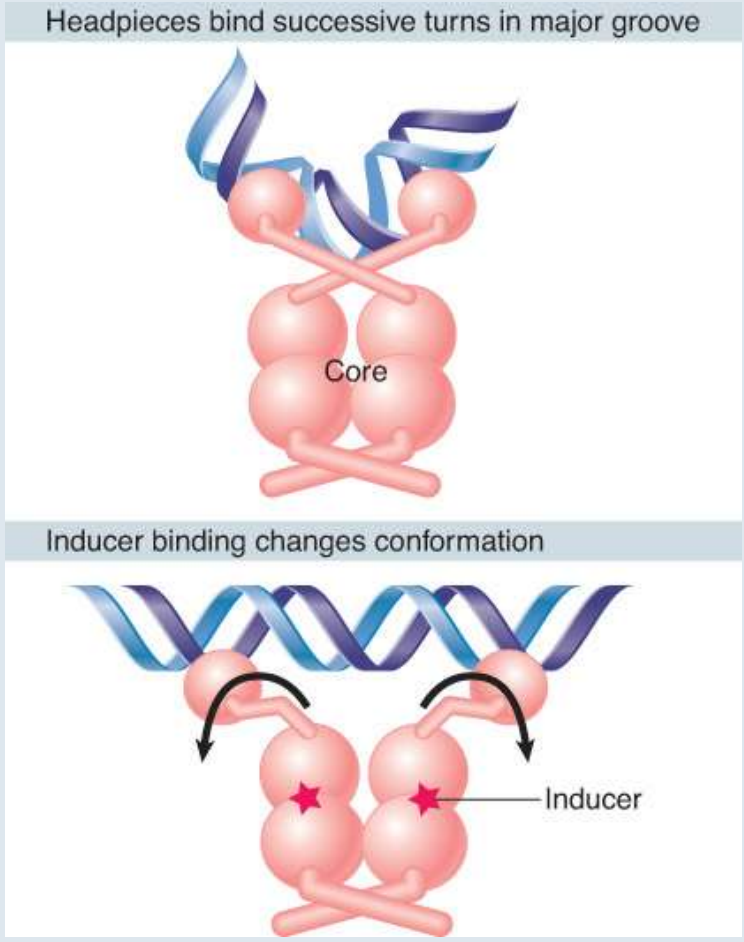


Figure 26.18: The inducer changes the structure of the core.

- Inducer binding causes a change in repressor conformation that reduces its affinity for DNA and releases it from the operator.

26.8 *lac* Repressor Binding to the Operator Is Regulated by an Allosteric Change in Conformation

- *lac* repressor protein binds to the double-stranded DNA sequence of the operator.
- The operator is a **palindromic** sequence of 26 bp.
- Each inverted repeat of the operator binds to the DNA-binding site of one repressor subunit.

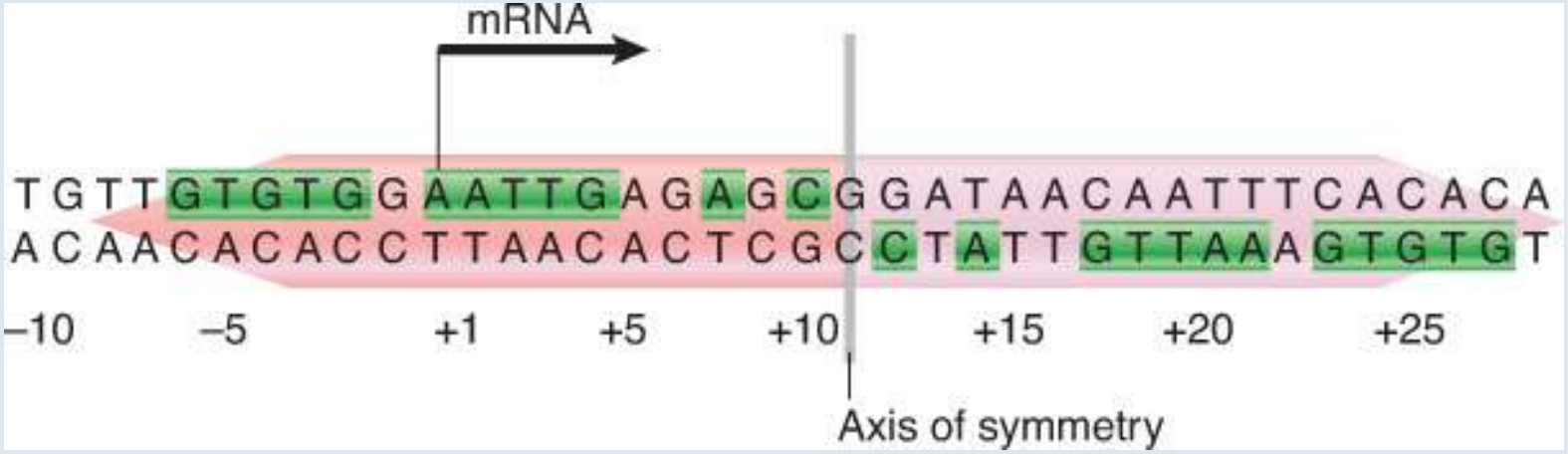


Figure 26.17: The lac operator has a symmetrical sequence.

Wide major groove means DNA sequence can be recognized by DNA-binding proteins.

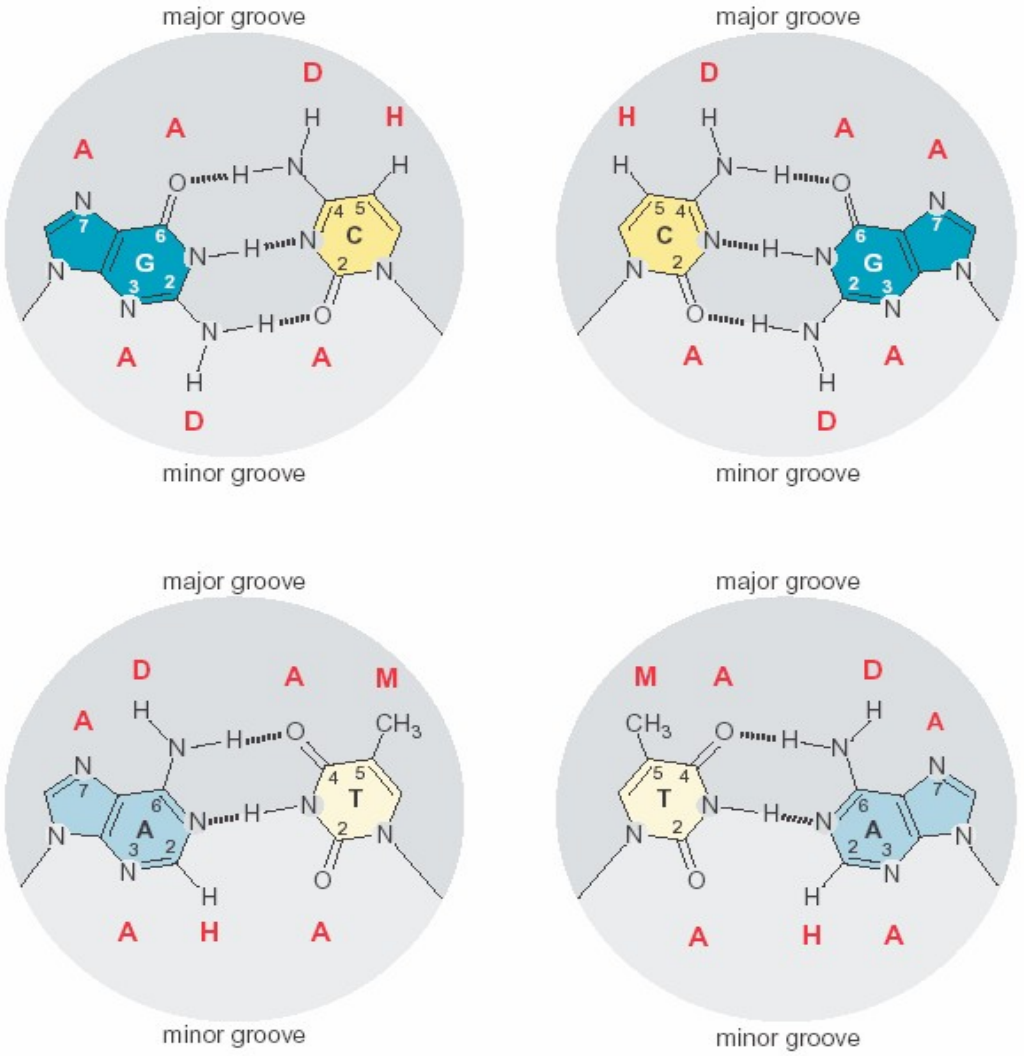


FIGURE 6-10 Chemical groups exposed in the major and minor grooves from the edges of the base pairs. The letters in red identify hydrogen bond acceptors (**A**), hydrogen bond donors (**D**), nonpolar hydrogens (**H**), and methyl groups (**M**).

26.9 *lac* Repressor Binds to Three Operators and Interacts with RNA Polymerase

- Each dimer in a repressor tetramer can bind an operator, so that the tetramer can bind two operators simultaneously.
- Full repression requires the repressor to bind to an additional operator downstream or upstream as well as to the primary operator at the *lacZ* promoter.
- Binding of repressor at the operator stimulates binding of RNA polymerase at the promoter but precludes transcription.

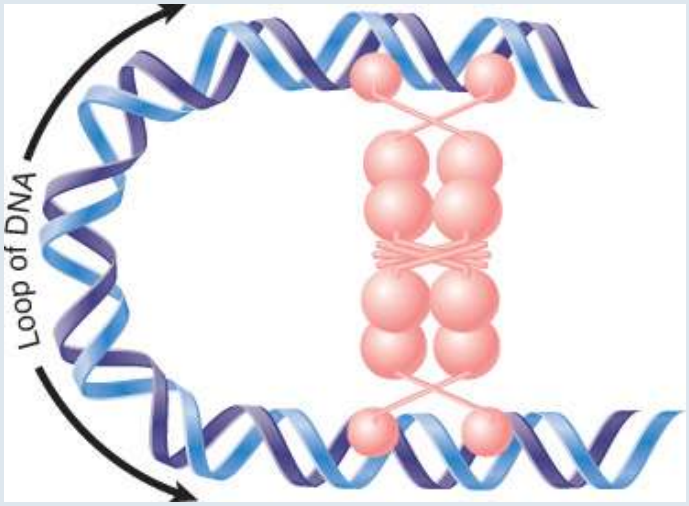


Figure 26.21: If both dimers in a repressor tetramer bind to DNA, the DNA between the two binding sites is held in a loop. Third operator not shown.

26.11 Positive regulators of transcription

The *lac* Operon Has a Second Layer of Control: Catabolite Repression

- **catabolite repression –** The ability of glucose to prevent the expression of a number of genes.
  - In bacteria this is a positive control system; in eukaryotes, it is completely different.
- **Catabolite repressor protein (CRP)** is an activator protein that binds to a target sequence at a promoter.
- CRP is or was also known as cAMP-dependent, **Catabolite Activator Protein (CAP),** a newer name I find clearer.

26.11 The *lac* Operon Has a Second Layer of Control: Catabolite Repression

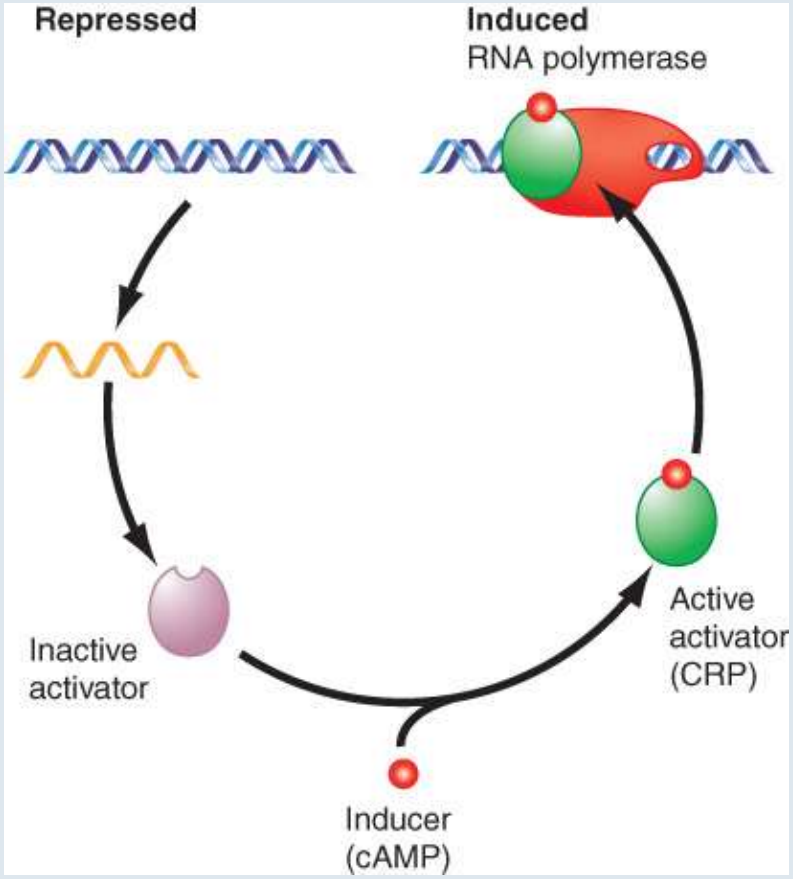


Figure 26.25: cAMP converts an activator protein CRP to a form that binds the promoter and assists RNA polymerase in initiating transcription.

26.11 The *lac* Operon Has a Second Layer of Control: Catabolite Repression

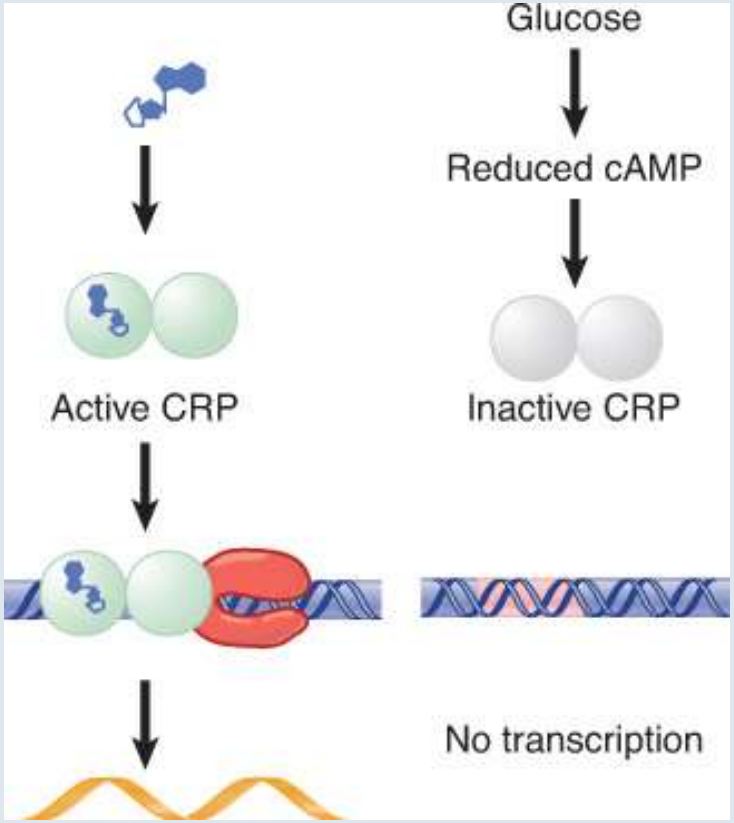
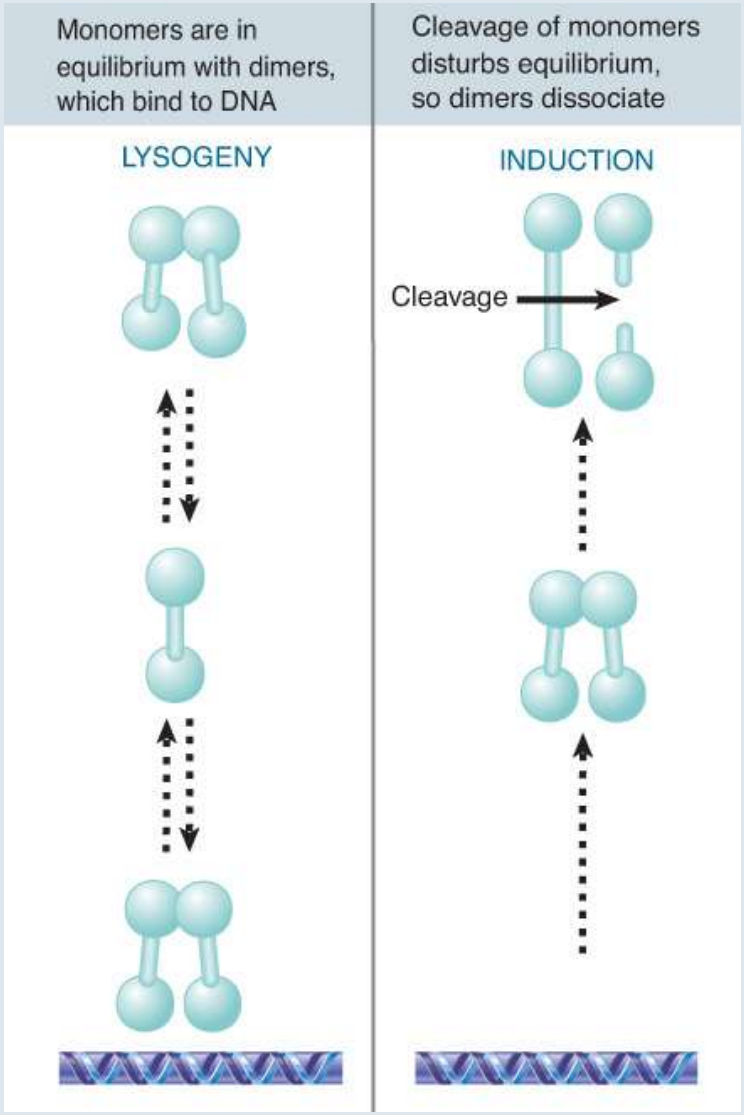


Figure 26.27: By reducing the level of cyclic AMP, glucose inhibits the transcription of operons that require CRP activity.

- A dimer of CRP is activated by a single molecule of **cyclic AMP (cAMP)**.
- cAMP is controlled by the level of glucose in the cell; a low glucose level allows cAMP to be made.
- CRP interacts with the C-terminal domain of the α subunit of RNA polymerase to activate it.

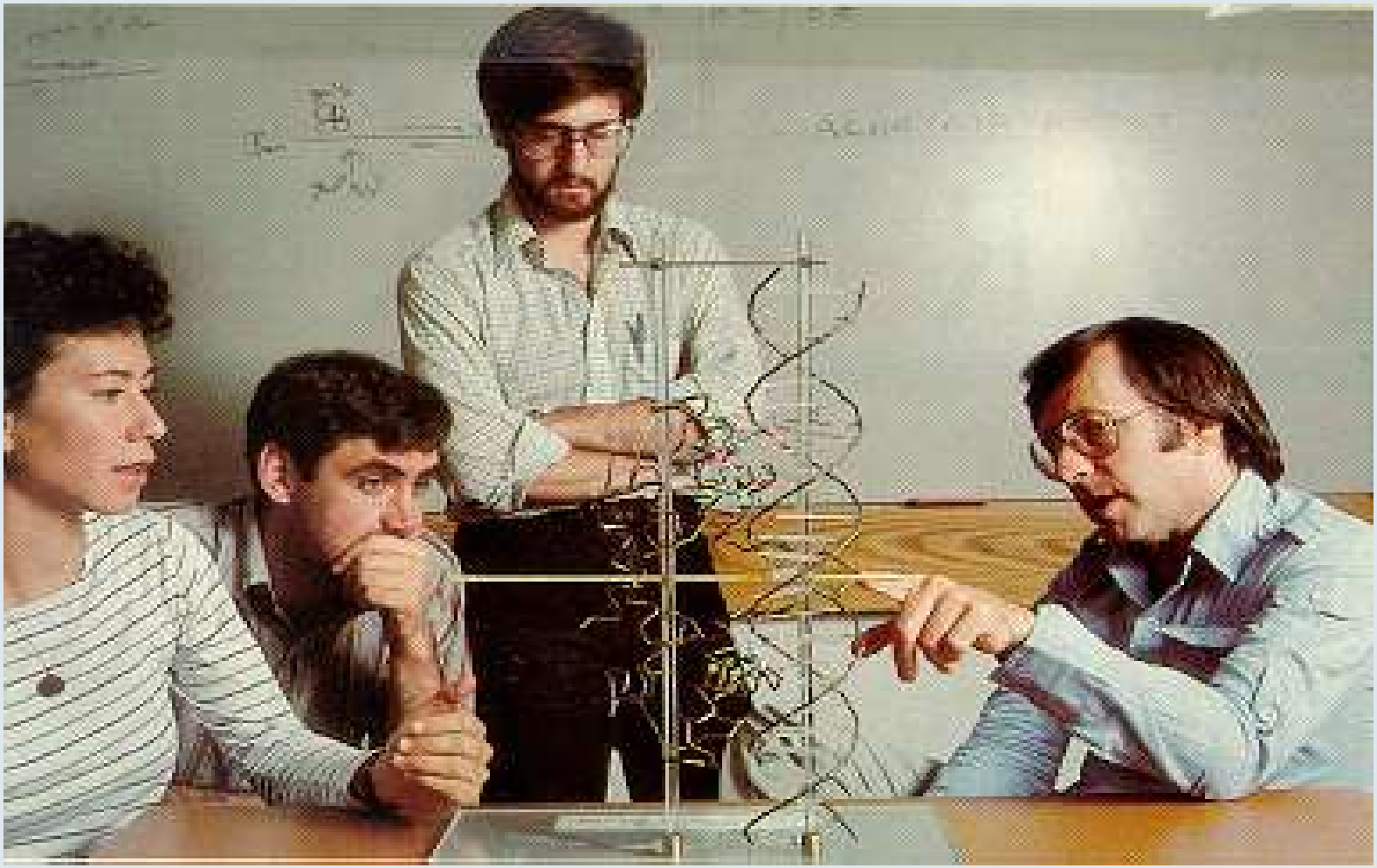
27.10 The DNA-Binding Form of the Lambda Repressor Is a Dimer



- A repressor monomer has two distinct domains.
- The N-terminal domain contains the DNA-binding site.
- The C-terminal domain dimerizes.
- Binding to the operator requires the dimeric form so that two DNA-binding domains can contact the operator simultaneously.
- Cleavage of the repressor between the two domains reduces the affinity for the operator and induces a lytic cycle.

Figure 27.18: Repressor dimers bind to the operator.

Sequence-specific DNA recognition is the key to differential gene regulation.



Mark Ptashne with graduate students Cynthia Wohlberger, Liam Keegan, Ed Giniger at Harvard, 1982

Need to purify scarce gene regulator proteins drove biotechnology of protein overexpression

- Take *E. coli* cell as a cube 1 micron on each side. A liter is a cube 10 cm on each side. What is the concentration of the lac operator? (Answer. nanomolar; about 10^{-9} M). 10-100 lac repressor molecules per cell and the k_d for operator-binding is 10^{-10} M or lower (tighter).

- Lambda repressor was first isolated from an overproducer mutant virus. Later the repressor gene was expressed from *lac* or *tac* promoters. Also, a hybrid ribosome binding site (Shine-Dalgarno sequence), upstream of ATG gave strong translation initiation.

27.11 Lambda Repressor Uses a Helix-Turn-Helix Motif to Bind DNA

- Each DNA-binding region in the repressor contacts a half-site in the DNA.
- The DNA-binding site of the repressor includes two short α-helical regions that fit into the successive turns of the major groove of DNA (**helix-turn-helix**).
- A DNA-binding site is a (partially) palindromic sequence of 17 bp.

Figure 27.19: The operator is a 17-bp sequence with an axis of symmetry through the central base pair.

Lambda repressor binds as a symmetrical protein dimer to a nearly symmetrical 17 bp DNA sequence.

Symmetric repressor site

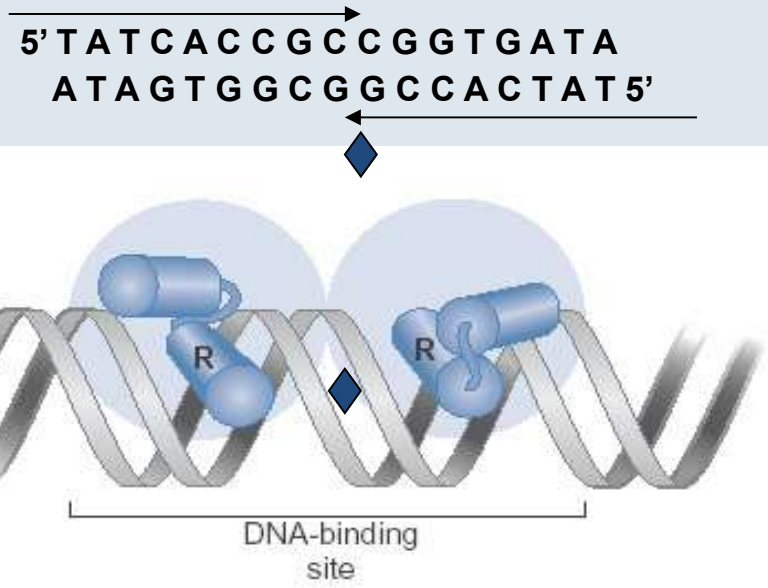


FIGURE 16-11 Binding of a protein with a helix-turn-helix domain to DNA.
The protein, as is typically the case, binds as a dimer, and the two subunits are indicated by the shaded circles. The helix-turn-helix motif on each monomer is indicated; the "recognition helix" is labeled R.

Recognition helix amino acid side-chains make many sequence-specific contacts to base pairs.

5' T A T C A C C G C C G G T G A T A
A T A G T G G C G G C C A C T A T 5'

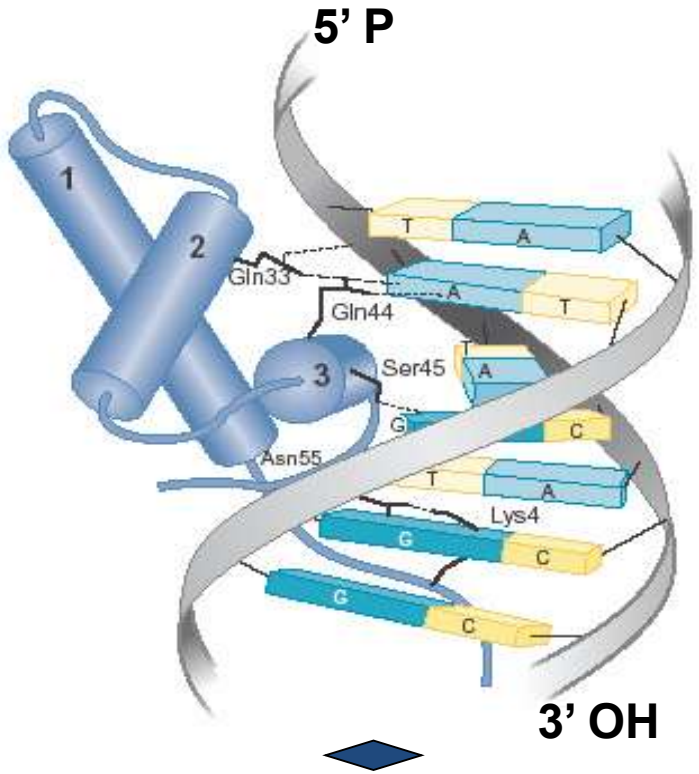
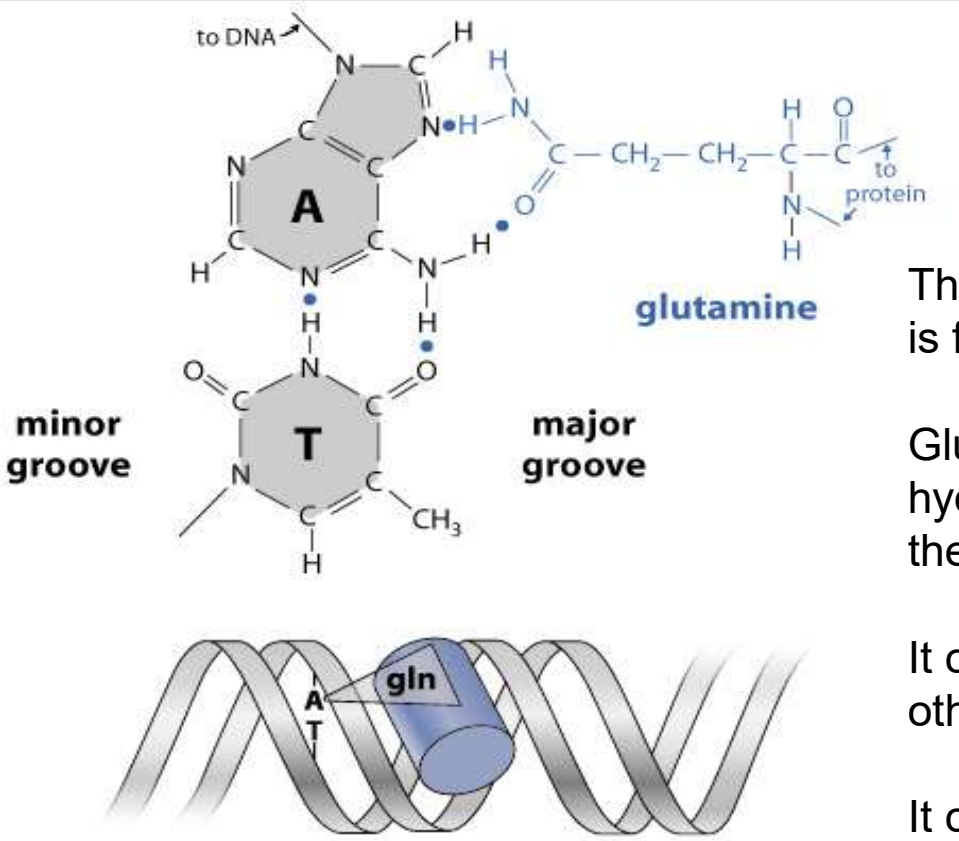


FIGURE 16-12 Hydrogen bonds between λ repressor and base pairs in the major groove of its operator. Diagram of the repressor-operator complex, showing hydrogen bonds (in dotted lines) between amino acid side chains and bases in the consensus half-site. Only the relevant amino acid side chains are shown. In addition to Gln44 and Ser45 in the recognition helix, Asn55 in the loop following the recognition helix also makes contact with a specific base. Furthermore (and unusual to this case, see later in the text) Lys4 in the N-terminal arm of the protein makes a contact in the major groove on the opposite face of the DNA helix. Gln33 contacts the backbone. (Source: Redrawn from Jordan, S. and Pabo, C. *Science* 242: 896, Fig. 3B.)

Example of an amino acid contact that can discriminate between bases in a binding site.



A Genetic Switch, 3rd edition, 2004
© Cold Spring Harbor Laboratory Press
Chapter 2, Figure 12

The position of the glutamine is fixed when repressor binds.

Glutamine side chain needs to find the hydrogen donor and acceptor sites in just the right places.

It cannot reach the paired base on the other strand.

It cannot reach the other bases on the same strand.

Only an A base meets the criteria.

Transcription activation in multicellular organisms.

Liam.Keegan@ceitec.muni.cz

Eukaryotic total RNA.

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

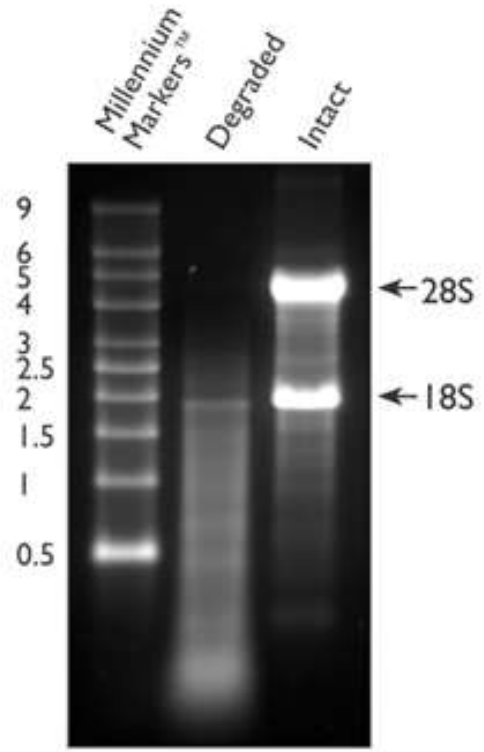
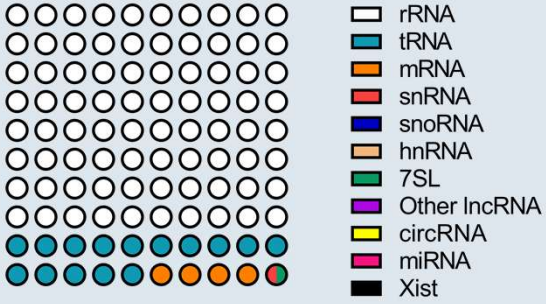
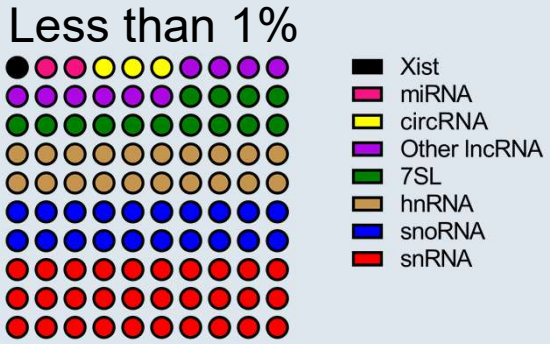


Figure 1. Intact vs. Degraded RNA. Two µg of degraded total RNA and intact total RNA were run beside Ambion's RNA Millennium Markers™ on a 1.5% denaturing agarose gel. The 18S and 28S ribosomal RNA bands are clearly visible in the intact RNA sample. The degraded RNA appears as a lower molecular weight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

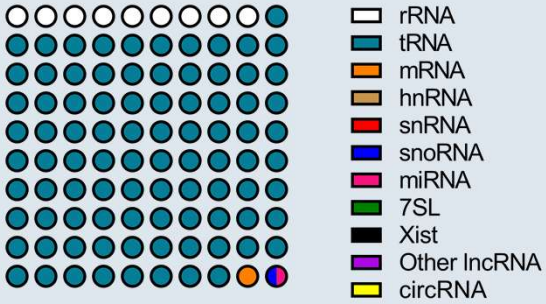
Total RNA

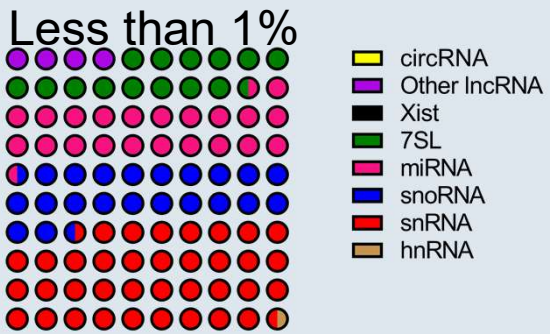
RNA by mass





RNA by number of molecules





Adapted from doi: 10.3389/fgene.2015.00002

• Eukaryotic RNA Polymerases

- Three nuclear RNA polymerases (*E. coli* has only one).
-
- RNA pol I (in nucleolus)
- transcribes rRNA genes → 50 – 70 % cell's RNA synthesis
- resistant to > 500 μg/ml α-amanitin, an octapeptide from
- *Amanita phalloides* (Death Cap Mushroom) that grows near Oak trees.
-
- RNA pol II (in nucleoplasm)
- transcribes all protein-encoding genes & most small nuclear RNAs
- → 20 – 40% cell's RNA synthesis
- inhibited by low ~ 0.03 μg/ml α-amanitin
-
- RNA pol III (in nucleoplasm)
- transcribes 5S, tRNA genes & some small nuclear RNAs
- → < 10% cell's RNA synthesis
- inhibited by 20 μg/ml α-amanitin in animal cells
- resistant to α-amanitin in yeast and insects

20.2 Eukaryotic RNA Polymerases Consist of Many Subunits

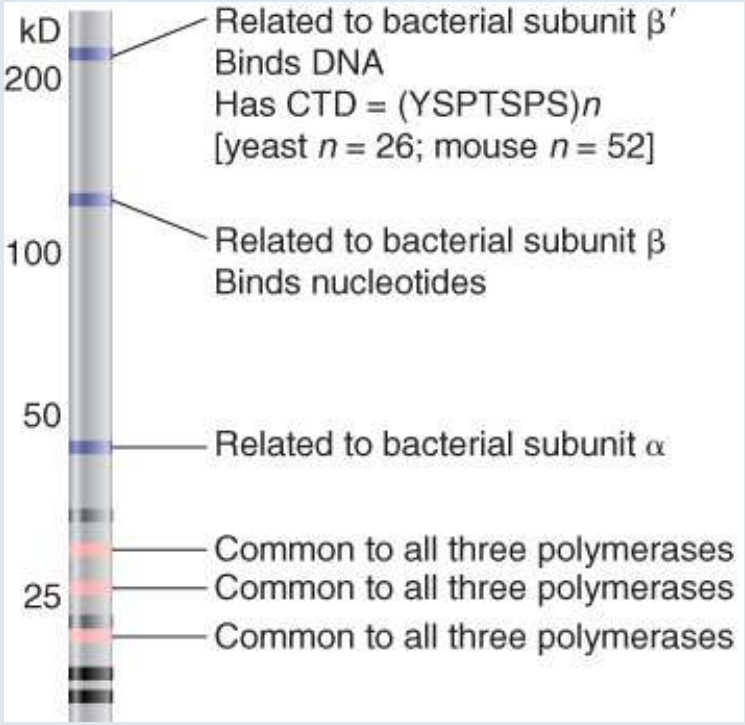
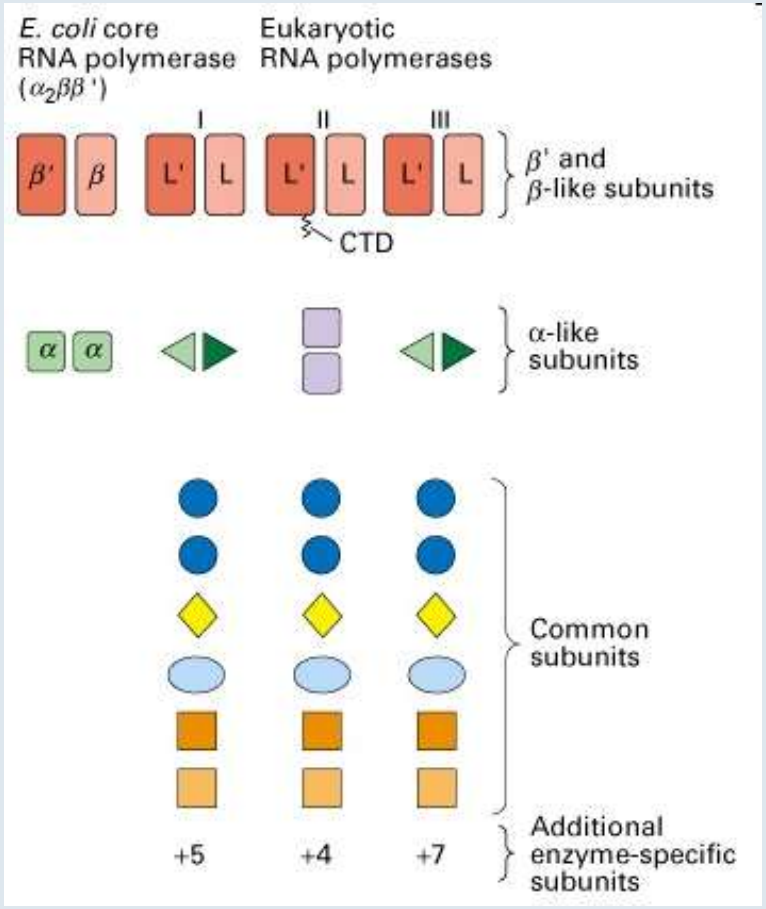


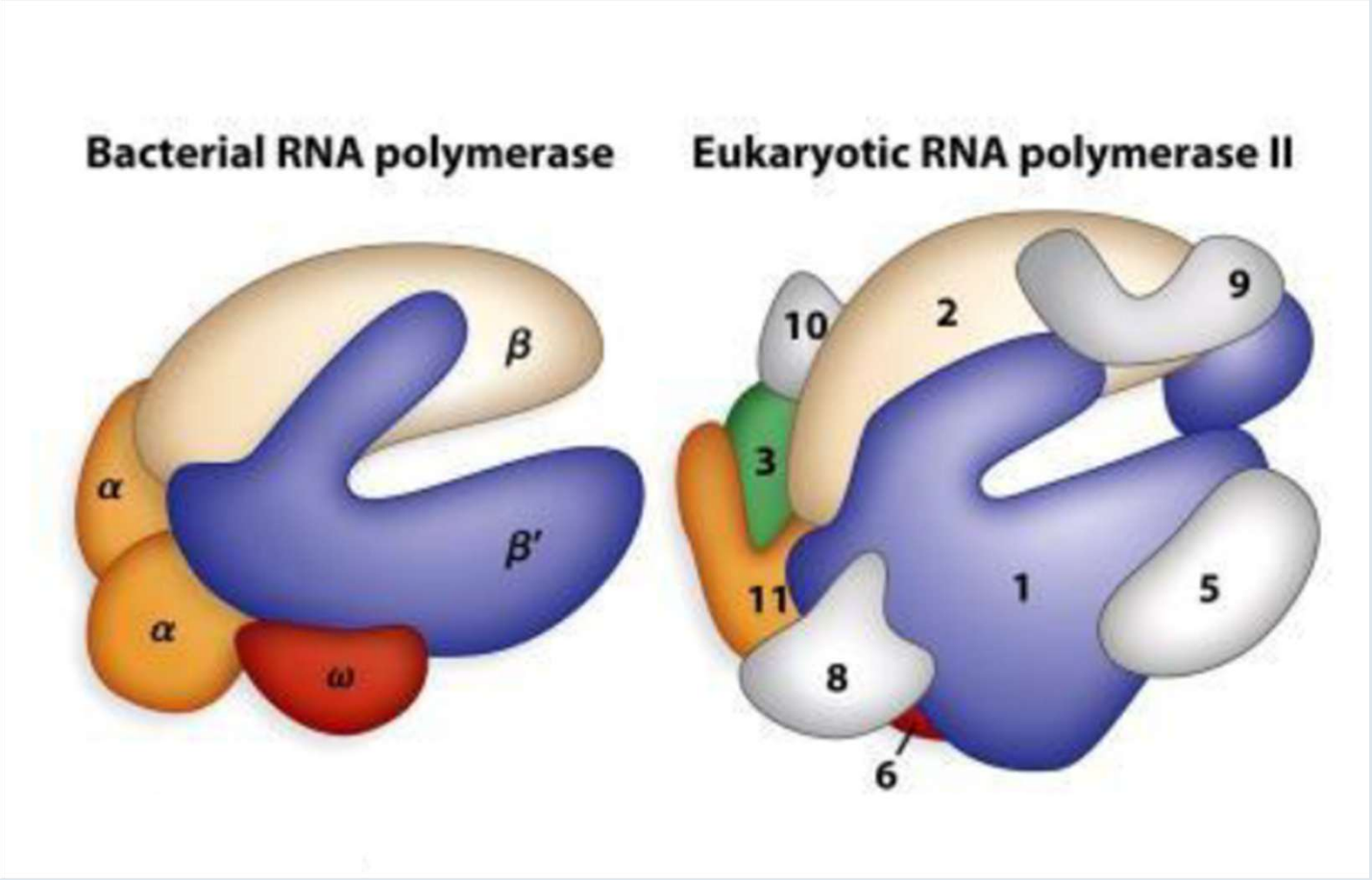
Figure 20.02: Some subunits are common to all classes of eukaryotic RNA polymerases and some are related to bacterial RNA polymerase.

- All eukaryotic RNA polymerases have ~12 subunits and are complexes of ~500 kD.
- Some subunits are common to all three RNA polymerases.
- The largest subunit in RNA polymerase II has a **CTD (carboxy-terminal domain)** consisting of multiple repeats of a heptamer.

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

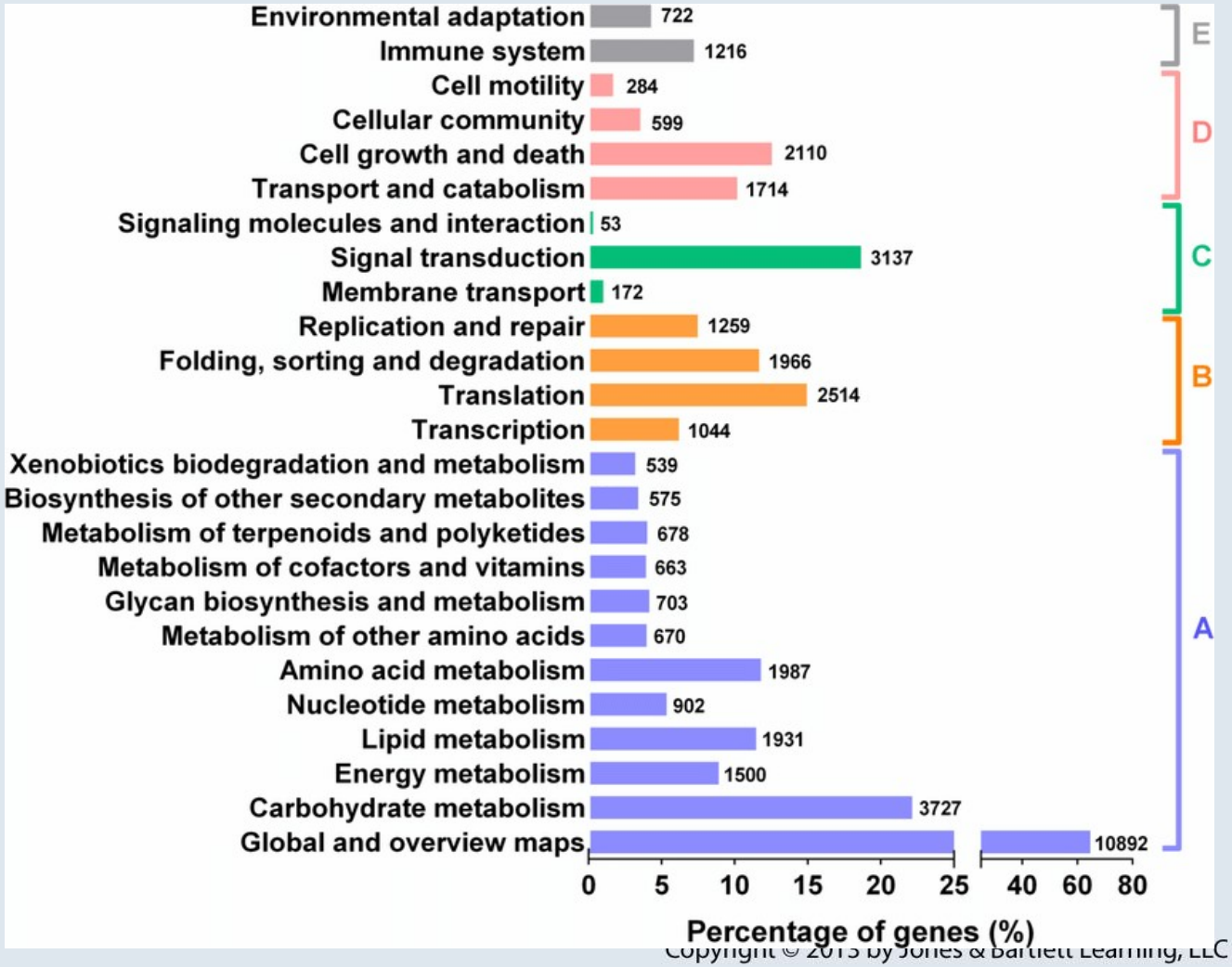


Conserved and similar subunits between eukaryotic RNA pol II and *E. coli* RNA polymerase

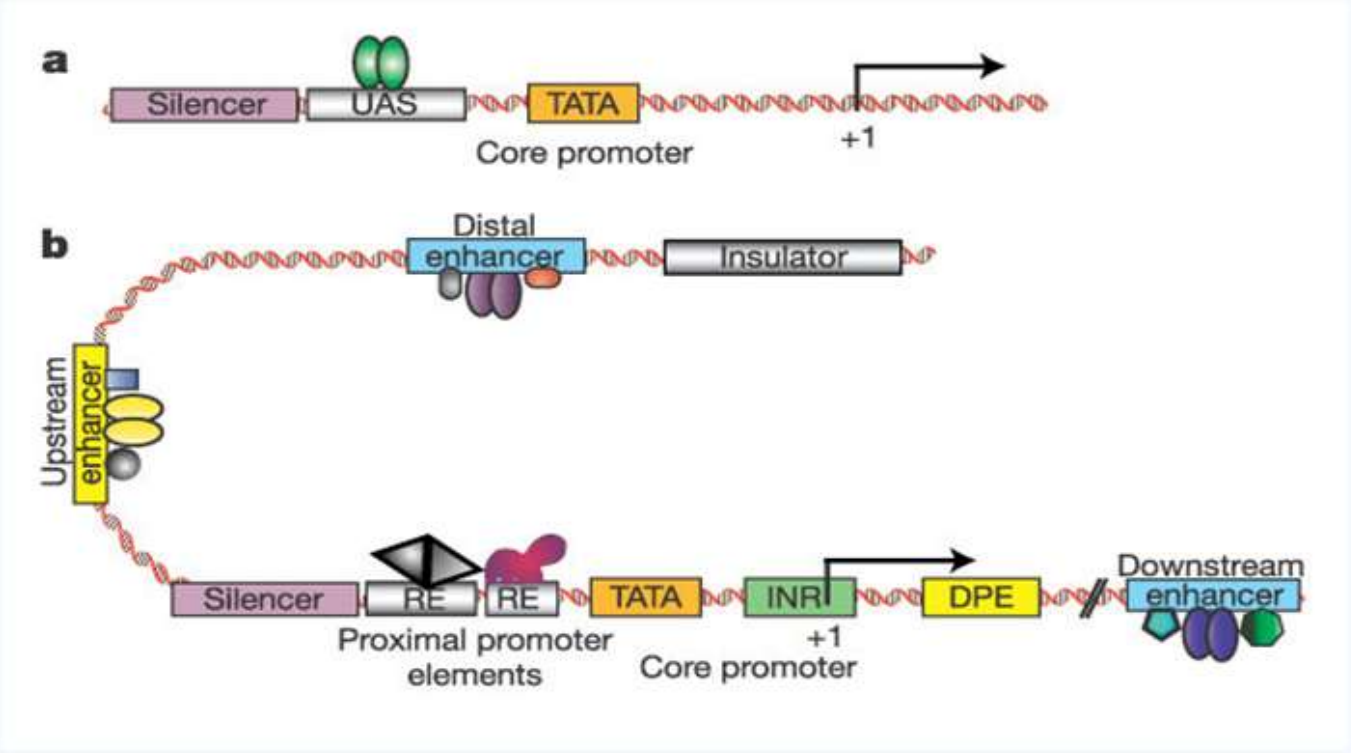


Housekeeping genes, metabolic genes (60%) (in blue). Inducible, developmental, tissue-specific genes (some of the orange, green, pink).

Kyoto Encyclopedia of Genes and Genomes (KEGG)



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.

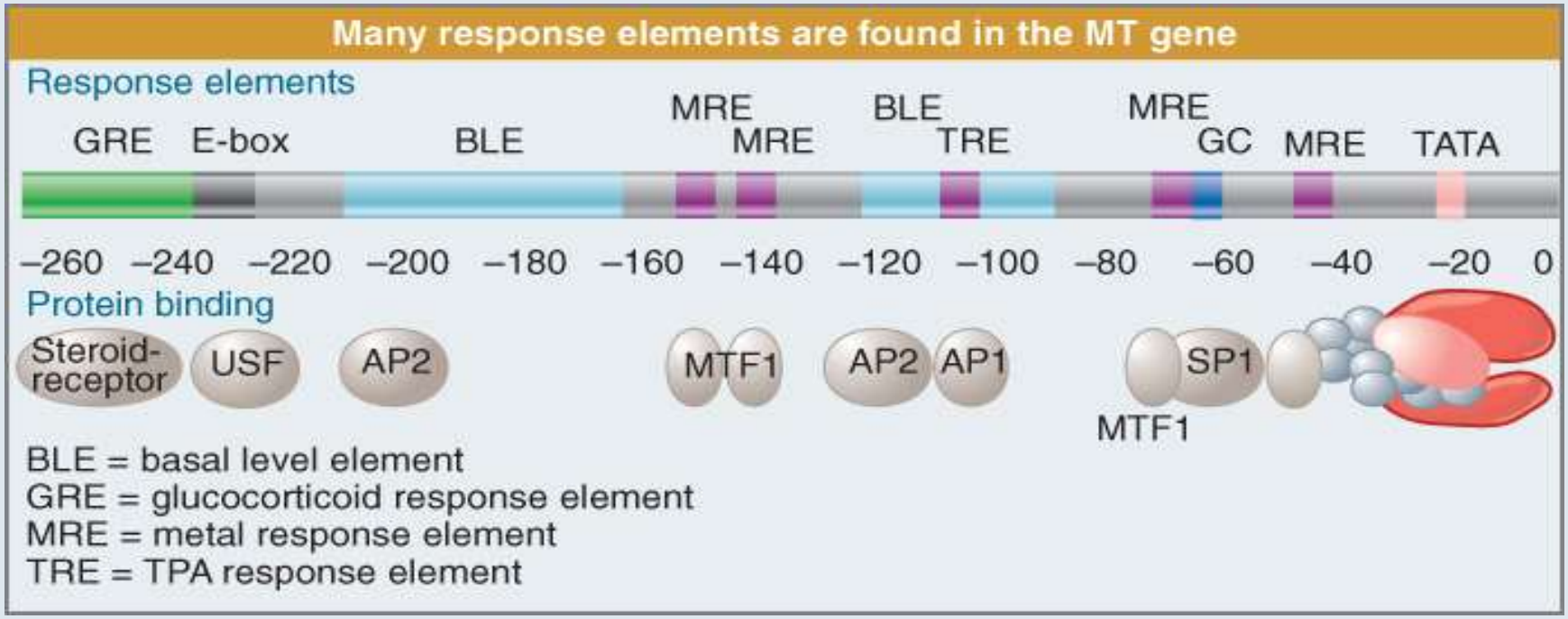


Figure 25.11

Transcription of protein-coding mRNAs by RNA polymerase II.

Defining RNA polymerase II promoters.

20.5 **The Start Point for RNA Polymerase II**

- RNA polymerase II requires general transcription factors (called $TF_{II}X$) to initiate transcription.
- RNA polymerase II promoters frequently have a short conserved sequence Py_2CAPy_5 (the **initiator Inr**) at the start point.
- The **TATA box** is a common component of RNA polymerase II promoters and consists of an A-T-rich octamer located ~25 bp upstream of the start point.

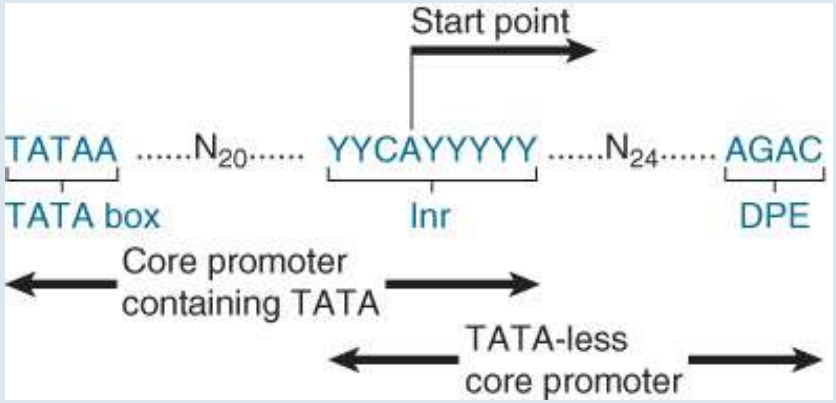


Figure 20.07: A minimal pol II promoter may have a TATA box ~25 bp upstream of the Inr.

20.5 **The Start Point for RNA Polymerase II**

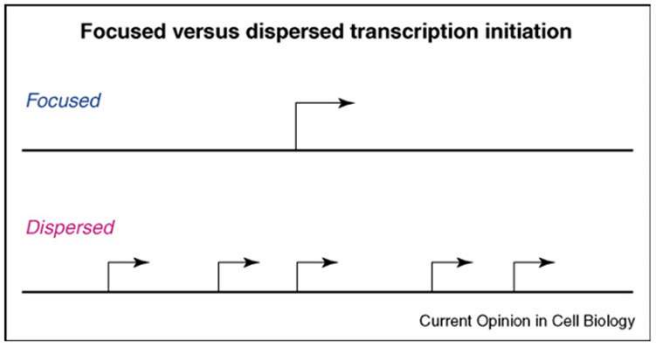
- The **downstream promoter element (DPE)** is a common component of RNA polymerase II promoters that do not contain a TATA box (**TATA-less promoters**).
- A core promoter for RNA polymerase II includes the Inr and, commonly, either a TATA box or a DPE.
  - It may also contain other minor elements.

Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Cell Biology
ELSEVIER

The RNA polymerase II core promoter — the gateway to transcription

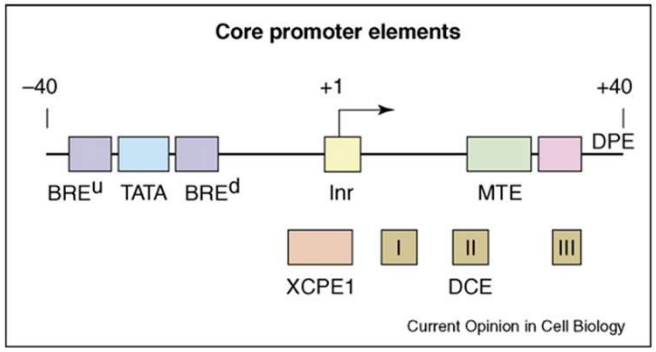
Tamar Juven-Gershon, Jer-Yuan Hsu, Joshua WM Theisen and James T Kadonaga

Figure 1



Focused versus dispersed core promoters. In focused core promoters, transcription initiates at a single site or in a cluster of sites in a narrow region of several nucleotides. Dispersed core promoters are typically found in CpG islands in vertebrates and usually yield multiple weak start sites over a region of 50–100 nucleotides. Focused core promoters are more ancient and widespread throughout nature than dispersed core promoters. In vertebrates, however, dispersed promoters are more common than focused promoters. There may be fundamental differences in the basic mechanisms of transcription from focused versus dispersed core promoters.

Figure 2



Core promoter motifs. This diagram, which is drawn roughly to scale, shows some of the known core promoter elements for transcription by RNA polymerase II. There are no universal core promoter elements. Each of these elements is found in only a fraction (typically estimated to be from 1% to 30%, depending on the motif) of all core promoters. The Inr is probably the most commonly occurring core promoter motif. There are additional core promoter elements that remain to be discovered. The TATA box, Inr, MTE, DPE, and DCE are recognition sites for the binding of transcription factor TFIID. It should be noted, however, that there are multiple forms of TFIID and TFIID-related protein complexes that could potentially interact with the core promoter. BRE^u and BRE^d interact with TFIIB.

In vertebrate promoters

In *Drosophila* promoters

The RNA polymerase II core promoter: a key component in the regulation of gene expression

Jennifer E.F. Butler[1] and James T. Kadonaga[2,3,4]

[1]Vollum Institute, Oregon Health Sciences University, Portland, Oregon 97201, USA; [2]Section of Molecular Biology, University of California, San Diego, La Jolla, California 92093, USA

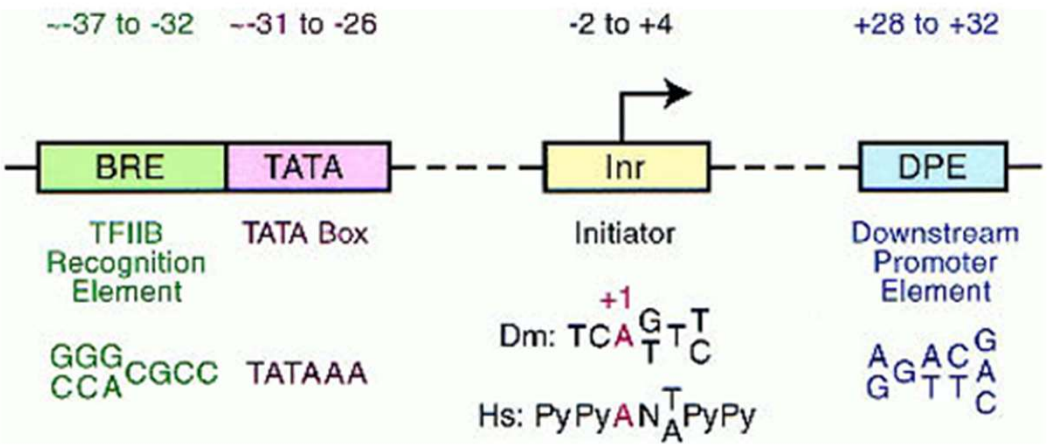


Figure 1. Core promoter elements. Some core promoter motifs that can participate in transcription by RNA polymerase II are depicted. Each of these elements is found in only a subset of core promoters. Any specific core promoter may contain some, all, or none of these motifs. The BRE is an upstream extension of a subset of TATA boxes. The DPE requires an Inr, and is located precisely at +28 to +32 relative to the A_{+1} nucleotide in the Inr. The DPE consensus was determined with *Drosophila* transcription factors and core promoters. The Inr consensus sequence is shown for both *Drosophila* (Dm) and humans (Hs).

Rational design of a super core promoter that enhances gene expression

Tamar Juven-Gershon, Susan Cheng & James T Kadonaga

Transcription is a critical component in the expression of genes. Here we describe the design and analysis of a potent core promoter, termed super core promoter 1 (SCP1), which directs high amounts of transcription by RNA polymerase II in metazoans. SCP1 contains four core promoter motifs—the TATA box, initiator (Inr), motif ten element (MTE) and downstream promoter element (DPE)—in a single promoter, and is distinctly stronger than the cytomegalovirus (CMV) IE1 and adenovirus major late (AdML) core promoters both *in vitro* and *in vivo*. Each of the four core promoter motifs is needed for full SCP1 activity. SCP1 is bound efficiently by TFIID and exhibits a high propensity to form productive transcription complexes. SCP1 and related super core promoters (SCPs) with multiple core promoter motifs will be useful for the biophysical analysis of TFIID binding to DNA, the biochemical investigation of the transcription process and the enhancement of gene expression in cells.

enhancers that function via the TATA box[8,9]. A preference for activators to function with either a TATA box or an Inr has also been observed[10,11]. Moreover, canonical versus noncanonical TATA box elements can exhibit distinct properties (for reviews, see refs. 2,4).

In this study, we explored the use of the core promoter to achieve high levels of transcription by RNA polymerase II. To this end, we created and characterized an optimized core promoter, SCP1, which contains TATA, Inr, MTE and DPE motifs, and exhibits high transcriptional activity both *in vitro* and *in vivo*. SCP1 should be a versatile core promoter in eukaryotic expression vectors as well as a high-affinity TFIID recognition site and potent core promoter for the biochemical and biophysical analysis of transcription.

RESULTS

Creation of SCP1

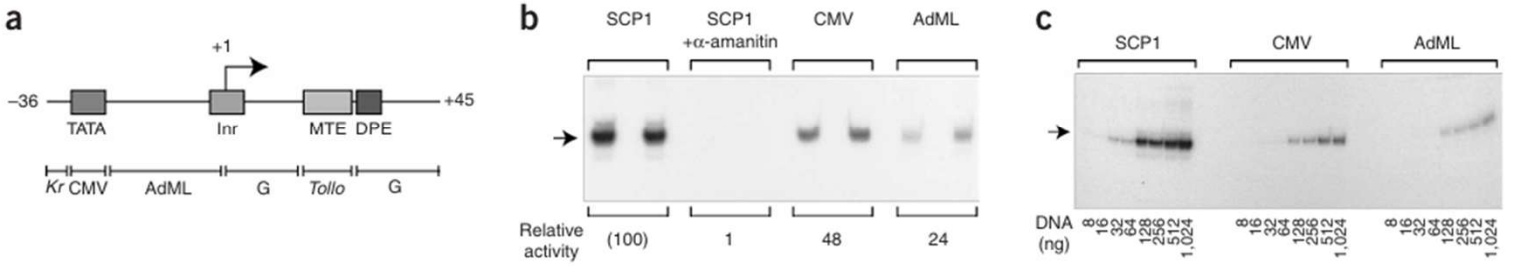


Figure 1 | SCP1 is stronger than CMV and AdML core promoters *in vitro*. (**a**) Diagram of SCP1. *Kr*, *D. melanogaster Krüppel* gene; G, *Drosophila* G retrotransposon. The nucleotide sequence is given in Methods. (**b**) SCP1, CMV and AdML core promoters (each of which contains their respective sequences from −36 to +45 relative to the +1 start site cloned into pUC119) were subjected to *in vitro* transcription analysis (in duplicate reactions) with a standard HeLa transcription extract[21]. The resulting transcripts were detected by primer extension. To test whether transcription was catalyzed by RNA polymerase II, α-amanitin (4 μg/ml) was included, as indicated. (**c**) Transcription reactions were performed as in **b** with the indicated amounts of template DNA in a volume of 50 μl.

A nuclear extract from HeLa cells that gave accurate initiation at Pol II promoters was fractionated by ion-exchange column chromatography. (M. Sawadogo and R. Roeder)

Increasing salt concentration

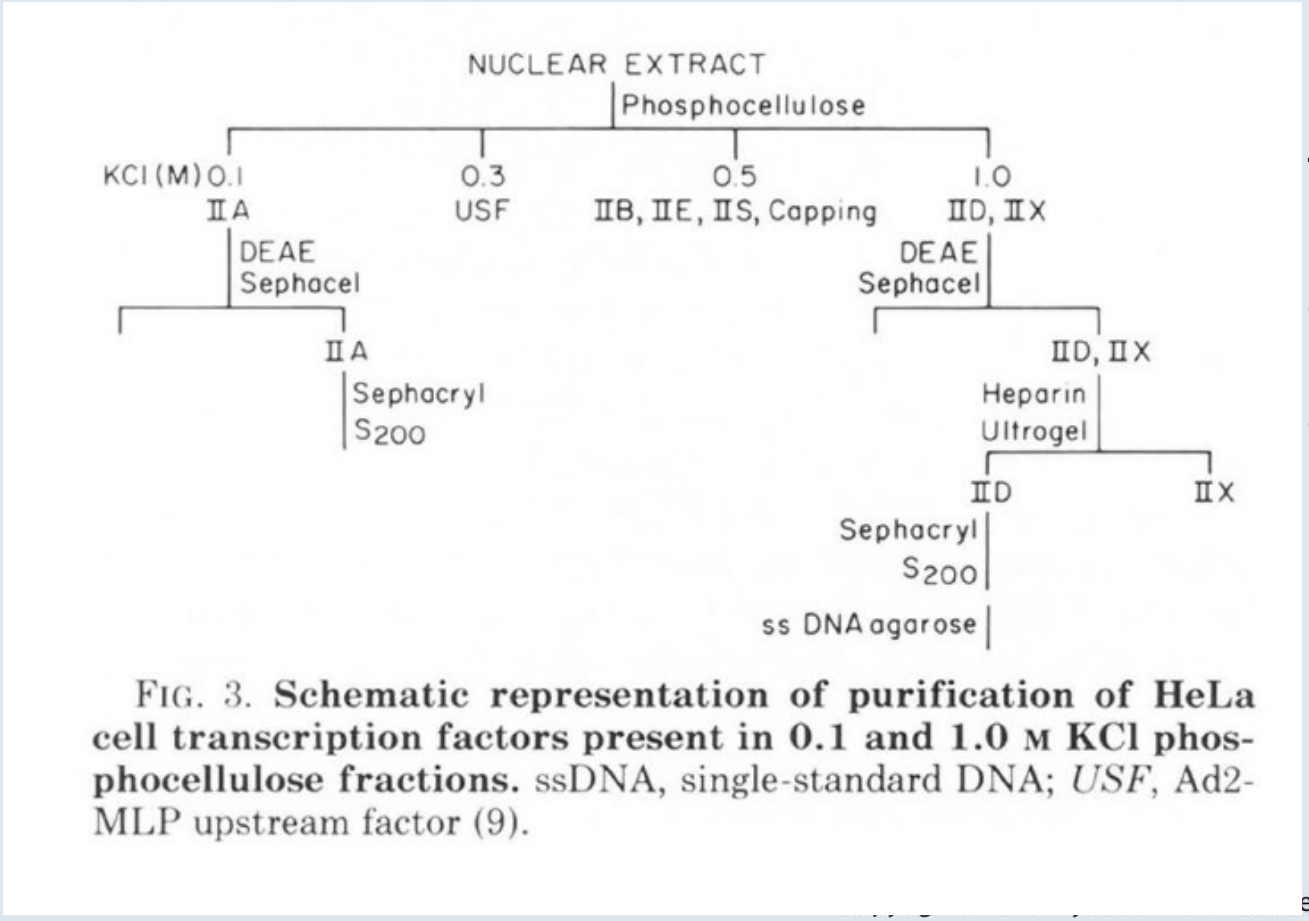


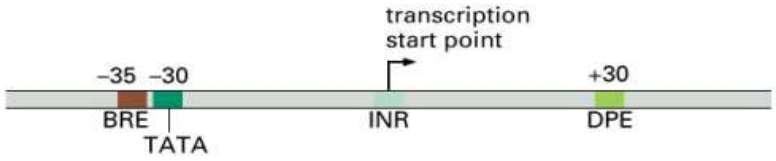
FIG. 3. **Schematic representation of purification of HeLa cell transcription factors present in 0.1 and 1.0 M KCl phosphocellulose fractions.** ssDNA, single-standard DNA; *USF*, Ad2-MLP upstream factor (9).

- Charge on column

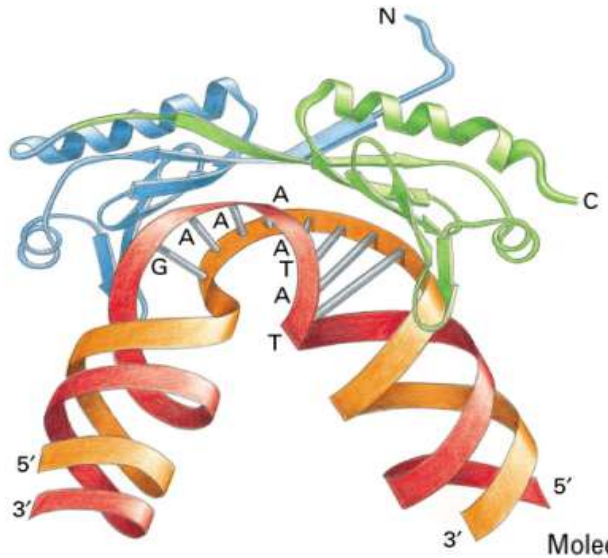
+ charge on column

+ charge on column

TATA-binding protein (TBP) and 12 TBP associated factors (TAFs) comprise TFIID.



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

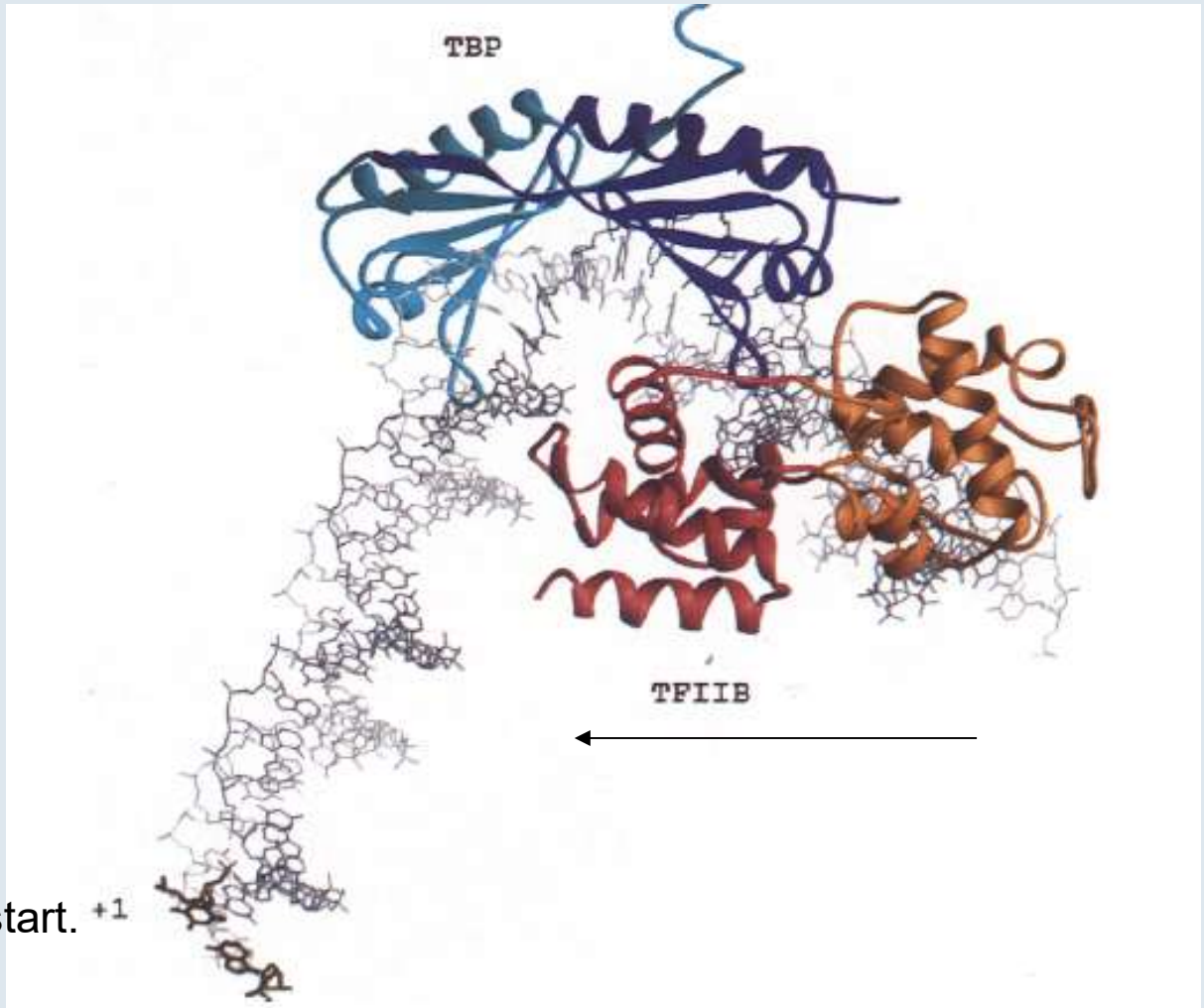


TBP

TATA

Molecular Biology of the Cell, 4th Edition.
By Alberts

TFIIB binds asymmetrically and sets direction of transcription.



Transcription start.

Figure 3. [*A*] Crystal structure (*left*) and solution structure

20.6 **TBP Is a Universal Factor**

- TBP binds to the TATA box in the minor groove of DNA.
- TBP forms a saddle around the DNA and bends it by ~80°.

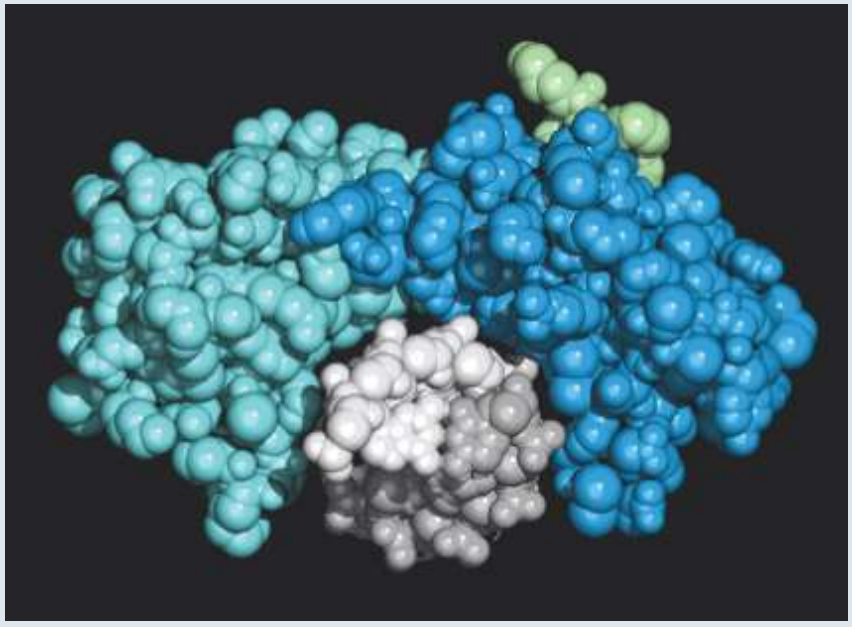
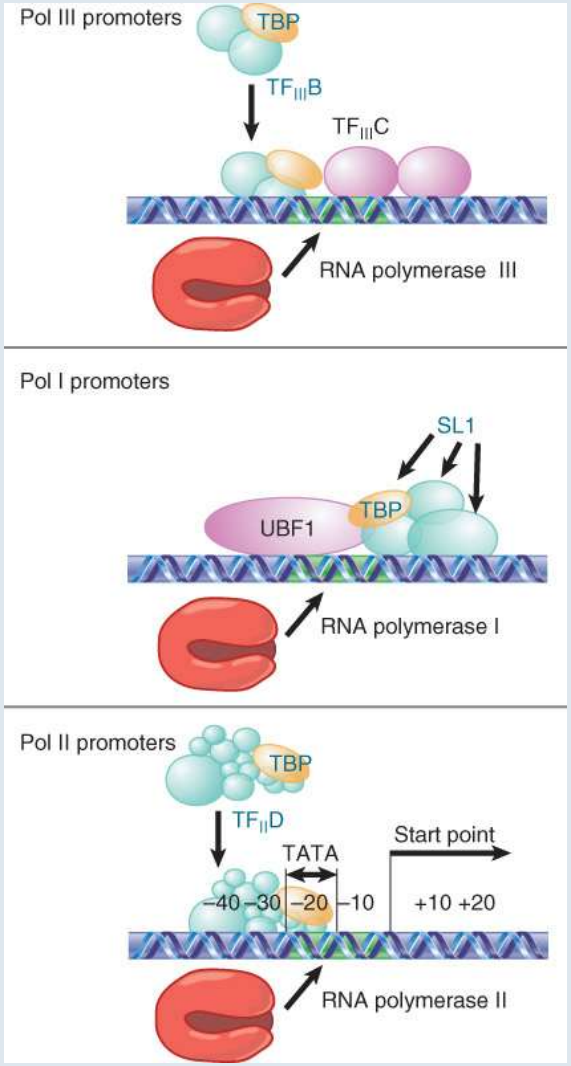


Figure 20.09: A view in cross-section shows that TBP surrounds DNA from the side of the narrow groove.

Photo courtesy of Stephen K. Burley

20.6 General transcription factors (GTFs) TBP in TFIID Is a Universal Factor



- TATA-binding protein (TBP) is a component of each of the different positioning factors required for each type of RNA polymerase to bind its promoter.
- The factor for RNA polymerase II is **$TF_{II}D$**, which consists of TBP and ~14 **TAFs**, (TBP-associated factors) with a total mass ~800 kD.

Figure 20.08: RNA polymerases are positioned at all promoters by a factor that contains TBP.

20.7 **The Basal Apparatus Assembles at the Promoter**

- The upstream elements and the factors that bind to them increase the frequency of initiation.
- Binding of $TF_{II}D$ to the TATA box or Inr is the first step in initiation.

20.7 The Basal Apparatus Assembles at the Promoter

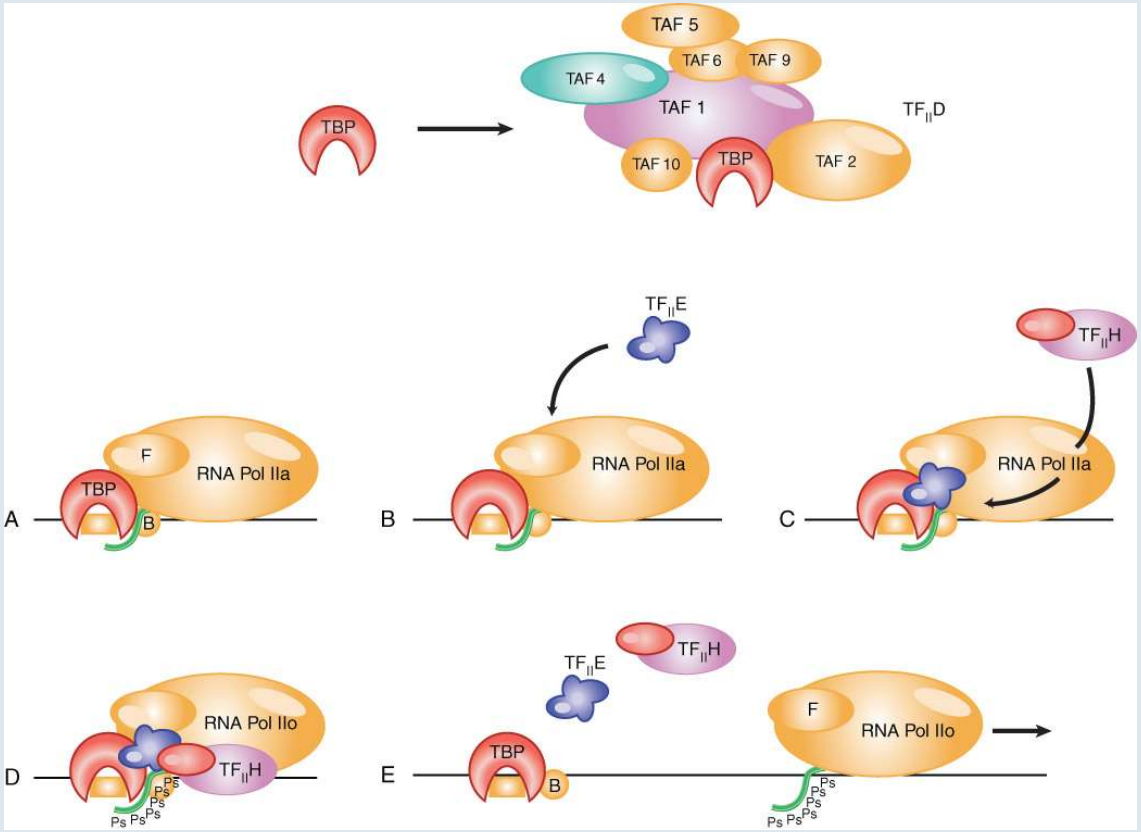


Figure 20.11: An initiation complex assembles at promoters for RNA polymerase II by an ordered sequence of association with transcription factors.

Adapted from M. E. Maxon, J. A. Goodrich, and R. Tijan, Genes Dev. 8 (1994): 515-524.

TFIID binds Inr-DPE and drops off TBP at TATA box

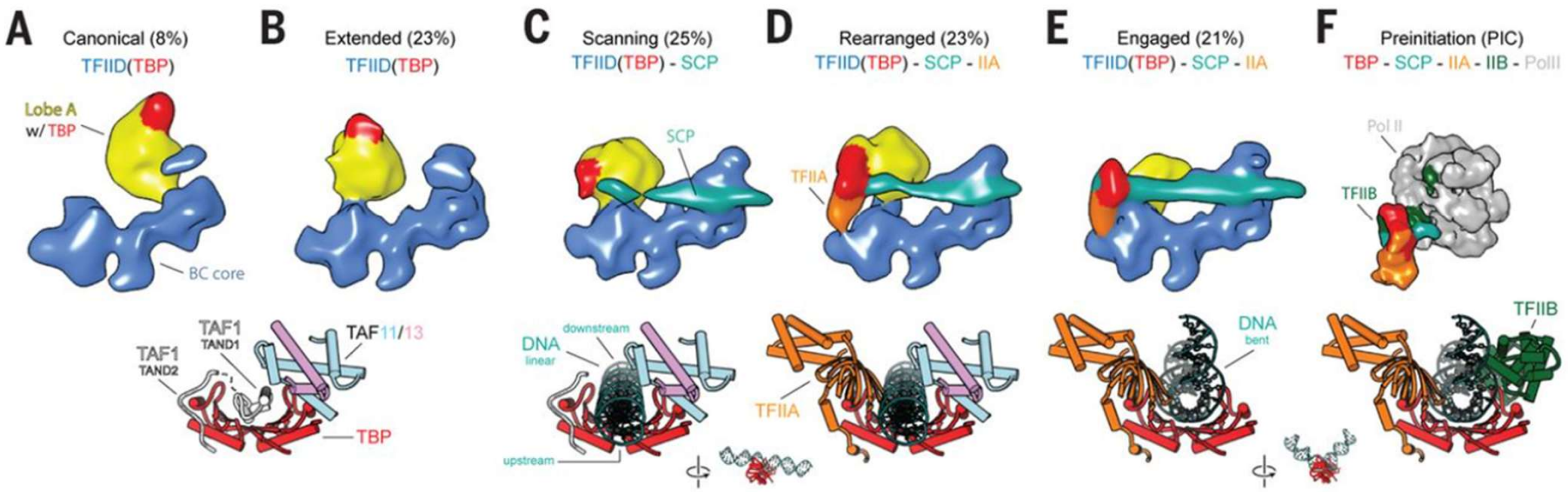


Fig. 4 Regulation of TBP DNA-binding activity by lobe A.
Reconstructions of TFIID from the mixed dataset (which includes SCP and TFIIA), showing TFIID in the canonical (**A**), extended (**B**), scanning (**C**), rearranged (**D**), and engaged (**E**) states. (**F**) Human PIC cryo-EM map (EMD-2304) containing Pol II, TFIIA, TFIIB, TBP, and promoter DNA (*6*). Models for TBP (PIC: PDB 5IYA) and its interacting partners are shown below each corresponding reconstruction. See also Movie 2.

TFIID lobes A and B have nucleosome-like cores

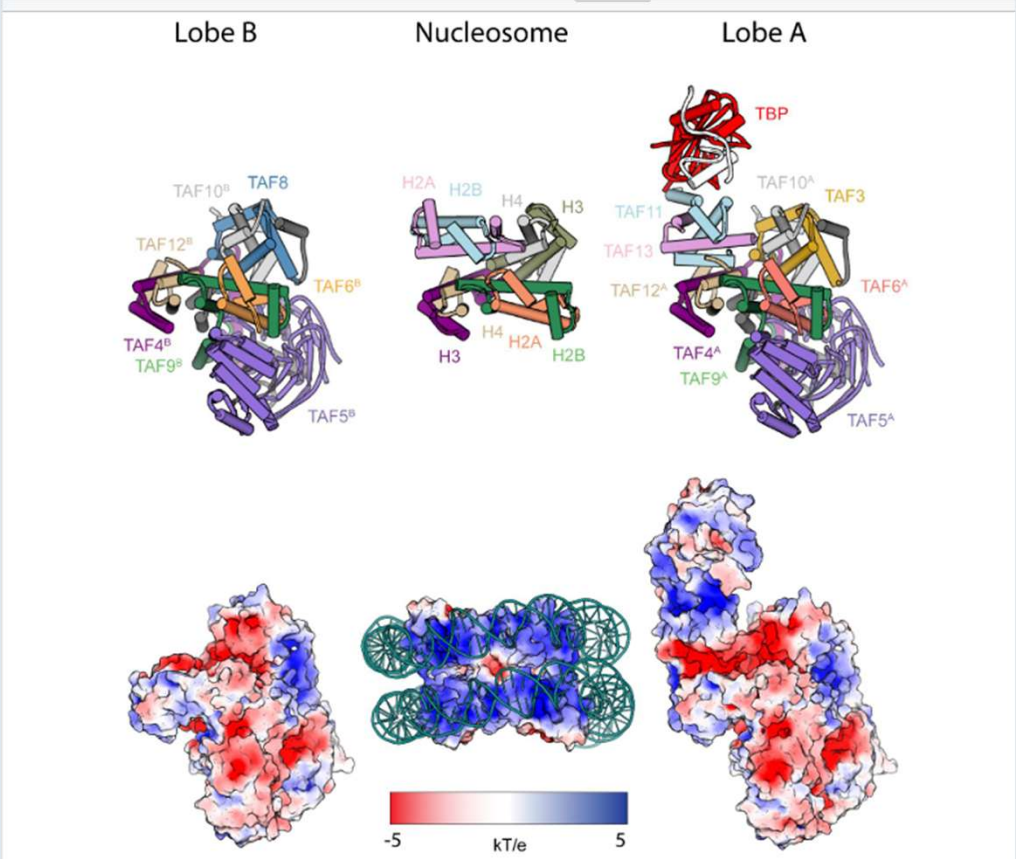
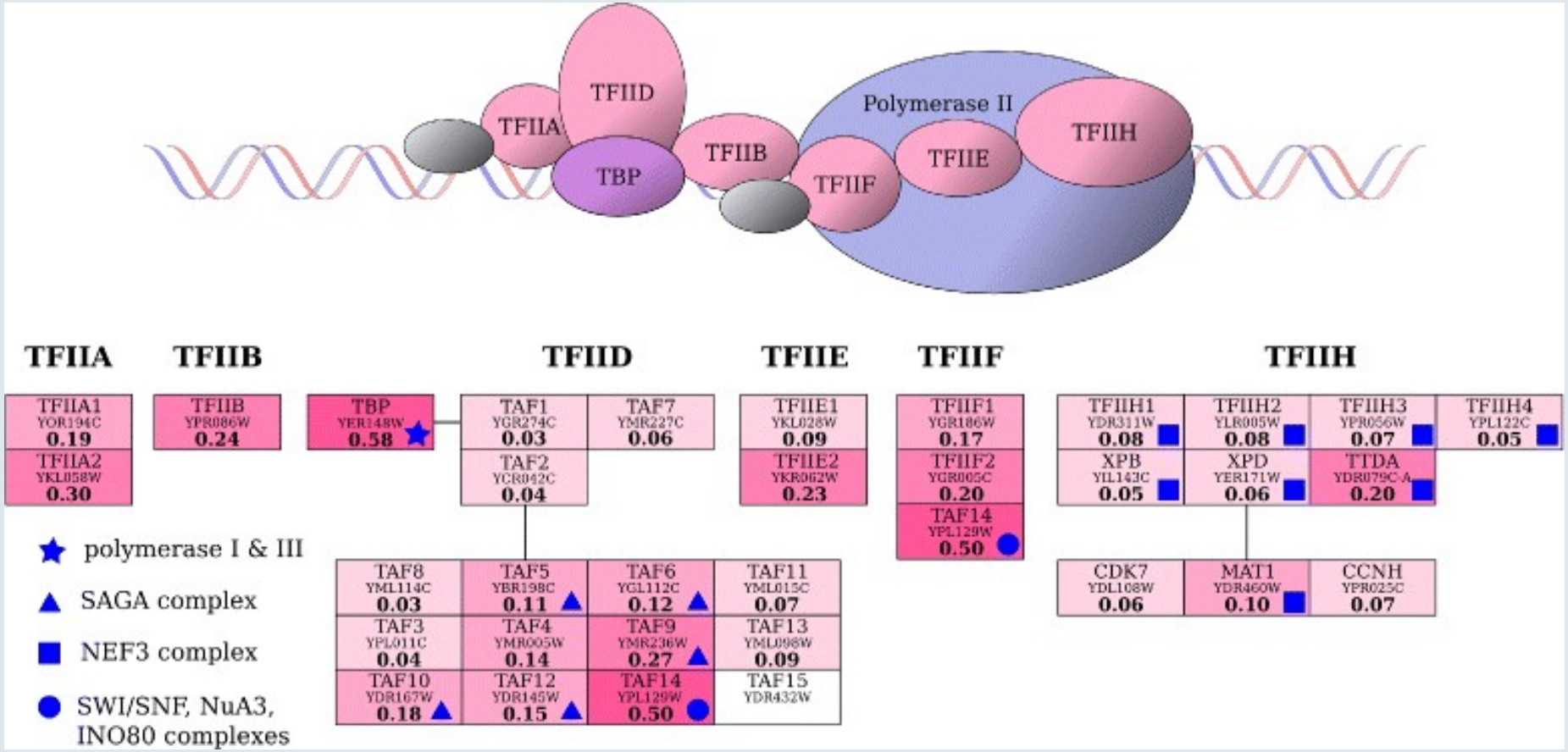


Fig. S8. Comparison of the TFIID lobe A and lobe B structures with the nucleosome core particle. Top: Rendering of the TFIID lobe A and B HFDs aligned to the structure of the histone octamer-containing nucleosome core particle (PDB ID 1AOI (*84*)). Bottom: Electrostatic surface potentials (generated using ABPS (*94*) in PYMOL; blue: positive electrostatic potential; white: neutral; red, negative electrostatic potential). The surface of the nucleosome core particle also shows the DNA wrapping around the histone octamer.

GTFs help TBP to hold promoter firmly, like *E. coli* sigma 70.

TFIIH pulls the downstream DNA back and 'scrunches' it open and into the polymerase active site.



Upstream GTFs help to hold promoter during DNA scrunching and opening by TFIIH DNA translocase

DNA contacts of GTFs

Inr, DPE, TATA
Just 5' of TATA

BRE$^{U, D}$ -38 -32, -23 -17
-23 to -13

-14 to -2

Table 1. Summary of the general transcription factors and RNA polymerase II.

Protein	Subunits	Size (kDa)	Main Binding Partners	Function
TFIID	TBP, 13 TAFs	~1300	promoter, Pol II	Nucleates PIC assembly by binding multiple core promoter elements
TFIIA	TFIIAα, TFIIAβ, TFIIAɣ	35, 19, and 12	TBP, TFIID	Stabilizes the TFIID-DNA interaction; enhances the effects of transcriptional co-activators
TFIIB	TFIIB	33	promoter, TBP, Pol II	Helps to define the start site of transcription and orient Pol II in the proper direction
TFIIF	RAP30, RAP74	30 and 74	promoter, Pol II, GTFs	Guides Pol II to the PIC and facilitates elongation
Pol II	Rpb1–Rpb12	~514	promoter, all GTFs	Catalyzes RNA synthesis; phosphorylation of the CTD tail of Rpb1 serves a regulatory role
TFIIE	TFIIEα, TFIIEβ	56 and 34	promoter, TFIIH, Pol II, TFIIF	Recruits TFIIH to the PIC; stimulates enzymatic activities of TFIIH; stabilizes the open DNA conformation
TFIIH	Core domain: XPD, XBP, p62, p52, p44, p34, p8; CAK domain: CDK7, MAT1, cyclin H	~500	downstream DNA, TFIIE, Pol II	CDK7 kinase phosphorylates the CTD; ATP-dependent XPB translocase opens the promoter DNA

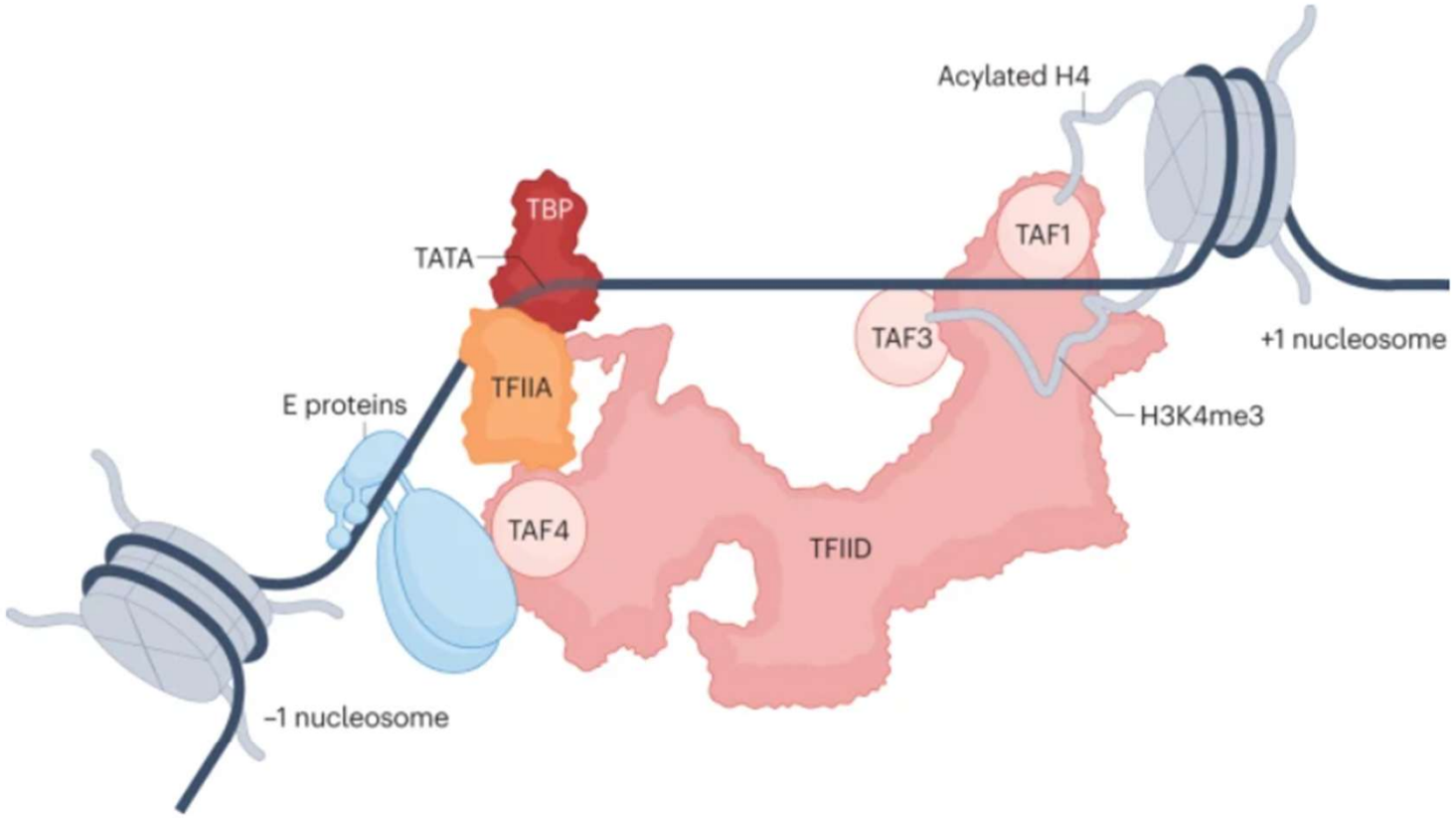
biomolecules MDPI

Review

Mechanisms and Functions of the RNA Polymerase II General Transcription Machinery during the Transcription Cycle

Stephen R. Archuleta, James A. Goodrich * and Jennifer F. Kugel *

Fig. 5: TFIID recruitment to promoters in the context of a nucleosome-depleted region.



A composite model showing multiple potential interactions of TFIID just upstream of a positioned +1 nucleosome. In addition to the promoter DNA interactions (Fig. 4), TATA binding protein (TBP)-associated factors (TAFs) in TFIID can have stabilizing interactions with an activator (such as TAF4 with an E protein) bound to a proximal element or with post-translationally modified histone tails in the +1 nucleosome. Documented tail interactions include recognition of acetylated histone H4 by TAF1 and of histone H3 trimethylated at lysine 4 (H3K4me3) by TAF3. A location for the highly mobile

20.8 **Initiation Is Followed by Promoter Clearance and Elongation**

- $TF_{II}B$, $TF_{II}E$, and $TF_{II}H$ are required to melt DNA to allow polymerase movement.
- Phosphorylation of the CTD is required for promoter clearance and elongation to begin.
- Further phosphorylation of the CTD is required at some promoters to end pausing and abortive initiation.

Promoter-proximal pausing. Elongation

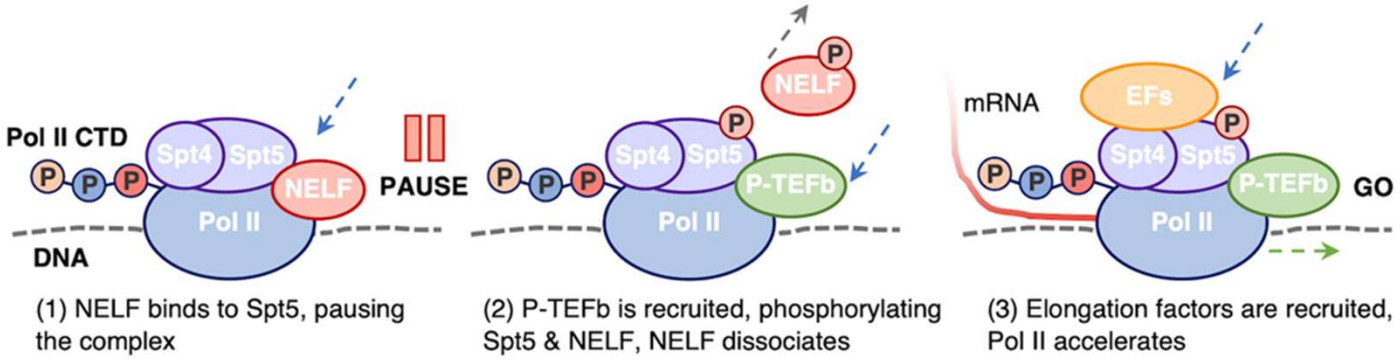


Figure 3. Promoter proximal pausing mechanism of Pol II. Around 30–100 bases downstream of the TSS, Pol II pauses transcription due to the recruitment (blue arrow) of NELF and DSIF (Spt4 and Spt5) (**1**). Pausing is released when P-TEFb is recruited (blue arrow), Spt5 and NELF are phosphorylated by the CDK9 subunit of P-TEFb, and NELF dissociates (gray arrow) (**2**). As Pol II transitions to productive elongation, elongation factors (EFs) are recruited (blue arrow), thereby increasing Pol II elongation efficiency (green arrow) (**3**).

biomolecules

MDPI

Review

Mechanisms and Functions of the RNA Polymerase II General Transcription Machinery during the Transcription Cycle

Stephen R. Archuleta, James A. Goodrich * and Jennifer F. Kugel *

Department of Biochemistry, University of Colorado Boulder, 596 UCB, Boulder, CO 80309, USA; stephen.r.archuleta@colorado.edu

* Correspondence: james.goodrich@colorado.edu (J.A.G.); jennifer.kugel@colorado.edu (J.F.K.)

other important Pol II CTD residues provides numerous areas for future study.

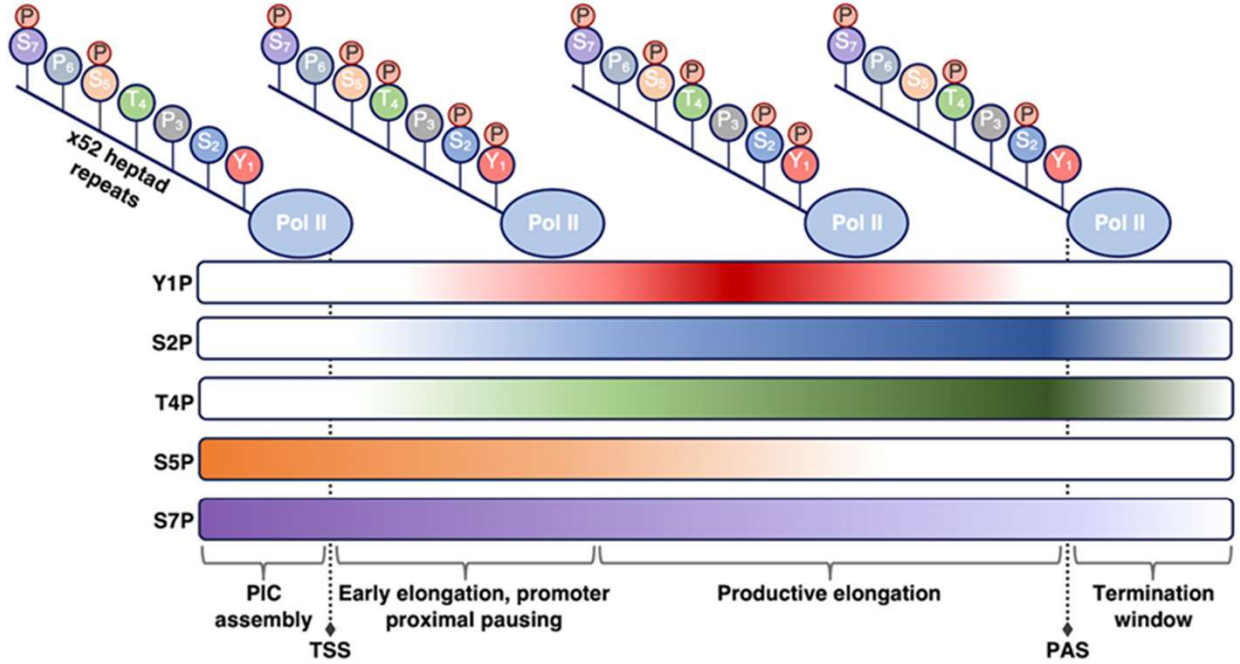


Figure 1. Phosphorylation state of the Pol II CTD is regulated during transcription. As Pol II transcribes through a gene and progresses through the stages of transcription (shown from left to right), different phosphorylation marks are added or removed to promote unique functions. The phosphorylation patterns shown here pertain to human Pol II; other organisms may exhibit slight differences in these patterns. TSS, transcription start site; PAS, polyadenylation site.

Further rounds of transcription initiation are from an activator-dependent reinitiation complex that remains at promoter.

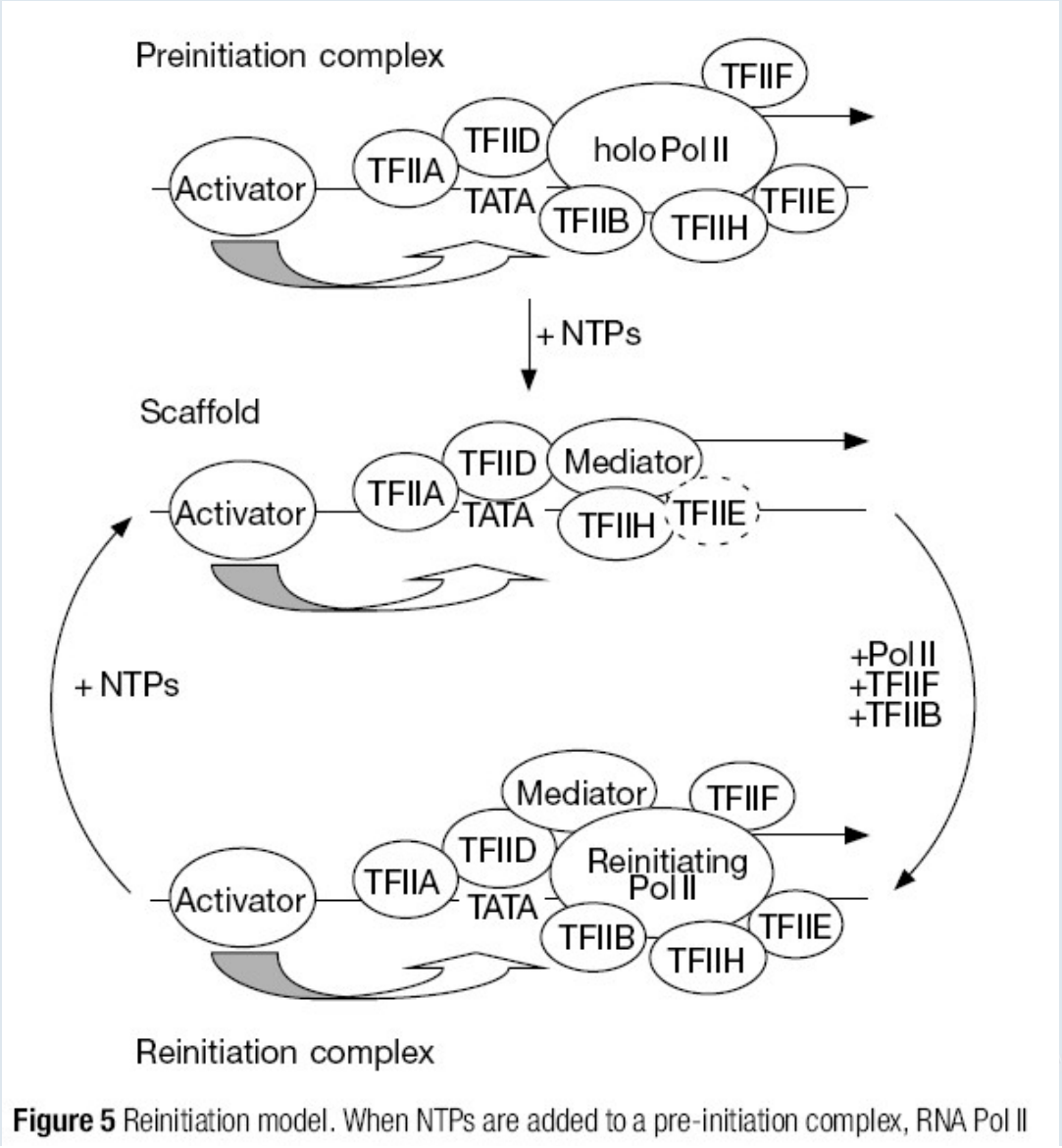


Figure 5 Reinitiation model. When NTPs are added to a pre-initiation complex, RNA Pol II

Activators maintain a scaffold of proteins at the promoter that allows further polymerases to bind and initiate transcription.

This most likely involves a DNA loop between enhancer and promoter.