

# VU Medical Zone

ADMIN:

AMAAAN KHAN

---

BIO 302 FINAL TERM

PPT

LECTURE 102 TO 202

# Transcription Cycle in Bacteria

- The bacterial core RNA polymerase can, in principle, initiate transcription at any point on a DNA molecule.
- It, however, initiates transcription only at promoters.

# Transcription Cycle in Bacteria

- It is the addition of an initiation factor called  $\sigma$  that converts core enzyme ( $\alpha_2 \beta \beta' \omega$ ) into the form that initiates only at promoters. This form of the enzyme is called the RNA polymerase **holoenzyme**.

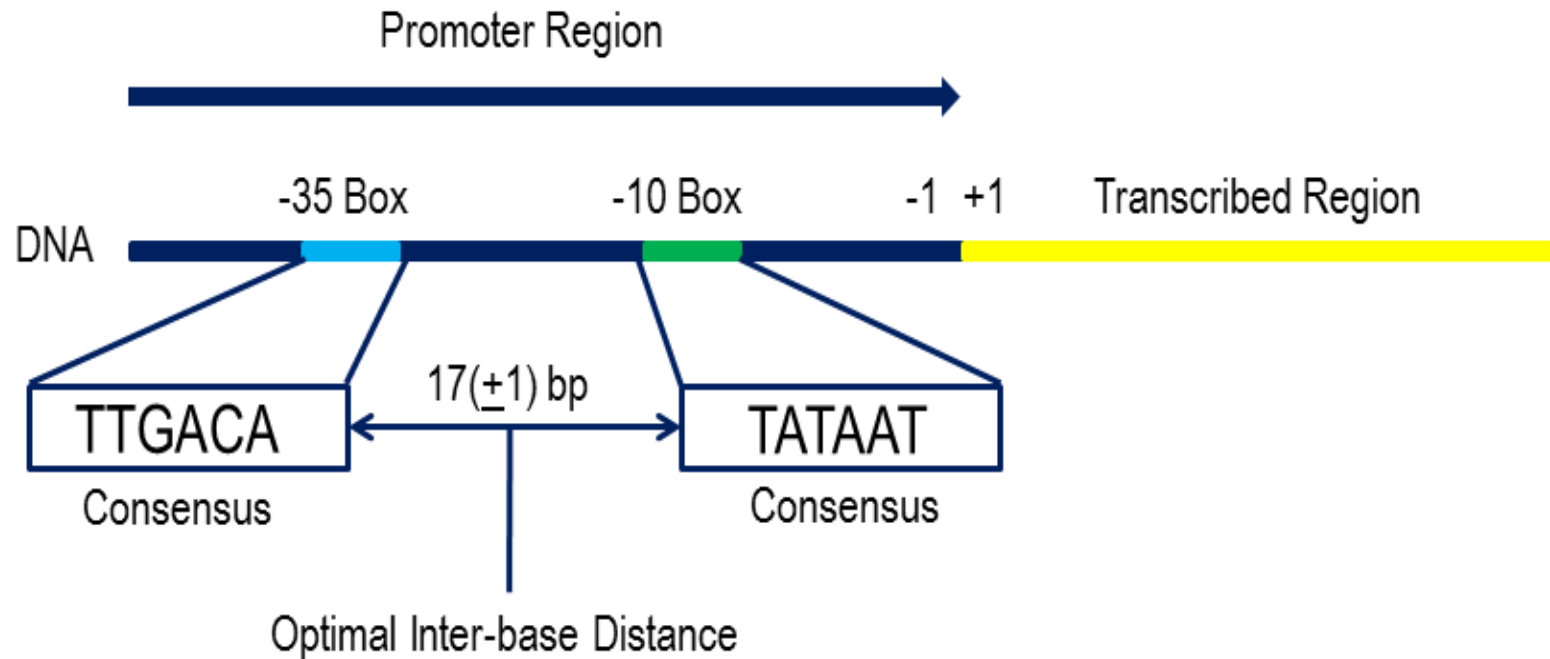
# Transcription Cycle in Bacteria

- In the case of *Escherichia coli*, the predominant  $\sigma$  factor is called  $\sigma^{70}$ .
- Promoters recognized by polymerase have two conserved sequences each of 6 nucleotides, separated by a nonspecific stretch of 17–19 nucleotides.

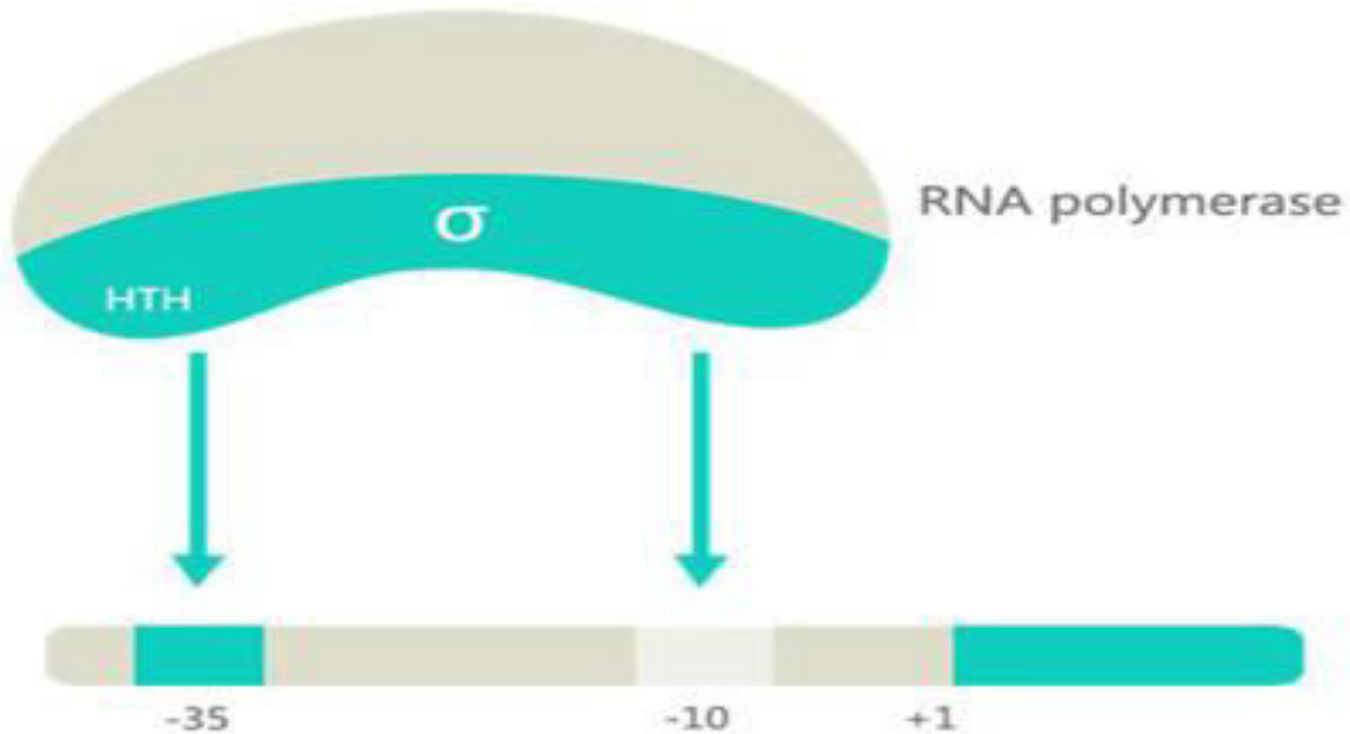
# Transcription Cycle in Bacteria

- The two defined sequences are centered, respectively, at ~10 bp and at ~35 bp upstream of the site where RNA synthesis starts.
- The sequences are thus called the **minus 35** and **minus 10** regions or element.

# Transcription Cycle in Bacteria



# Transcription Cycle in Bacteria





# Transcription Cycle in Bacteria

- The  $\sigma^{70}$  factor can be divided into four regions called  $\sigma$  region 1 through  $\sigma$  region 4.
- The regions that recognize the  $-10$  and  $-35$  elements of the promoter are regions 2 and 4, respectively.



# Transcription Cycle in Bacteria

- Two helices within region 4 form a common DNA-binding motif called a **helix-turn-helix**.
- One of these helices inserts into the major groove and interacts with bases in the  $-35$  region.

# Transcription Cycle in Bacteria

- The other helix lies across the top of the groove, making contacts with the DNA backbone.

# Transcription Cycle in Bacteria

## Helix-turn-helix



# Transcription Cycle in Bacteria

- The  $-10$  region has a more elaborate role in transcription initiation, because it is within that element that DNA melting is initiated in the transition from the closed to the open complex.

# Transcription Cycle in Bacteria

- Thus, the region of  $\sigma$  that interacts with the  $-10$  region is doing more than simply binding DNA.

# Transcription Cycle in Bacteria

- The  $\sigma$  helix involved in recognition of the  $-10$  region contains several essential aromatic amino acids that can interact with bases on the non-template strand in a manner that stabilizes the melted DNA.



# Transcription Cycle in Bacteria

- Two bases in the non-template strand are flipped out and inserted into pockets within the  $\sigma$  protein where they make favorable contacts that stabilize the unwound state of the promoter region.



# Transcription Cycle in Bacteria

- The  $\sigma$  subunit is positioned within the holoenzyme structure in such a way as to make feasible the recognition of various promoter elements.

# Transcription Cycle in Bacteria

- Thus, the DNA-binding regions point away from the body of the enzyme, rather than being embedded.

END

# Transition to the Open Complex

- Transition to the Open Complex involves structural changes in RNA Polymerase and in the Promoter DNA.
- The initial binding of RNA polymerase to the promoter DNA in the closed complex leaves the DNA in double-stranded form.

# Transition to the Open Complex

- The next stage in initiation requires the enzyme to become more intimately engaged with the promoter, in the open complex.

# Transition to the Open Complex

- The transition from the closed to the open complex involves structural changes in the enzyme and the opening of the DNA double helix to reveal the template and non-template strands.

# Transition to the Open Complex

- This “melting” occurs between positions –11 and +2, with respect to the transcription start site.
- In case of the bacterial enzyme bearing  $\sigma 70$ , this transition is often called **isomerization**.

# Transition to the Open Complex

- This transition does not require energy derived from ATP hydrolysis and is instead the result of a spontaneous conformational change in the DNA–enzyme complex to a more energetically favourable form.



# Transition to the Open Complex

- Two bases in the non-template strand of the –10 element flip out from their base-stacking interactions and instead insert into pockets within the  $\sigma$  protein where they make more favourable interactions.

# Transition to the Open Complex

- By stabilizing the single-stranded form of the  $-10$  element, these interactions drive melting of the promoter region.

# Transition to the Open Complex

- Isomerization is essentially irreversible and, once complete, typically guarantees that transcription will subsequently initiate.
- Formation of the closed complex, in contrast, is readily reversible.

# Transition to the Open Complex

- The examination of the structure of the holoenzyme in more detail reveals the active site of the enzyme, which is made up of regions from both the  $\beta$  and  $\beta'$  subunits, is found at the base of the pincers within the active center cleft.

END

# Transition to the Open Complex

- There are five channels into the enzyme, as shown in the illustration of the open complex.
- The NTP-uptake channel allows ribonucleotides to enter the active center.

# Transition to the Open Complex

- The RNA-exit channel allows the growing RNA chain to leave the enzyme as it is synthesized during elongation.
- The remaining three channels allow DNA entry and exit from the enzyme.



# Transition to the Open Complex

- The downstream DNA (i.e., DNA ahead of the enzyme, yet to be transcribed) enters the active center cleft in double-stranded form through the downstream DNA channel (between the pincers).



# Transition to the Open Complex

- Within the active center cleft, the DNA strands separate from position +3.

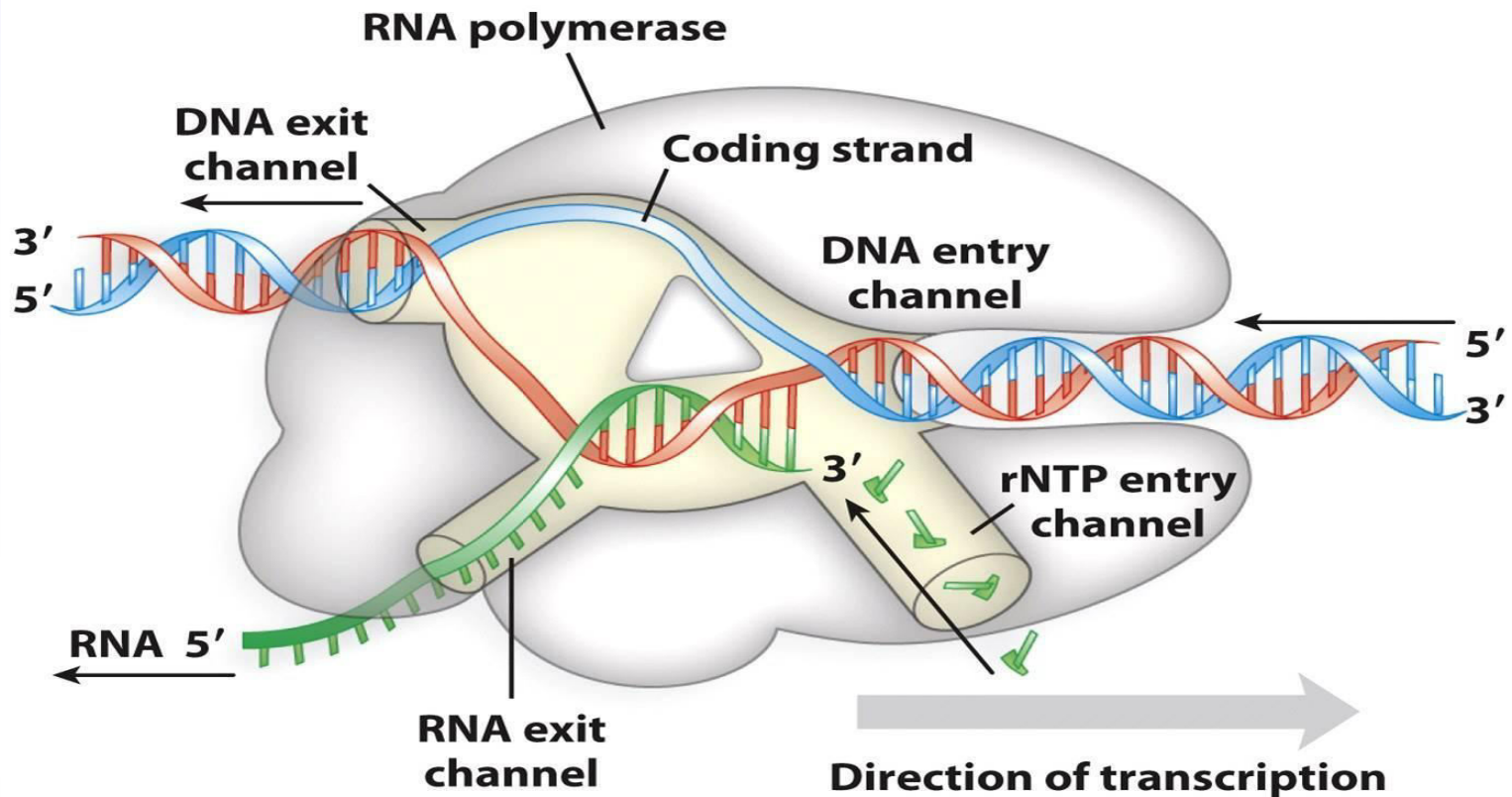
# Transition to the Open Complex

- The non-template strand exits the active center cleft through the non-template-strand (NT) channel and travels across the surface of the enzyme.

# Transition to the Open Complex

- The template strand, in contrast, follows a path through the active center cleft and exits through the template-strand (T) channel.
- The double helix reforms at  $-11$  in the upstream DNA behind the enzyme.

# Transition to the Open Complex



**Figure 15-14**  
*Molecular Biology: Principles and Practice*  
© 2012 W. H. Freeman and Company

# Transition to the Open Complex

- Two striking structural changes are seen in the enzyme upon isomerization from the closed to the open complex.

# Transition to the Open Complex

- First, the pincers at the front of the enzyme clamp down tightly on the downstream DNA.
- Second, there is a major shift in the position of the amino-terminal region of  $\sigma$ .



# Transition to the Open Complex

- When not bound to DNA,  $\sigma$  region 1.1 lies within the active center cleft of the holoenzyme, blocking the path that, in the open complex, is followed by the template DNA strand.



# Transition to the Open Complex

- In the open complex, region 1.1 shifts some 50Å° and is now found on the outside of the enzyme, allowing the DNA access to the cleft.

# Transition to the Open Complex

- Region 1.1 of  $\sigma$  is highly negatively charged ( just like DNA). Thus, in the holoenzyme, region 1.1 acts as a **molecular mimic** of DNA.

# Transition to the Open Complex

- The space in the active center cleft, which may be occupied either by region 1.1 or by DNA, is highly positively charged.

END

# Initial Transcription

- RNA polymerase can initiate a new RNA chain on a DNA template and thus does not need a primer.

# Initial Transcription

- This requires that the DNA template be brought into the polymerase active site and held stably in a helical conformation.

# Initial Transcription

- And that the initiating ribonucleotide be brought into the active site and held stably on the template while the next NTP is presented with correct geometry for the polymerization to take place.



# Initial Transcription

- This is particularly difficult because RNA polymerase starts most transcripts with an A, and that ribonucleotide binds the template nucleotide (T) with only two hydrogen bonds.

# Initial Transcription

- Thus, the enzyme has to make specific interactions with the DNA template strand, the initiating ribonucleotide, and the second ribonucleotide to hold them all rigidly in the correct orientation to allow chemical attack by the incoming NTP.

# Initial Transcription

- This is the reason that most transcripts start with the same nucleotide.

# Initial Transcription

- During initial transcription, RNA polymerase produces and releases short RNA transcripts of  $<10$  nucleotides (abortive synthesis) before escaping the promoter, entering the elongation phase, and synthesizing the proper transcript.

# Initial Transcription

- It has long been unclear how the enzyme's active site translocates along the DNA template during initial abortive cycles of transcription.
- Three general models were proposed to explain this.

# Initial Transcription

- 1. The “transient excursion” model proposes transient cycles of forward and reverse translocation of RNA polymerase.

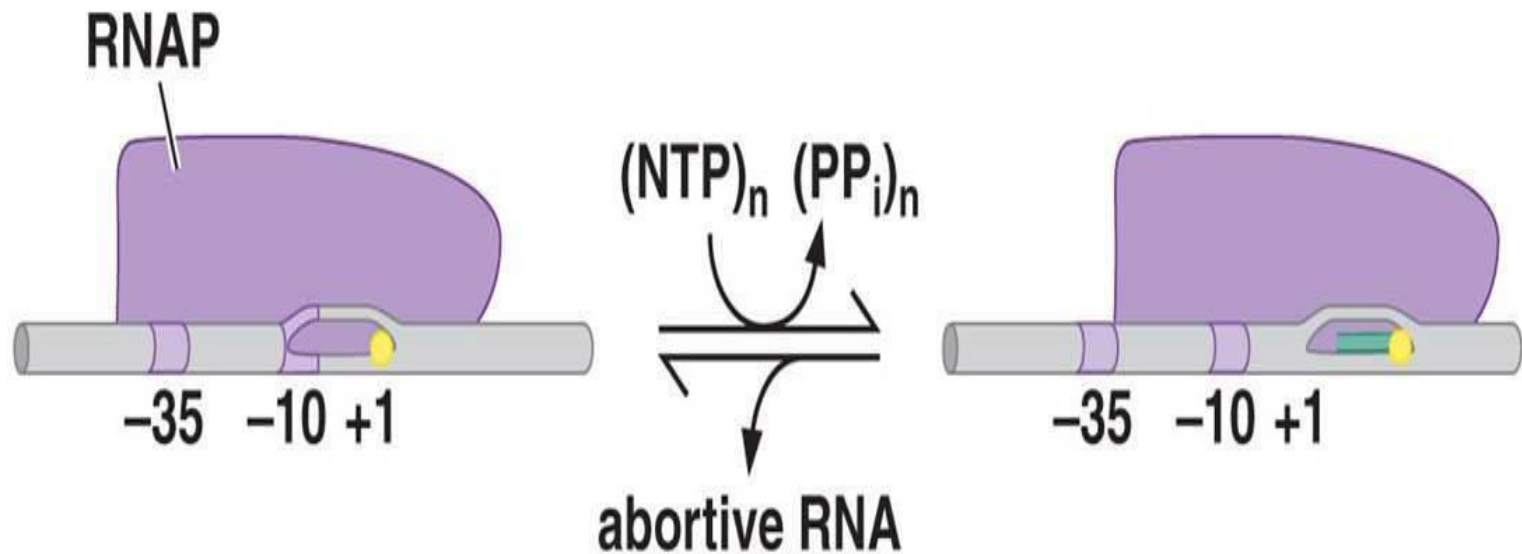


# Initial Transcription

- Thus, polymerase is thought to leave the promoter and translocate a short way along the DNA template, synthesizing a short transcript before aborting transcription, releasing the transcript, and returning to its original location on promoter.

# Initial Transcription

“transient excursions”



# Initial Transcription

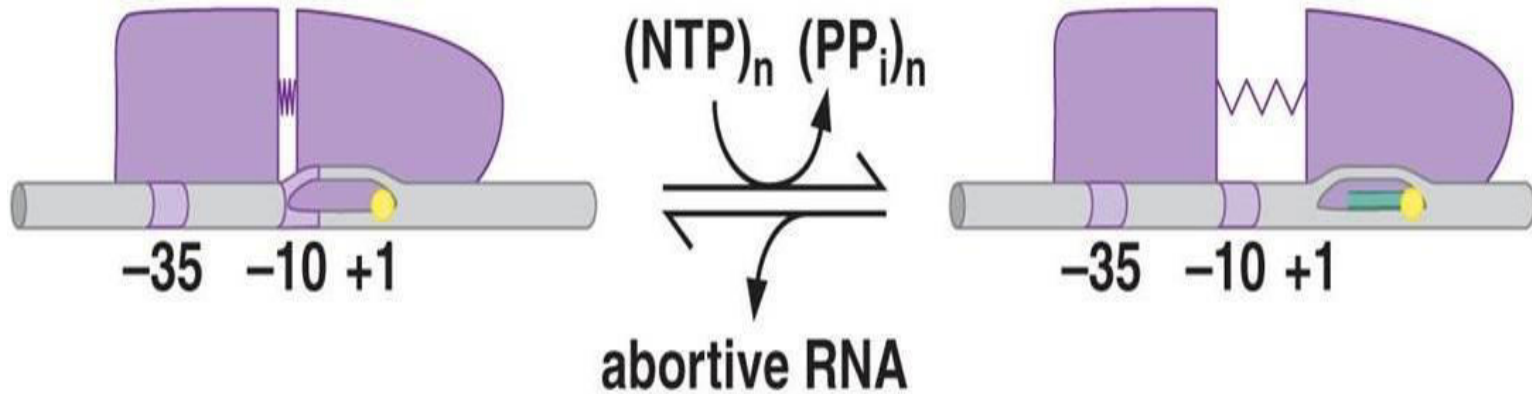
- 2. “Inchworming” invokes a flexible element within the polymerase that allows a module at the front of the enzyme, containing the active site.

# Initial Transcription

- The module moves downstream, synthesizing a short transcript before aborting and retracting to the body of the enzyme still at the promoter.

# Initial Transcription

“inchworming”



# Initial Transcription

- 3. “Scrunching”  
proposes that DNA downstream from the stationary, promoter-bound, polymerase is unwound and pulled into the enzyme.



# Initial Transcription

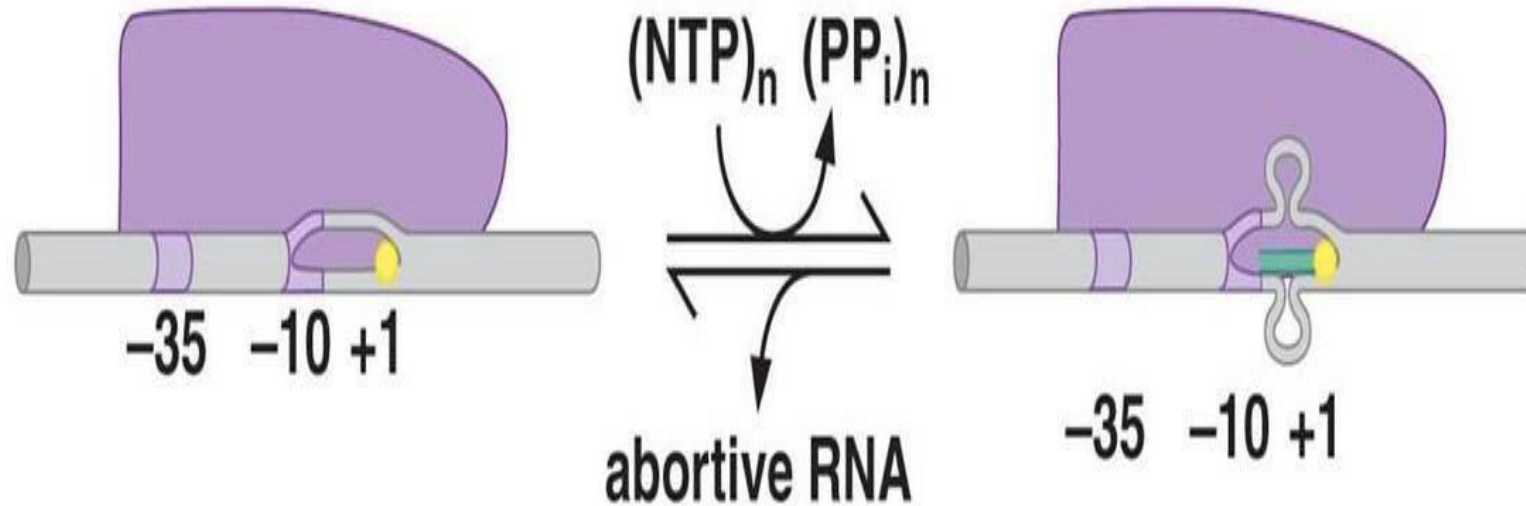
- The DNA thus accumulated within the enzyme is accommodated as single-stranded bulges.

# Initial Transcription

- The DNA thus accumulated within the enzyme is accommodated as single-stranded bulges.

# Initial Transcription

“scrunching”



Alberts et al. Molecular Biology of the Cell, 6th Edition, © Garland Science 2015

# Initial Transcription

- It is now believed that the third model—scrunching—reflects what actually happens.

# Initial Transcription

- The experiments have shown that during initial transcription, the polymerase remains stationary on the promoter, unwinds downstream DNA, and pulls that DNA into itself.

# Initial Transcription

- Only the scrunching model is consistent with these results.

END



# The Elongating Polymerase

- The Elongating Polymerase Is a processive machine that synthesizes and proofreads RNA.
- DNA passes through the elongating enzyme in a manner very similar to its passage through the open complex.

# The Elongating Polymerase

- Thus, double-stranded DNA enters the front of the enzyme between the pincers.
- At the opening of the catalytic cleft, the strands separate to follow different paths through the enzyme.

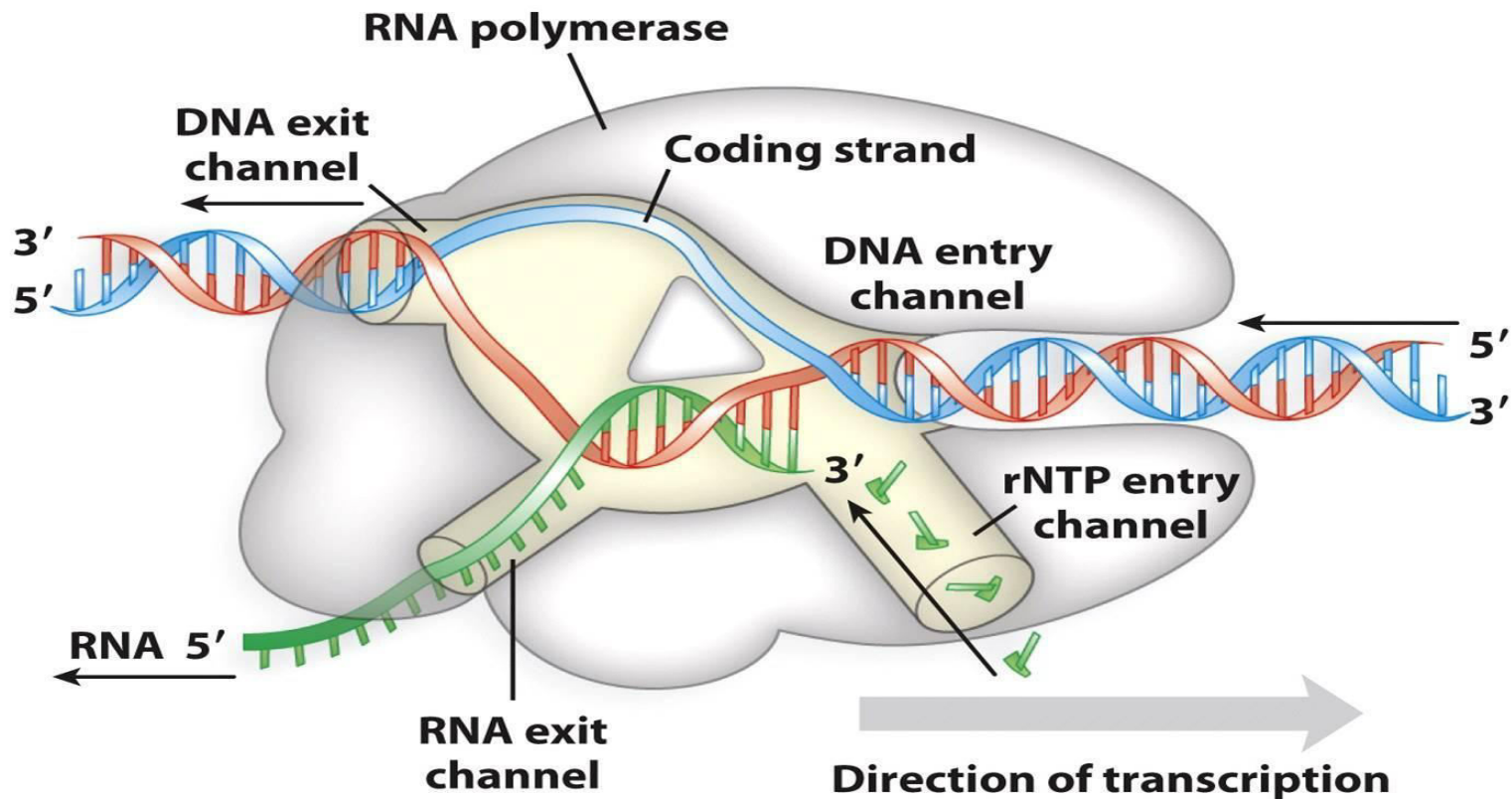
# The Elongating Polymerase

- The strands exit via their respective channels and re-form a double helix behind the elongating polymerase.

# The Elongating Polymerase

- Ribonucleotides enter the active site through their defined channel and are added to the growing RNA chain under the guidance of the template DNA strand.

# The Elongating Polymerase



**Figure 15-14**  
*Molecular Biology: Principles and Practice*  
© 2012 W. H. Freeman and Company

# The Elongating Polymerase

- Only 8 or 9 nucleotides of the growing RNA chain remain base-paired to the DNA template at any given time.
- The remainder of the RNA chain is peeled off and directed out of the enzyme through the RNA exit channel.



# The Elongating Polymerase

- During elongation, the enzyme adds one nucleotide at a time to the growing RNA transcript.
- During elongation, polymerase uses a step mechanism: using single-molecule techniques.



# The Elongating Polymerase

- It was shown that the enzyme steps forward as a molecular motor, advancing in a single step a distance equivalent to a base pair for every nucleotide it adds to the growing RNA chain.

# The Elongating Polymerase

- In addition, the size of the bubble, i.e., the length of DNA that is not double-helical, remains constant throughout elongation.
- 1 bp is separated ahead of the processing enzyme, 1 bp is formed behind it.

# The Elongating Polymerase

- As well as synthesizing the transcript, RNA polymerase performs two proofreading functions on that growing transcript.
- The first of these is called pyrophosphorolytic editing.

# The Elongating Polymerase

- In this, the enzyme uses its active site, in a simple back-reaction, to catalyze the removal of an incorrectly inserted ribonucleotide, by reincorporation of P<sub>Pi</sub>.

# The Elongating Polymerase

- The enzyme can then incorporate another ribonucleotide in its place in the growing RNA chain.

# The Elongating Polymerase

- In the second proofreading mechanism, called **hydrolytic editing**, the polymerase backtracks by one or more nucleotides and cleaves the RNA product, removing the error-containing sequence.

# The Elongating Polymerase

- Another group of proteins—the **Nus proteins**—joins polymerase in the elongation phase and promotes the processes of elongation and termination.

END



# Termination of Transcription

- The termination of transcription is a normal and important function at the ends of gene/s.
- However, an unusual termination may be triggered by damaged DNA or by other unanticipated hindrances.

# Termination of Transcription

- The sequences called **terminators** trigger the elongating polymerase to dissociate from the DNA and release the RNA chain it has made.

# Termination of Transcription

- In bacteria, terminators come in two types:-
  - 1. Rho-Dependant
  - 2. Rho-Independent

# Termination of Transcription

- The first, requires a protein called Rho to induce termination.
- The second causes termination without the involvement of other factors.

# Termination of Transcription

- Rho-dependent terminators have rather ill-defined RNA elements called **rut sites**.
- The work of these sites requires the action of the Rho factor.

# Termination of Transcription

- Rho, which is a ring-shaped protein with six identical subunits, binds to single-stranded RNA as it exits the polymerase.

# Termination of Transcription

- The protein also has an ATPase activity, and once attached to the transcript, Rho uses the energy derived from ATP hydrolysis to induce termination.



# Termination of Transcription

- The precise mechanism of termination remains to be determined, and models include the following:-

# Termination of Transcription

- i. Rho pushes polymerase forward relative to the DNA and RNA, resulting in termination in a manner analogous to termination by the protein TRCF (transcription-repair coupling factor).

# Termination of Transcription

- **Rho** binds double-stranded DNA upstream of the polymerase and translocates along the DNA until it collides the RNA polymerase.

# Termination of Transcription

- The collision pushes polymerase forward, causing dissociation of the ternary complex of RNA polymerase, template DNA, and RNA transcript.

END

# Termination of Transcription

- ii. Rho pulls RNA out of the polymerase, resulting in termination; or
- iii. Rho induces a conformational change in polymerase, causing the enzyme to terminate the transcription.

# Termination of Transcription

- Most recent experiments suggest that the last of the above may be correct and that the conformational change causes the elongating complex to stall, with dissociation following slowly.

# Termination of Transcription

- Recent studies also have suggested that Rho binds to RNA polymerase throughout the transcription cycle.
- Thus, Rho doesn't reach polymerase by translocating along a nascent, rut-containing transcript like TRCF.



# Termination of Transcription

- Rather, Rho binds polymerase early in transcription and then at some point also binds the RNA transcript being exuded from that elongating enzyme.

# Termination of Transcription

- The role of translocation by Rho is thus perhaps to tighten the resulting RNA loop, and when sufficiently tight, polymerase elongation ceases.
- How is Rho directed to work on particular RNA transcripts?

# Termination of Transcription

- First, there is some specificity in the sites it binds (the rut sites).
- These sites consist of stretches of ~40 nucleotides that do not fold into a secondary structure. They are also rich in C residues.

# Termination of Transcription

- The second level of specificity is that Rho fails to bind any transcript that is being translated.

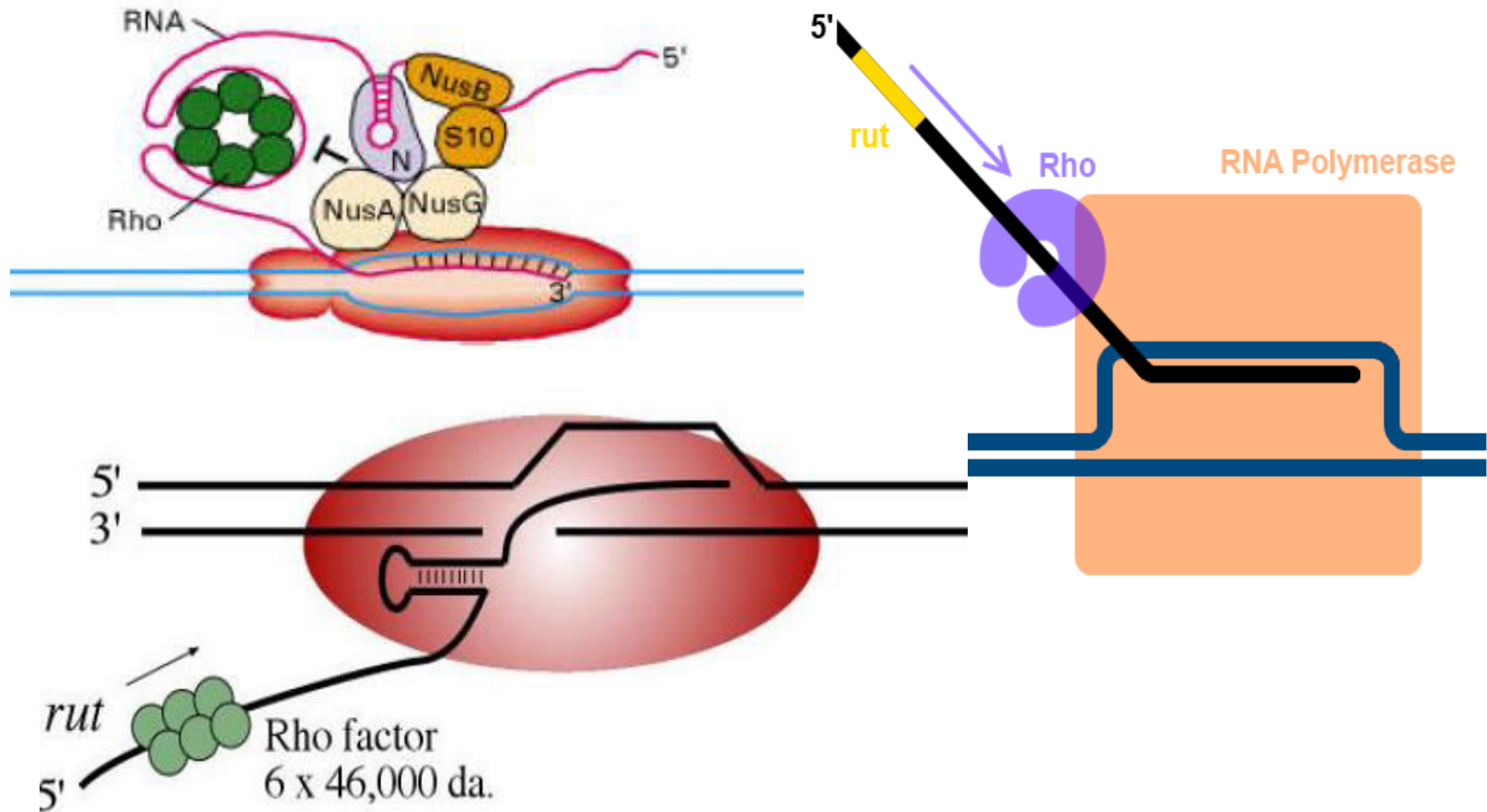
# Termination of Transcription

- In bacteria, transcription and translation are tightly coupled—translation initiates on growing RNA transcripts as soon as they start exiting polymerase, while they are still being synthesized.

# Termination of Transcription

- Thus, Rho typically terminates only those transcripts which are still being transcribed beyond the end of a gene or operon.

# Termination of Transcription





# Rho-independent Terminators

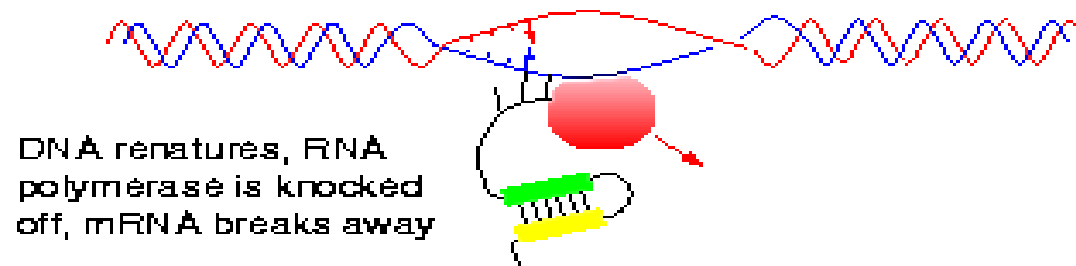
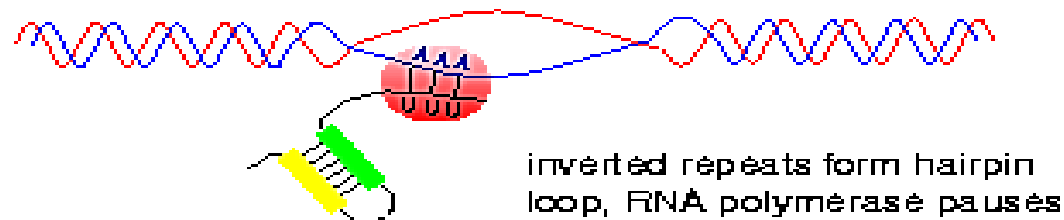
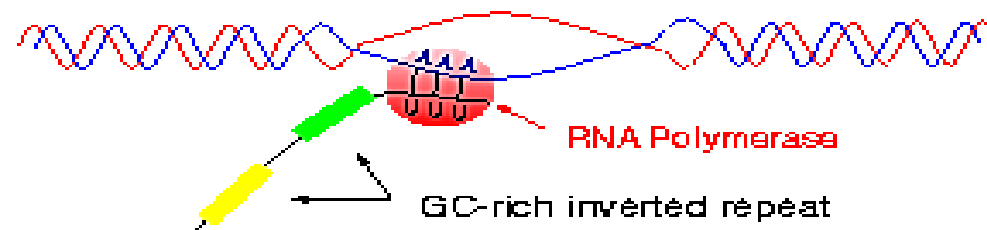
- Rho-independent terminators are also called **intrinsic terminators** because they need no other factors to work.

# Rho-independent Terminators

- They consist of two sequence elements:-
- A short inverted repeat of 20 nucleotides followed by a stretch of about eight A:T base pairs.

# Rho-independent Terminators

## Chain Termination



# Rho-independent Terminators

- These elements do not affect the polymerase until they have been transcribed—that is, they function in the RNA rather than in the DNA.

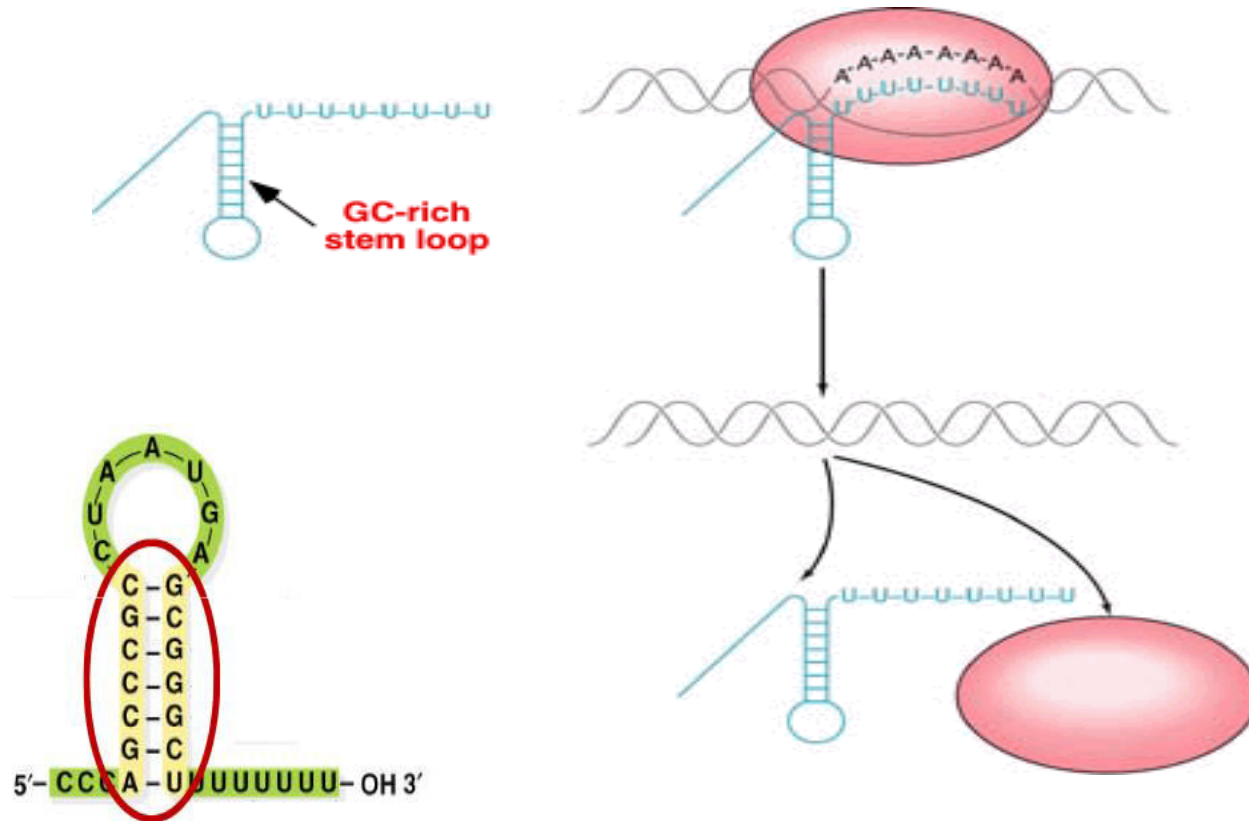
# Rho-independent Terminators

- When polymerase transcribes an inverted repeat sequence, the resulting RNA can form a stem-loop structure often called a “hairpin” by base-pairing with itself.

# Rho-independent Terminators

- Formation of the hairpin causes termination by disrupting the elongation complex.

# Rho-independent Terminators





# Rho-independent Terminators

- The hairpin induces termination by either pushing polymerase forward relative to the DNA and RNA, wresting the transcript from polymerase, or inducing a conformational change in polymerase.

# Rho-independent Terminators

- The hairpin works as an efficient terminator only when it is followed by a stretch of A:U base pairs.

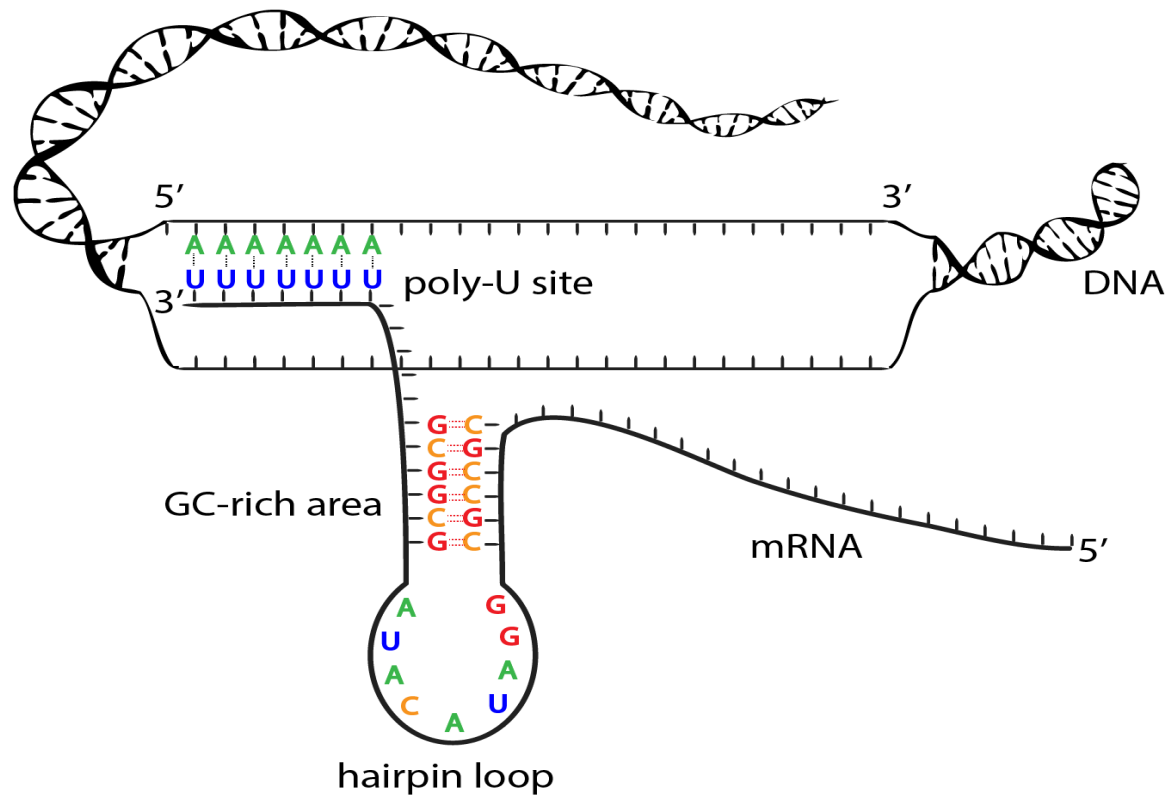
# Rho-independent Terminators

- This is because, under those circumstances, at the time the hairpin forms, the growing RNA chain will be held on the template at the active site by only A:U base pairs.

# Rho-independent Terminators

- Because A:U base pairs are the weakest of all base pairs, they are more easily disrupted by the effects of the stem-loop on the transcribing polymerase, and thus the RNA will more readily dissociate.

# Rho-independent Terminators



# TRANSCRIPTION IN EUKARYOTES

- Transcription in eukaryotes is undertaken by polymerases closely related to the RNA polymerases found in prokaryotes.

# TRANSCRIPTION IN EUKARYOTES

- The process of transcription is identical in both the prokaryotes and eukaryotes. There are, however, differences in the machinery used in each case.



# TRANSCRIPTION IN EUKARYOTES

- Bacteria have only one RNA polymerase but all eukaryotes have at least three different ones i.e., Pol I, II, and III; and plants also have a Pol IV and a Pol V.

# TRANSCRIPTION IN EUKARYOTES

- In addition, whereas bacteria require only one additional initiation factor ( $\sigma$ ), several initiation factors are required for efficient and promoter-specific initiation in eukaryotes.

# TRANSCRIPTION IN EUKARYOTES

- These are called the general transcription factors (GTFs).
- In vitro, the general transcription factors are all that are required, together with Pol II, to initiate transcription on a DNA template.

# TRANSCRIPTION IN EUKARYOTES

- In vivo, however, the general transcription factors are not alone sufficient to bind promoter sequences and elicit significant expression.

# TRANSCRIPTION IN EUKARYOTES

- Rather, additional factors are required, including DNA-binding regulatory proteins, the so-called Mediator complex, and often chromatin-modifying enzymes.

# TRANSCRIPTION IN EUKARYOTES

- The eukaryotic **core promoter** refers to the minimal set of sequence elements required for accurate transcription initiation by the Pol II machinery.

# TRANSCRIPTION IN EUKARYOTES

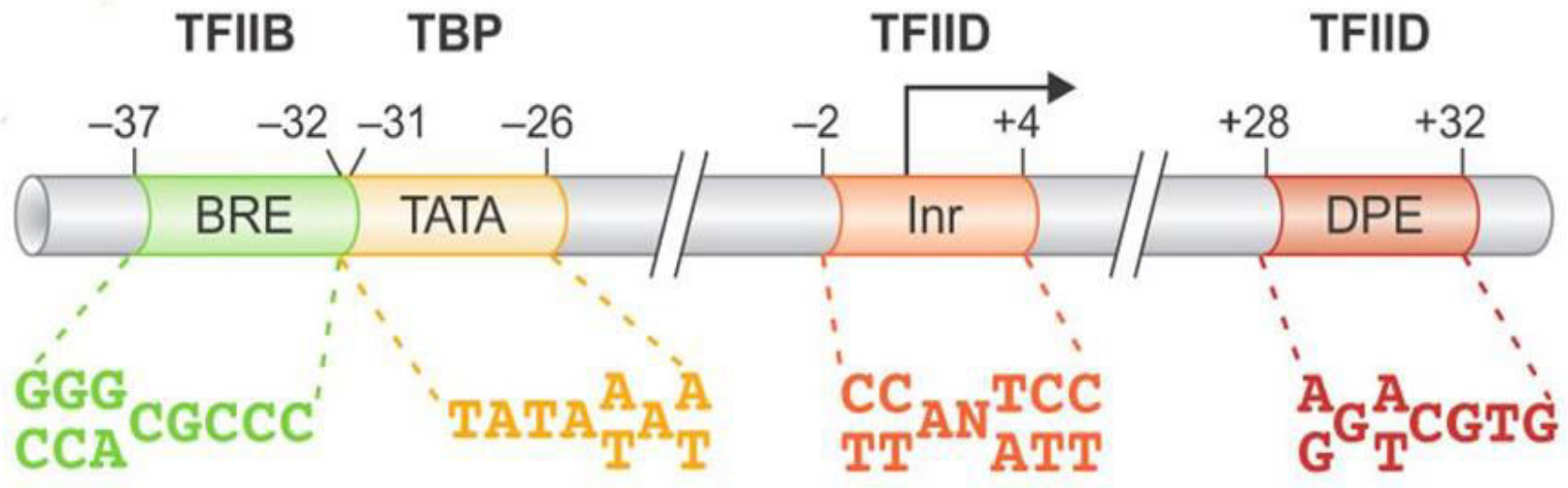
- A core promoter is typically ~40–60 nucleotides long, extending either upstream or downstream from the transcription start site.



# TRANSCRIPTION IN EUKARYOTES

- The elements found in Pol II core promoter include the TFIIB recognition element (BRE), the TATA element (or box), the initiator (Inr), and the downstream promoter elements (known as DPE, DCE, and MTE).

# TRANSCRIPTION IN EUKARYOTES



**Pol II core promoter**

# TRANSCRIPTION IN EUKARYOTES

- Typically, a promoter includes some subset of these elements. Thus, for example, promoters typically have either a TATA element or a DPE element, not both.

# TRANSCRIPTION IN EUKARYOTES

- Often, a TATA-containing promoter also contains a DCE.
- The Inr is the most common element, found in combination with both TATA and DPEs.

# TRANSCRIPTION IN EUKARYOTES

- Typically upstream of the core promoter, there are other sequence elements required for accurate and efficient transcription in vivo.
- Together, these elements constitute the **regulatory sequences**.

# TRANSCRIPTION IN EUKARYOTES

- These elements include promoter proximal elements, upstream activator sequences (UASs), enhancers, and a series of other elements called silencers, boundary elements, and insulators.

# TRANSCRIPTION IN EUKARYOTES

- All of these DNA elements bind regulatory proteins (activators and repressors), which help or hinder transcription from the core promoter.



# TRANSCRIPTION IN EUKARYOTES

- Some of these regulatory sequences can be located many tens or even hundreds of kilobases from the core promoters on which they act.

END

# Formation of pre-initiation Complex

- The general transcription factors collectively perform the functions performed by  $\sigma$  in bacterial transcription.

# Formation of pre-initiation Complex

- Thus, the general transcription factors help polymerase bind to the promoter and melt the DNA.
- This is comparable to the transition from the closed to the open complex in the bacterial case.

# Formation of pre-initiation Complex

- They also help polymerase escape from the promoter and embark on the elongation phase.

# Formation of pre-initiation Complex

- The complete set of general transcription factors and polymerase, bound together at the promoter and poised for initiation, is called the **pre-initiation complex**.

# Formation of pre-initiation Complex

- Many Pol II promoters contain a so-called TATA element (some 30 bp upstream of the transcription start site).
- This is where preinitiation complex formation begins.

# Formation of pre-initiation Complex

- The TATA element is recognized by the general transcription factor called **TFIID** (“TFII” denotes a transcription factor for Pol II).



# Formation of pre-initiation Complex

- Like many of the general transcription factors, TFIID is, in fact, a multi-subunit complex.
- The component of TFIID that binds to the TATA DNA sequence is called **TBP** (TATA-binding protein).

# Formation of pre-initiation Complex

- The other subunits in this complex are called **TAFs**, for TBP-associated factors.
- Some TAFs recognize other core promoter elements such as the Inr, DPE, and DCE, although the strongest binding is between TBP and TATA.

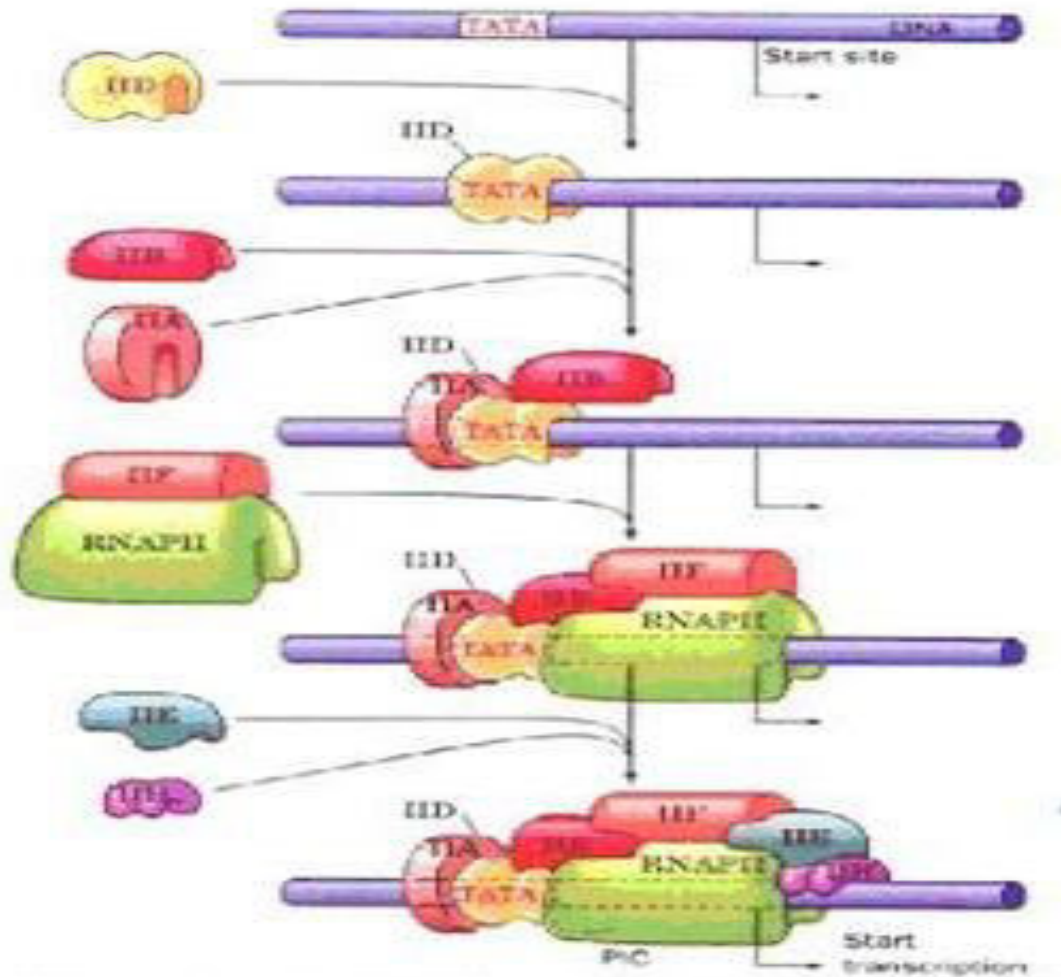
# Formation of pre-initiation Complex

- Thus, TFIID is a critical factor in promoter recognition and establishment of pre-initiation complex.

# Formation of pre-initiation Complex

- Thus, TFIID is a critical factor in promoter recognition and establishment of pre-initiation complex.

# Formation of pre-initiation Complex



# Formation of pre-initiation Complex

- Upon binding DNA, TBP extensively distorts the TATA sequence.
- The resulting TBP–DNA complex provides a platform to recruit other general transcription factors and polymerase itself to the promoter.



# Formation of pre-initiation Complex

- In vitro, these proteins assemble at the promoter in the following order: TFIIA, TFIIB, TFIIF together with polymerase, and then TFIIE and TFIIH.



# Formation of pre-initiation Complex

- Formation of the preinitiation complex containing these components is followed by promoter melting.
- In contrast to the situation in bacteria, promoter melting in eukaryotes requires hydrolysis of ATP and is mediated by TFIID.

END

# Promoter Escape

- The formation of pre-initiation complex follows a period of abortive initiation before the polymerase escapes the promoter and enters the elongation phase.

# Promoter Escape

- During abortive initiation, the polymerase synthesizes a series of short transcripts.
- In eukaryotes, promoter escape involves two steps not seen in bacteria:-

# Promoter Escape

- One is ATP hydrolysis (in addition to the earlier ATP hydrolysis needed for DNA melting), and the other is phosphorylation of the polymerase.

# Promoter Escape

- The large subunit of Pol II has a carboxy-terminal domain (CTD), which is referred to as the “tail”.
- The CTD contains a series of repeats of the heptapeptide sequence: Tyr-Ser-Pro-Thr-Ser-Pro-Ser.

# Promoter Escape

- There are 27 of these repeats in the yeast Pol II CTD, 32 in the worm *Caenorhabditis elegans*, 45 in the fly *Drosophila*, and 52 in humans.
- The number of repeats correlate with the complexity of the genome.

# Promoter Escape

- Each repeat contains sites for phosphorylation by specific kinases, including one that is a subunit of TFIIF.



# Promoter Escape

- The form of Pol II recruited to the promoter initially contains a largely unphosphorylated tail, but the species found in the elongation complex bears multiple phosphoryl groups on its tail.

# Promoter Escape

- Addition of these phosphates helps polymerase shed most of the general transcription factors used for initiation, and which the enzyme leaves behind as it escapes the promoter.

# Promoter Escape

- Regulating the phosphorylation state of the CTD of Pol II controls subsequent steps—elongation and even processing of the RNA— as well.

# Promoter Escape

- Indeed, in addition to TFIIH, several other kinases have been identified that act on the CTD, as well as a number of phosphatases that remove the phosphates added by those kinases.

END

# General Transcription Factors

- We do not know in detail the functions of all of the other general transcription factors.
- Some of these factors are in fact complexes made up of two or more subunits.

# General Transcription Factors

- **TAFs**
- TBP is associated with about 10 TAFs. Two of the TAFs bind DNA elements at the promoter, for example, the initiator element (Inr) and the downstream promoter elements.

# General Transcription Factors

- Several of the TAFs have structural homology with histone proteins, and it has been proposed that they might bind DNA in a similar manner, although evidence for such a binding has not been obtained.



# General Transcription Factors

- For example, TAF42 and TAF62 from *Drosophila* have been shown to form a structure similar to that of the H3.H4 tetramer of histones.

# General Transcription Factors

- Another TAF appears to regulate the binding of TBP to DNA. It does this using an inhibitory flap that binds to the DNA-binding surface of TBP.
- This flap must be displaced for TBP to bind TATA.

# General Transcription Factors

- **TFIIB**
- This protein, a single polypeptide chain, enters the preinitiation complex after TBP.
- The crystal structure of the ternary complex of TFIIB–TBP–DNA shows specific TFIIB–TBP and TFIIB–DNA contacts.

# General Transcription Factors

- These include base-specific interactions with the major groove upstream (to the BRE) and the minor groove downstream of the TATA element.

# General Transcription Factors

- The asymmetric binding of TFIIB to the TBP–TATA complex accounts for the asymmetry in the rest of the assembly of the preinitiation complex and the unidirectional transcription that results.

# General Transcription Factors

- TFIIB also contacts Pol II in the preinitiation complex.
- Thus, this protein appears to bridge the TATA-bound TBP and polymerase.

# General Transcription Factors

- Structural studies suggest that segments of TFIIB insert into the RNA-exit channel and active center cleft of Pol II in a manner analogous to the  $\sigma$  region  $3/4$  linker in the bacterial case.



# General Transcription Factors

- These regions of TFIIB(called the **linker** and **reader**) aid in open complex formation, perhaps by stabilizing the melted DNA until the RNA:DNA hybrid takes over that role.

# General Transcription Factors

- **TFIIF**
- This two-subunit factor associates with Pol II and is recruited to the promoter together with that enzyme (and other factors).

# General Transcription Factors

- Binding of Pol II–TFIIF stabilizes the DNA–TBP–TFIIB complex and is required before TFIIE and TFIIH are recruited to the preinitiation complex.

# General Transcription Factors

- In yeast, this factor includes a third subunit, but the function of the third subunit is not known.

END

# General Transcription Factors

- **TFIIE and TFIIH**
- TFIIE, which consists of two subunits, binds next and has roles in the recruitment and regulation of TFIIH.

# General Transcription Factors

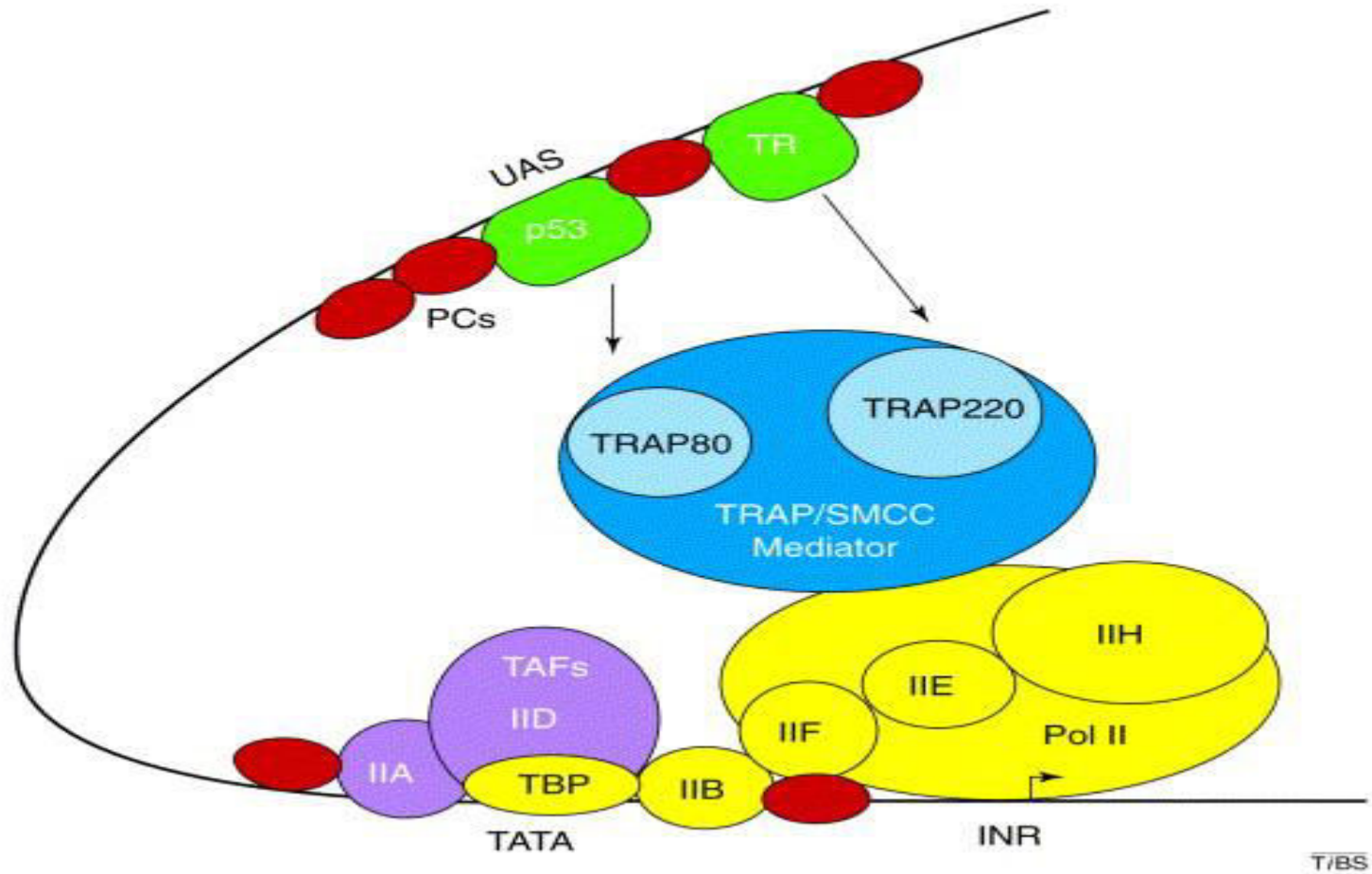
- TFIIH controls the ATP-dependent transition of the preinitiation complex to the open complex.

# General Transcription Factors

- It is the largest and most complex of the general transcription factors having 10 subunits and a molecular mass comparable to that of the polymerase itself.



# General Transcription Factors



# General Transcription Factors

- Within TFIIH are two subunits that function as ATPases and another that is a protein kinase, with roles in promoter melting and escape.

# General Transcription Factors

- Together with other factors, the ATPase subunits are also involved in nucleotide excision repair.
- How does TFIIH mediate promoter melting?

# General Transcription Factors

- It is now believed that a subunit of TFIIF acts as an ATP-driven translocator of double-stranded DNA.

# General Transcription Factors

- This subunit binds to DNA downstream from polymerase and feeds double-stranded DNA, with a right-handed threading, into the cleft of the polymerase.

# General Transcription Factors

- This action drives the melting of the DNA because the upstream promoter DNA is held in a fixed position by TFIID and the rest of the GTFs.

END

# The Mediator Complex

- High regulated levels of transcription in vivo, require transcriptional regulatory proteins, the **Mediator complex**, and nucleosome modifying enzymes in addition to the above described general transcription factors.



# The Mediator Complex

- One reason for these additional requirements is that the DNA template in vivo is packaged into chromatin.
- This condition complicates binding to the promoter of polymerase and its associated factors.

# The Mediator Complex

- Transcriptional regulatory proteins called **activators** help recruit polymerase to the promoter, stabilizing its binding there.

# The Mediator Complex

- This recruitment is mediated through interactions between DNA-bound activators, chromatin-modifying and -remodeling factors, and parts of the transcription machinery.

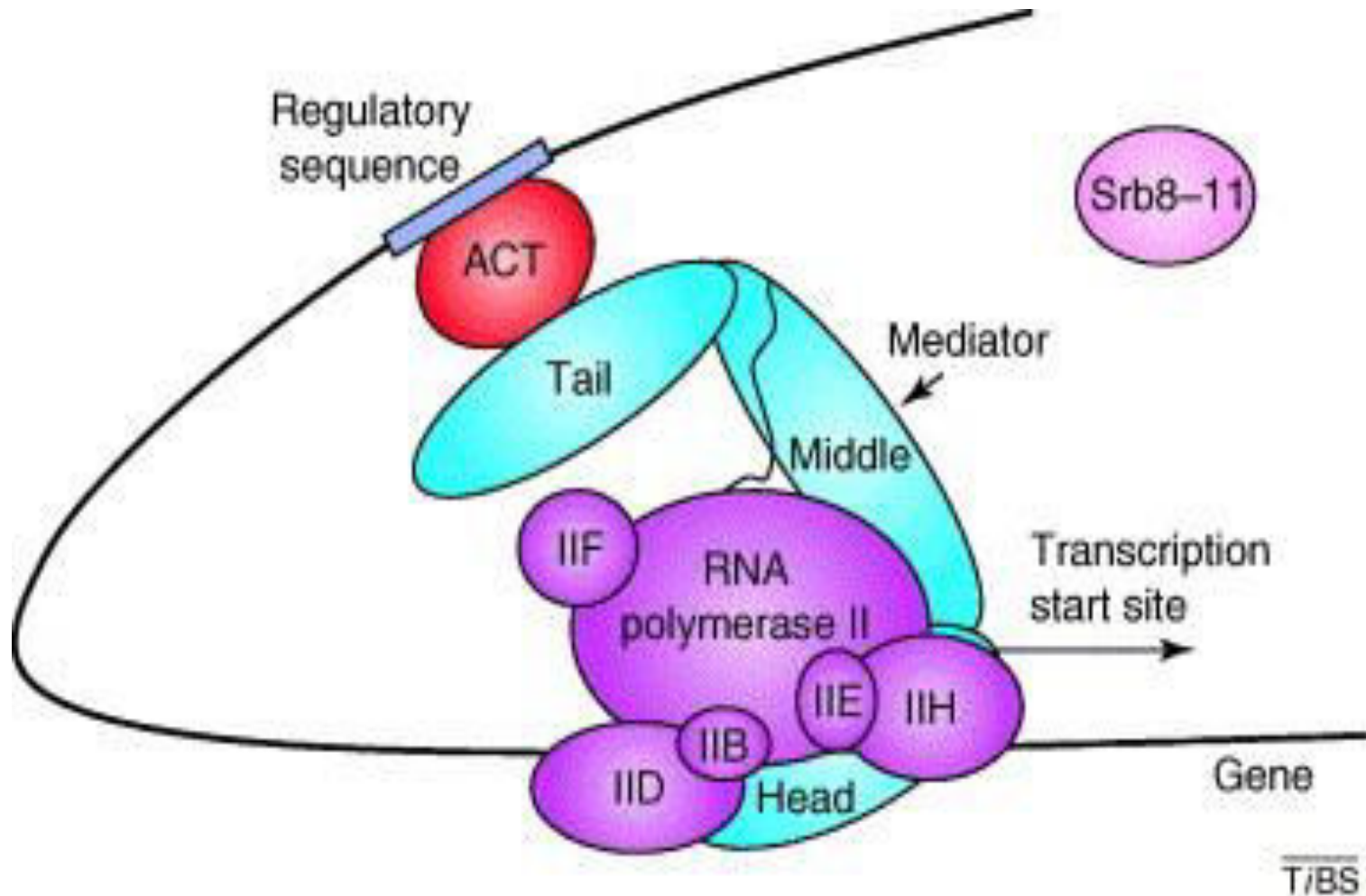
# The Mediator Complex

- One such interaction is with the Mediator complex.
- Mediator is associated with the basic transcription machinery, most likely touching the CTD “tail” of the large polymerase subunit through one surface,

# The Mediator Complex

- while presenting other surfaces for interaction with DNA-bound activators.
- This explains the need for Mediator to achieve significant transcription in vivo.

# The Mediator Complex





# The Mediator Complex

- Despite this central role in transcriptional activation, deletion of individual subunits of Mediator often leads to loss of expression of only a small subset of genes, different for each subunit.



# The Mediator Complex

- This result likely reflects the fact that different activators are believed to interact with different Mediator subunits to bring polymerase to different genes.
- Mediator also aids initiation by regulating the CTD kinase in TFIIF.

# The Mediator Complex

- The need for nucleosome modifiers and remodelers also differs at different promoters or even at the same promoter under different circumstances.

# The Mediator Complex

- When and where required, these complexes are also typically recruited by the DNA-bound activators, or sometimes by regulatory RNAs.

# The Mediator Complex

- The yeast and human Mediators each include more than 20 subunits, of which seven show significant sequence homology between the two organisms.

# The Mediator Complex

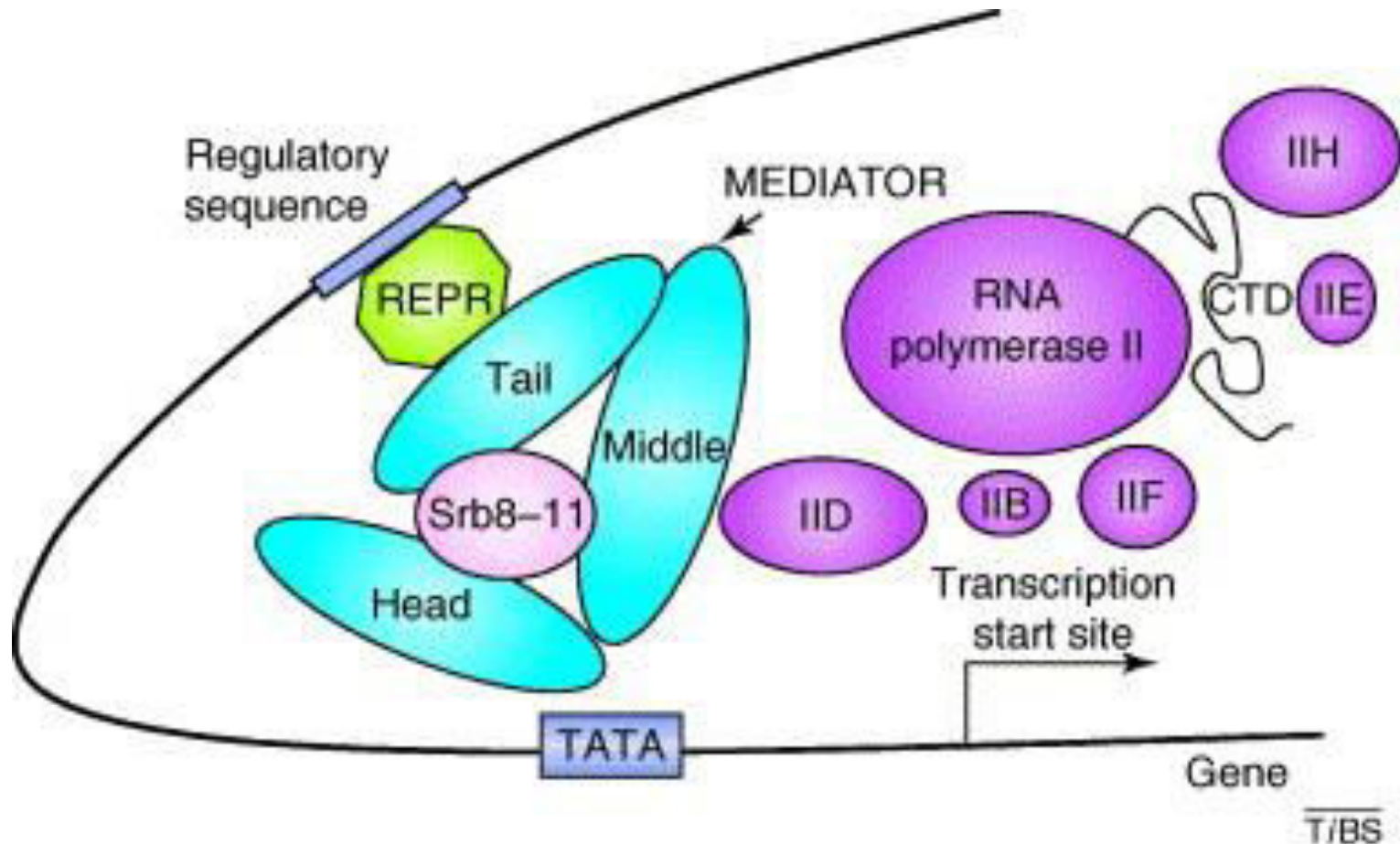
- The Mediator from both yeast and humans is organized in modules, each containing a subset of the subunits.

# The Mediator Complex

- These modules—called head, middle (or arm), and tail—can be dissociated from one another under certain conditions in vitro.



# The Mediator Complex





# The Mediator Complex

- Crystal structure of the head module of yeast Mediator reveals that it contains seven subunits (Med17/Srb4, Med11, Med22/Srb6, Med6, Med8, Med18/Srb5, and Med20/Srb2).

# The Mediator Complex

- It forms a three-domain structure that binds the transcription complex in such a way as to juxtapose TFIIF and the CTD tail of RNA polymerase, promoting phosphorylation of the latter by the former.

END

# RNA Elongation and Proofreading

- Once polymerase has escaped the promoter and initiated transcription, it shifts into the elongation phase.

# RNA Elongation and Proofreading

- This transition involves the Pol II enzyme shedding most of its initiation factors such as, the general transcription factors and Mediator.

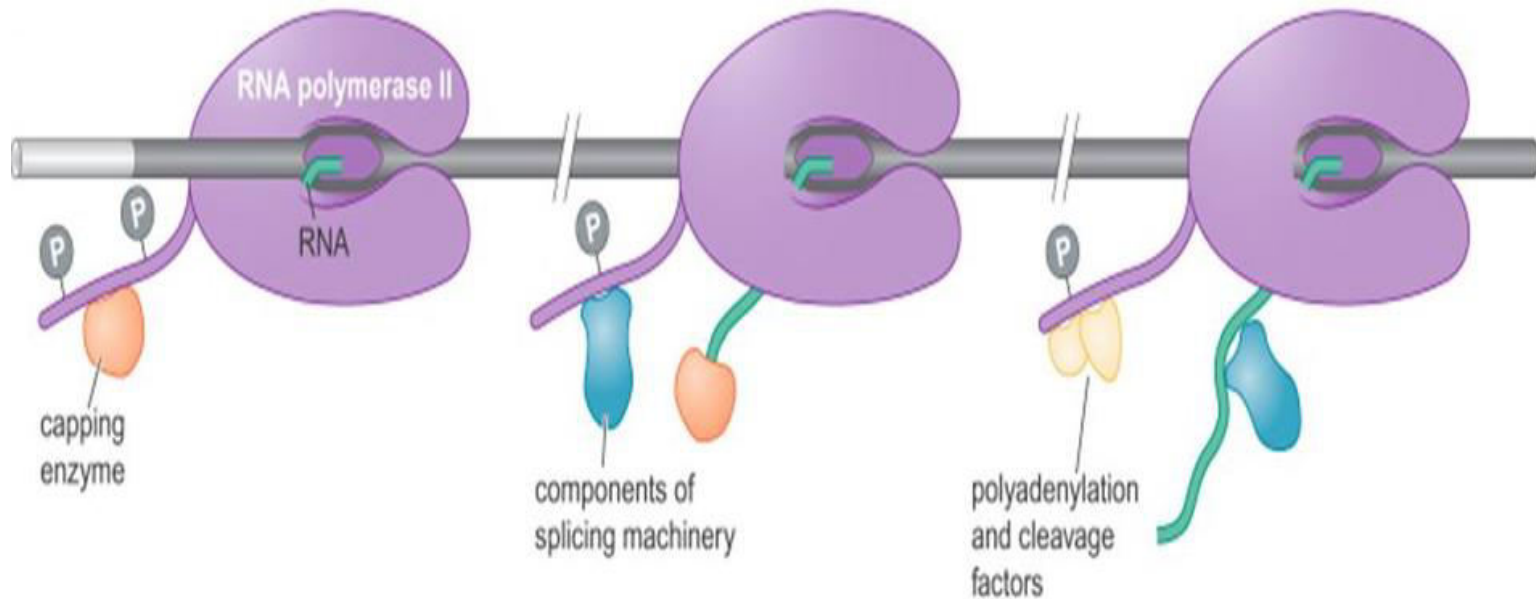
# RNA Elongation and Proofreading

- In their place, another set of factors is recruited. Some of these (such as TFIIS and SPT5) are **elongation factors** (i.e., factors that stimulate elongation).
- Others are required for RNA processing.

# RNA Elongation and Proofreading

- The enzymes involved in RNA processing are, like several of the initiation factors recruited to the carboxy-terminal (CTD) tail of the large subunit of Pol II

# RNA Elongation and Proofreading





# RNA Elongation and Proofreading

- In this case, however, the factors favor the phosphorylated form of the CTD. Thus, phosphorylation of the CTD leads to an exchange of initiation factors for those factors required for elongation and RNA processing.

# RNA Elongation and Proofreading

- As is evident from the crystal structure of yeast Pol II, the polymerase CTD lies directly adjacent to the channel through which the newly synthesized RNA exits the enzyme.

# RNA Elongation and Proofreading

- The CTD tail is also very long.
- It could potentially extend 800 Å from the body of the enzyme—that is, about seven times the length of the rest of the enzyme.

# RNA Elongation and Proofreading

- Together, these features allow the tail to bind several components of the elongation and processing machinery and deliver them to the emerging RNA.

# RNA Elongation and Proofreading

- Various proteins are thought to stimulate elongation by Pol II.
- One of these, the **kinase P-TEFb**, is recruited to polymerase by transcriptional activators.

# RNA Elongation and Proofreading

- Once bound to Pol II, this protein phosphorylates the serine residue at position 2 of the CTD repeats. That phosphorylation event correlates with elongation.

# RNA Elongation and Proofreading

- In addition, P-TEFb phosphorylates and thereby activates another protein, called **SPT5**, itself an elongation factor.
- Finally, **TAT-SF1**, yet another elongation factor, is recruited by P-TEFb.



# RNA Elongation and Proofreading

- Thus, P-TEFb stimulates elongation in three separate ways.
- SPT5 is comparable to the bacterial elongation factor NusG.

# RNA Elongation and Proofreading

- Indeed, this is the only universally conserved transcription factor across all three kingdoms of life—from bacteria, through Archaea, to eukaryotes.

END

# RNA Elongation and Proofreading

- NusG/SPT5 factors bind to their respective RNA polymerases at the tip of the clamp, overlapping the region contacted by  $\sigma$  region 4 (in bacteria) and TFIIB (in eukaryotes).

# RNA Elongation and Proofreading

- This overlapping and binding raises the interesting possibility that displacing initiation factors may be part of the function of these elongation regulators.

# RNA Elongation and Proofreading

- This also suggests that regulating the rate of elongation is an ancient mechanism of regulating gene expression.

# RNA Elongation and Proofreading

- There are some promoters in higher eukaryotes where the preinitiation complex is recruited effectively, but polymerase remains paused just after initiating transcription.

# RNA Elongation and Proofreading

- Such promoters seem to be associated with genes poised to be expressed either rapidly or in a highly coordinated fashion.



# RNA Elongation and Proofreading

- And their expression is regulated through recruitment by specific activators of the PTEFb kinase, which then releases them from their pause.

# RNA Elongation and Proofreading

- Yet another class of elongation factor is the so-called **ELL family**.
- These also bind to elongating polymerase and suppress transient pausing by the enzyme.

# RNA Elongation and Proofreading

- The first human ELL protein was originally identified as the product of a gene that undergoes translocations in acute myeloid leukemia.

# RNA Elongation and Proofreading

- Another factor that does not affect initiation, but stimulates elongation, is **TFIIS**.

# RNA Elongation and Proofreading

- This factor, like ELL, stimulates the overall rate of elongation by limiting the length of time that polymerase pauses when it encounters sequences that would otherwise tend to slow the enzyme's progress.

# RNA Elongation and Proofreading

- It is a feature of polymerase that it does not transcribe through all sequences at a constant rate.
- Rather, it pauses periodically, sometimes for rather long periods, before resuming transcription.

# RNA Elongation and Proofreading

- In the presence of TFIIIS, the length of time that polymerase pauses at any given site is reduced.
- TFIIIS also contributes to proofreading by polymerase.



# RNA Elongation and Proofreading

- TFIIIS stimulates an inherent Rnase activity in polymerase (not part of the active site), allowing an alternative approach to removing misincorporated bases through local limited RNA degradation.

# RNA Elongation and Proofreading

- This feature is comparable to the hydrolytic editing in the bacterial case stimulated by the Gre factors.

END

# Transcription Termination

- The final RNA processing event, polyadenylation of the 3' end of the mRNA, is intimately linked with the termination of transcription, although exactly how is still not quite clear.

# Transcription Termination

- Recently, however, an enzyme that degrades the second RNA as it emerges from the polymerase has been identified, and this enzyme may itself trigger termination. This is called the **torpedo model** of termination.

# Transcription Termination

- The free end of the second RNA is uncapped and thus can be distinguished from genuine transcripts.

# Transcription Termination

- This new RNA is recognized by an RNase called, in yeast, **Rat1** (in humans, **Xrn2**) that is loaded onto the end of the RNA by another protein (**Rtt103**) that binds the CTD of RNA polymerase.

# Transcription Termination

- The Rat1 enzyme is very processive and quickly degrades the RNA in a 5'-to-3' direction, until it catches up to the still-transcribing polymerase from which the RNA is being spewed.



# Transcription Termination

- Termination may not require very specific interaction between Rat1 and polymerase and might, in fact, be triggered in a manner rather similar to that of Rho-dependent termination in bacteria.

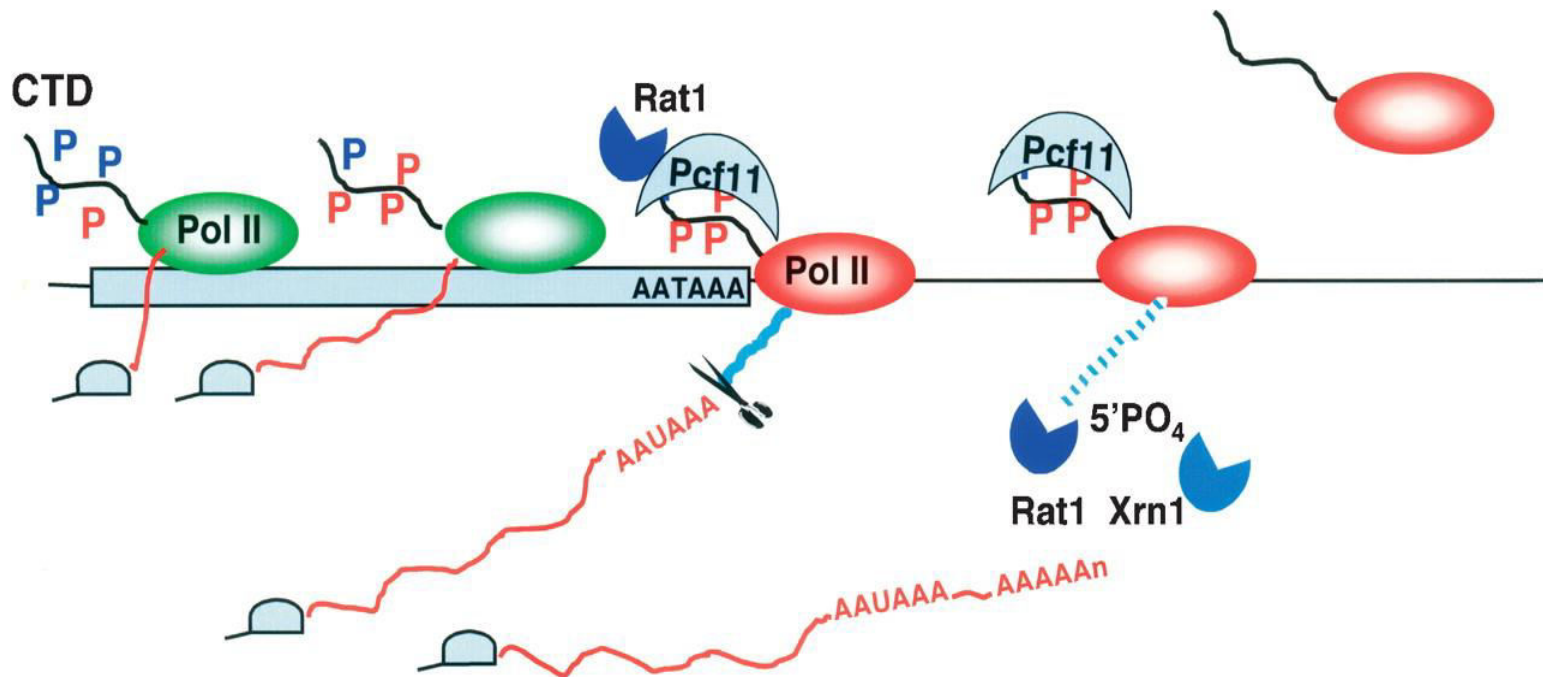
# Transcription Termination

- The highly processing RNase polymerase either pushes polymerase forward and/or pulls the remains of the nascent RNA transcript from the enzyme.

# Transcription Termination

- It is also possible that other factors are needed in addition to Rat1 to dislodge polymerase as, in vitro, Rat1 is alone insufficient to carry out this function, even after it has degraded the transcript.

# Transcription Termination



# Transcription Termination

- Although the torpedo model for termination is now the favored one, there is an alternative called the **allosteric model**.

# Transcription Termination

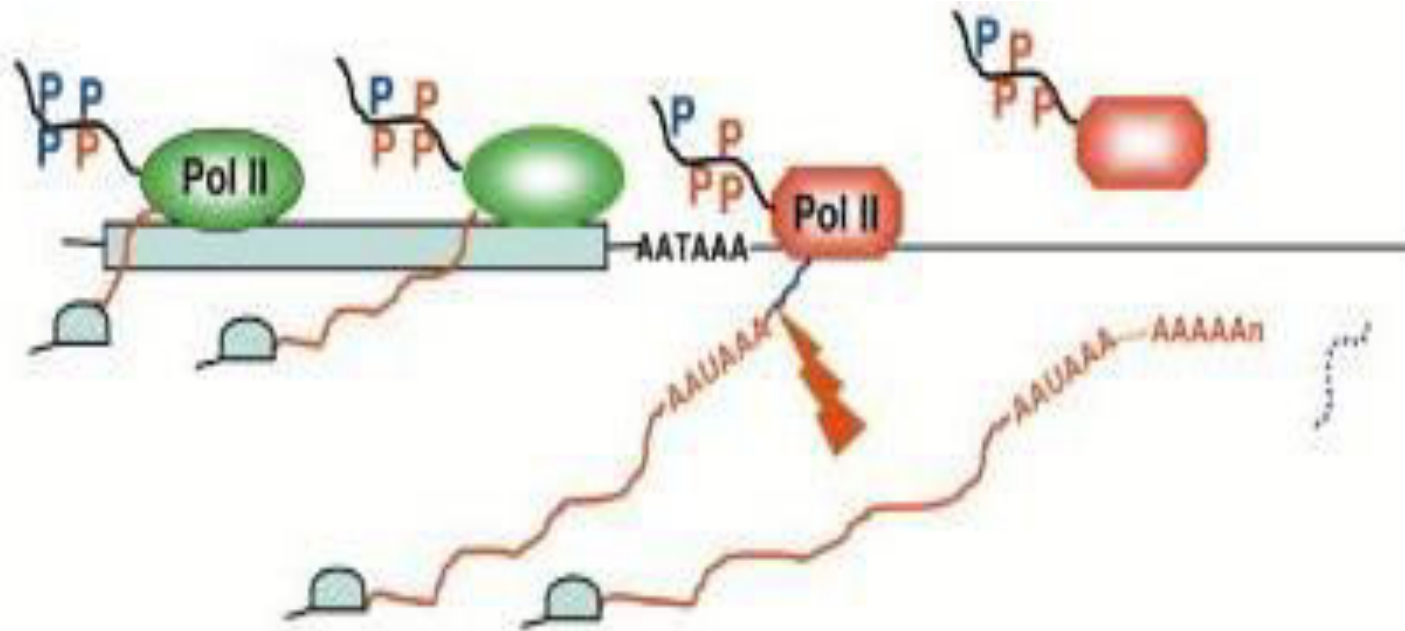
- According to this model, termination depends on a conformational change in the elongating polymerase that reduces the processivity of the enzyme leading to spontaneous termination soon afterward.

# Transcription Termination

- This conformational change would be linked to polyadenylation and could be triggered by the transfer of the 3'-processing enzymes from the CTD tail of polymerase to the RNA.



# Transcription Termination



# Polymerases I and III

- All eukaryotes have two other RNA polymerases— Pol I and Pol III—in addition to Pol II.

# Polymerases I and III

- These enzymes are related to Pol II and even share several subunits but they initiate transcription from distinct promoters and transcribe distinct genes.

# Polymerases I and III

- Those genes encode specialized RNAs rather than proteins. Each of these enzymes also works with its own unique set of general transcription factors.

# Polymerases I and III

- TBP, however, is universal—it is involved in initiating transcription by Pol I and Pol III, as well as Pol II.

# Polymerases I and III

- Although TBP is the only GTF that is used by Pol I and Pol III as well as by Pol II, it has emerged recently that some of the other GTFs, in fact, have structurally and functionally equivalent components in the other systems.

# Polymerases I and III

- Thus, for example, TFIIIF seems to have a counterpart in two subunits within Pol I (A49/34.5), and also in Pol II (C37/53).
- Likewise, TFIIIE-like subunits are found in Pol I and Pol III enzymes.



# Polymerases I and III

- In addition, both these other systems include additional factors comparable to TFIIB: the TAF1B factor in the Pol I system, and the Brf1 subunit of TFIIB in the case of Pol III.

# Polymerases I and III

- Pol I is required for the expression of only one gene, that encoding the rRNA precursor.

# Polymerases I and III

- There are many copies of that gene in each cell, and, indeed, it is expressed at far higher levels than any other gene, perhaps explaining why it has its own dedicated polymerase.

# Polymerases I and III

- The promoter for the rRNA gene comprises two parts: the core element and the UCE (upstream control element).

END

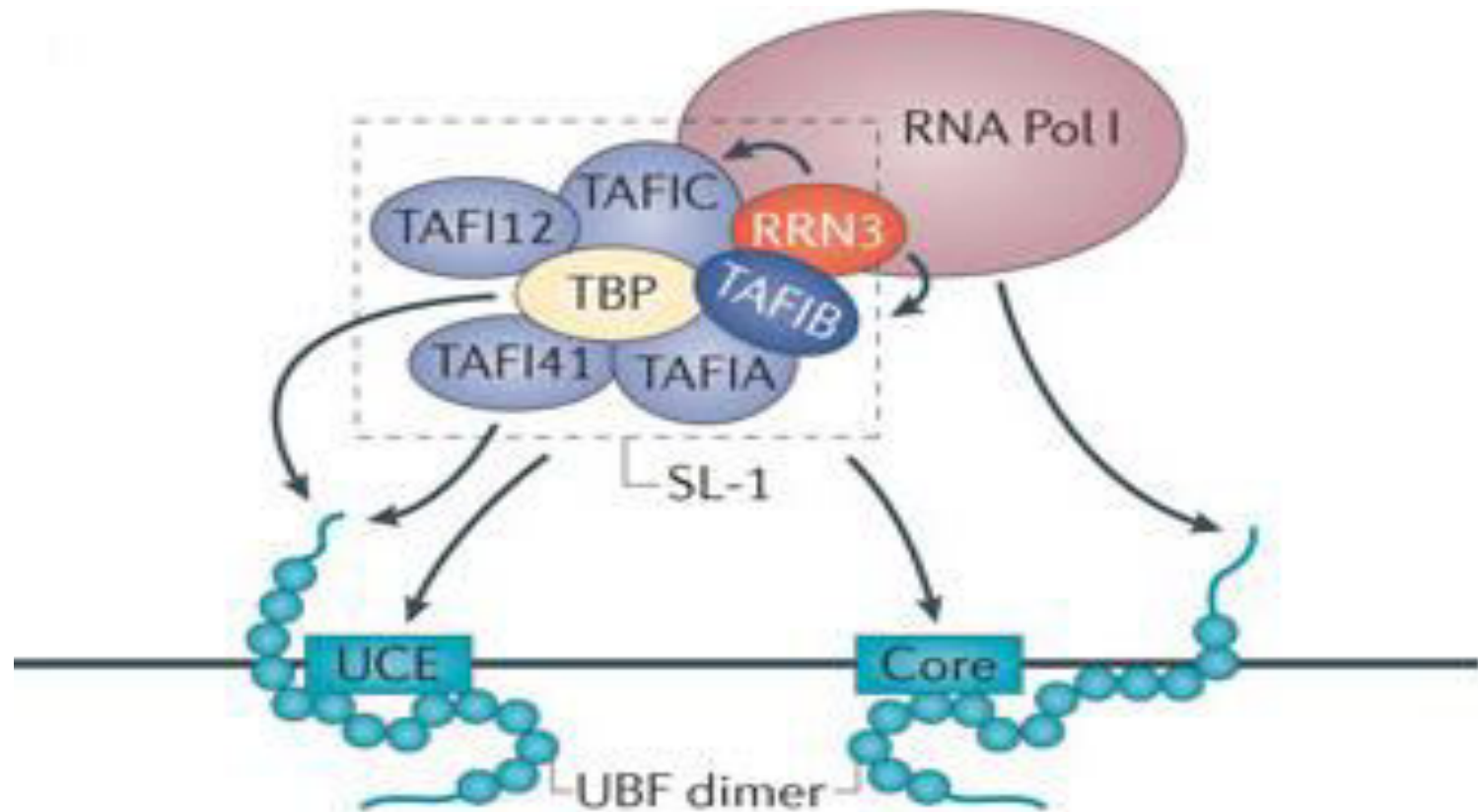
# Polymerases I and III

- The former is located around the start site of transcription, and the latter between 100 and 150 bp upstream (in humans).
- In addition to Pol I, initiation requires two other factors, called SL1 and UBF.

# Polymerases I and III

- SL1 comprises TBP and three TAFs specific for Pol I transcription. This complex binds to the core element.

# Polymerases I and III





# Polymerases I and III

- SL1 binds DNA only in the presence of UBF. This factor binds to UCE, bringing in SL1 and stimulating transcription from the core promoter by recruiting Pol I.

# Polymerases I and III

- Pol III promoters come in various forms, and the vast majority have the unusual feature of being located downstream from the transcription start site (i.e., within the coding region of the gene).

# Polymerases I and III

- Some Pol III promoters (e.g., those for the tRNA genes) consist of two regions, called Box A and Box B, separated by a short element.

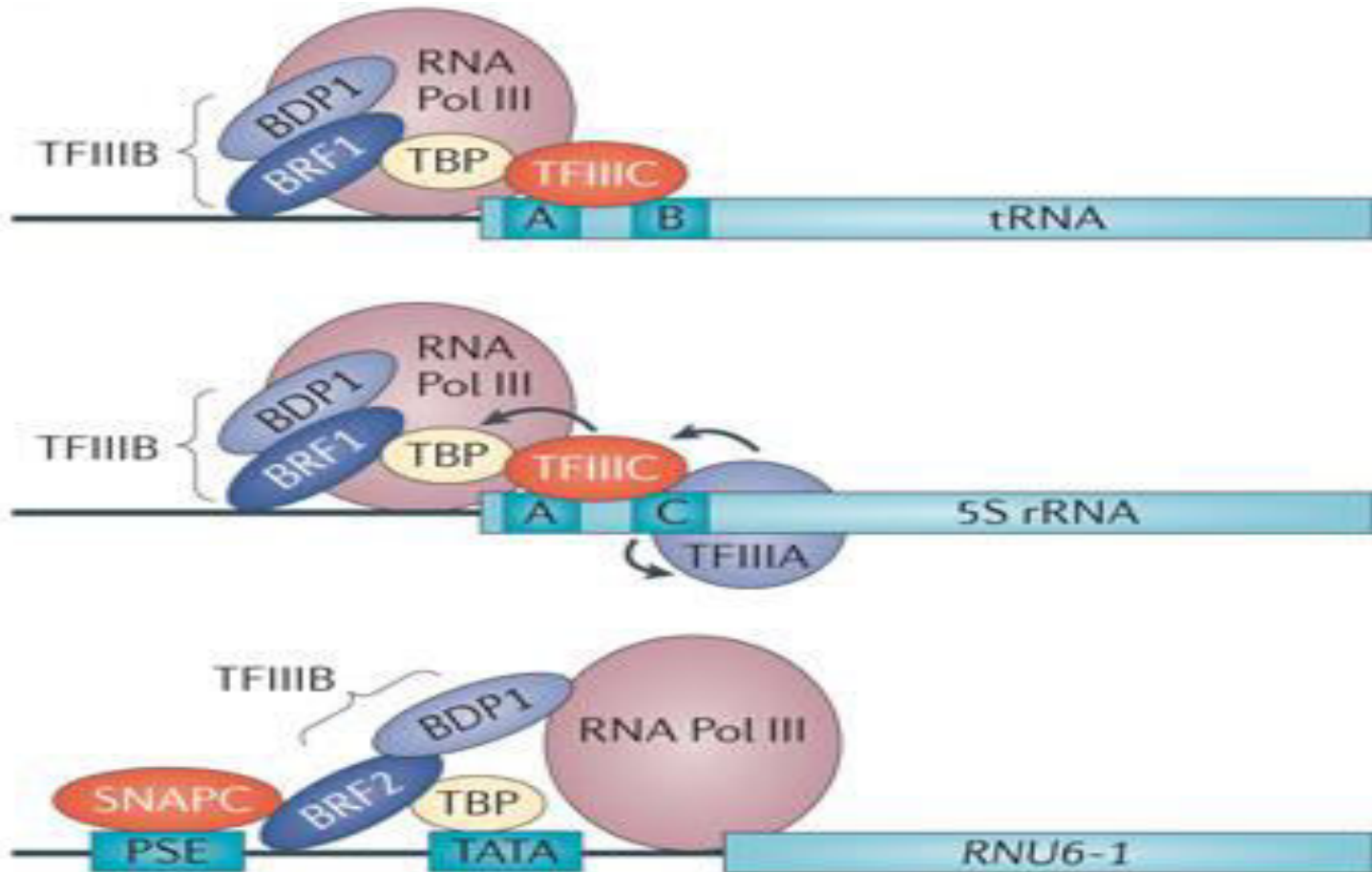
# Polymerases I and III

- Others contain Box A and Box C (e.g., the 5S rRNA gene); and still others contain a TATA element like those of Pol II.

# Polymerases I and III

- Transcription by Pol III also requires transcription factors in addition to polymerase.
- In this case, the factors are called TFIIB and TFIIC for the tRNA genes and those plus TFIIIA for the 5S rRNA gene.

# Polymerases I and III



# Polymerases I and III

- The TFIIIC complex binds to the promoter region. This complex recruits TFIIIB to the DNA just upstream of the start site, where it, in turn, recruits Pol III to the start site of transcription.



# Polymerases I and III

- The enzyme then initiates, presumably displacing TFIIIC from the DNA template as it goes.
- Pol III also uses TBP like other polymerases. In this case, that ubiquitous factor is found within the TFIIB complex.

END

# RNA Splicing

- The coding sequence of a gene is a series of three nucleotide codons that specifies the linear sequence of amino acids in its polypeptide product.

# RNA Splicing

- It is generally assumed that the coding sequence is contiguous; i.e., the codon for one amino acid is immediately adjacent to the codon for the next amino acid in the polypeptide chain.

# RNA Splicing

- This is true in the vast majority of cases in bacteria and their phage. But it is rarely so for eukaryotic genes.
- In those cases, the coding sequence is interrupted by stretches of non-coding sequences.

# RNA Splicing

- Many eukaryotic genes are thus mosaics, consisting of blocks of coding sequences separated from each other by blocks of non-coding sequences.

# RNA Splicing

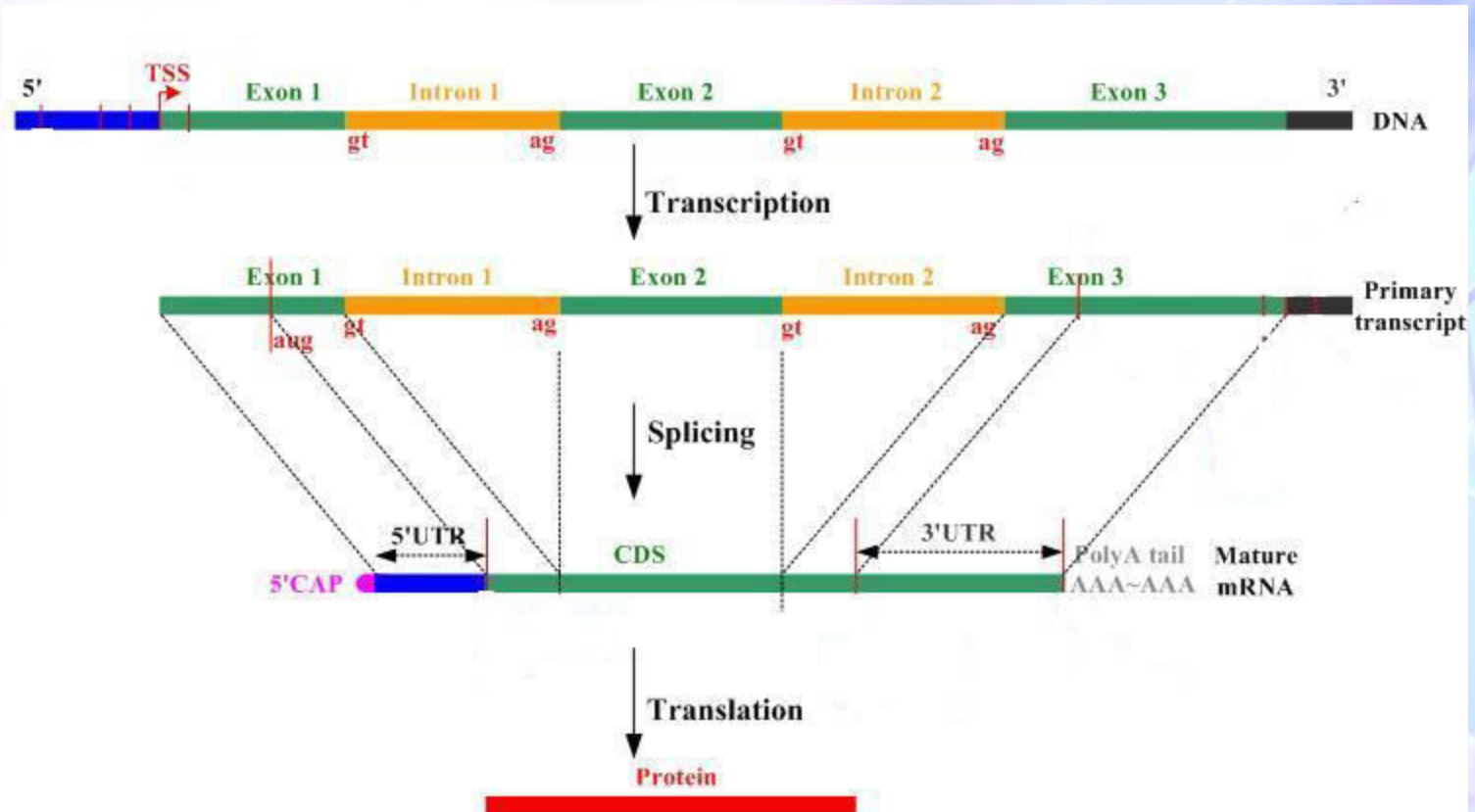
- The coding sequences are called **exons** and the intervening sequences are called **introns**.

# RNA Splicing

- Once transcribed into an RNA transcript, the introns must be removed and the exons joined together to create the mRNA for that gene.



# RNA Splicing



# RNA Splicing

- The number of introns found within a gene varies enormously—from one in the case of most yeast genes (and a few human genes), to 50 in the case of the chicken pro $\alpha$ 2 collagen gene,

# RNA Splicing

- to as many as 363 in the case of the Titin gene of humans.
- The sizes of the exons and introns vary as well.
- Indeed, introns are very often much longer than the exons they separate.

# RNA Splicing

- Thus, for example, exons are typically on the order of 150 nucleotides, whereas introns—although they too can be short—can be as long as 800,000 nucleotides (800 kb).

# RNA Splicing

- As another example, the mammalian gene for the enzyme dihydrofolate reductase is more than 31 kb long, and within it are dispersed six exons that correspond to 2 kb of mRNA.

# RNA Splicing

- Thus, in this case, the coding portion of the gene is, 10% of its total length.

# RNA Splicing

- Like the uninterrupted genes of prokaryotes, the split genes of eukaryotes are transcribed into a single RNA copy of the entire gene - the primary transcript that contains introns as well as exons.



# RNA Splicing

- Because the length and number of introns, the primary transcript (or **pre-mRNA**) can be very long indeed.

# RNA Splicing

- As already mentioned, the primary transcripts of intron-containing genes must have their introns removed before they can be translated into proteins.

# RNA Splicing

- The process of intron removal is called **RNA Splicing**.
- It converts the pre-mRNA into mature mRNA containing only exons.

# RNA Splicing

- RNA Splicing must occur with great precision to avoid the loss, or addition, of even a single nucleotide at the sites at which the exons are joined.

# RNA Splicing

- The triplet-nucleotide codons of mRNA are translated in a fixed reading frame that is set by the first codon.
- Lack of precision in splicing will change the reading frames of exons.

# RNA Splicing

- Some pre-mRNAs can be spliced in more than one way.
- Thus, mRNAs containing different selections of exons can be generated from a given pre-mRNA.

# RNA Splicing

- Alternative splicing strategy enables a gene to give rise to more than one polypeptide product.
- These alternative products are called **isoforms**.



# RNA Splicing

- It is estimated that 90% or more of the protein-coding genes in the human genome are spliced in alternative ways to generate more than one isoform.

# How splicing site is determined?

- Let us consider the molecular mechanisms of the splicing reaction.
- How are the introns and exons distinguished?
- How are introns removed?
- How are exons join with high precision?

# How splicing site is determined?

- The borders between introns and exons are marked by specific nucleotide sequences within the pre-mRNAs.
- These sequences delineate where splicing will occur.

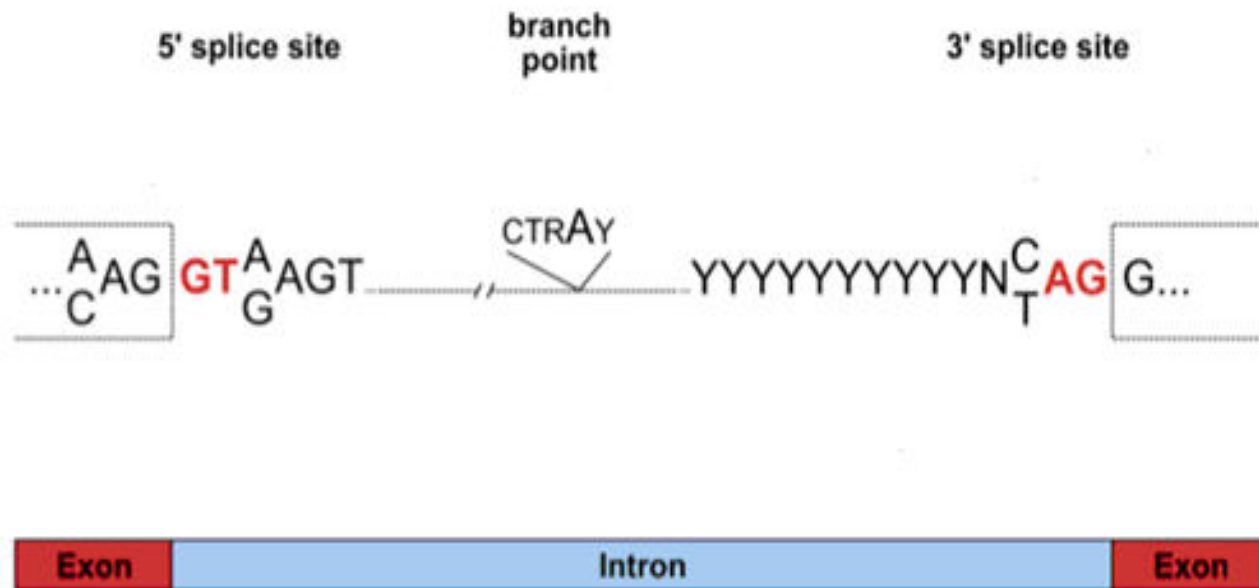
# How splicing site is determined?

- Thus the exon – intron boundary - that is, the boundary at the 5' end of the intron — is marked by a sequence called the **5' splice site**.
- The intron – exon boundary at the 3' end of the intron is marked by the **3' splice site**.

# How splicing site is determined?

- The 5' and 3' splice sites were sometimes referred to as the **donor** and **acceptor** sites, respectively.

# How splicing site is determined?



# How splicing site is determined?

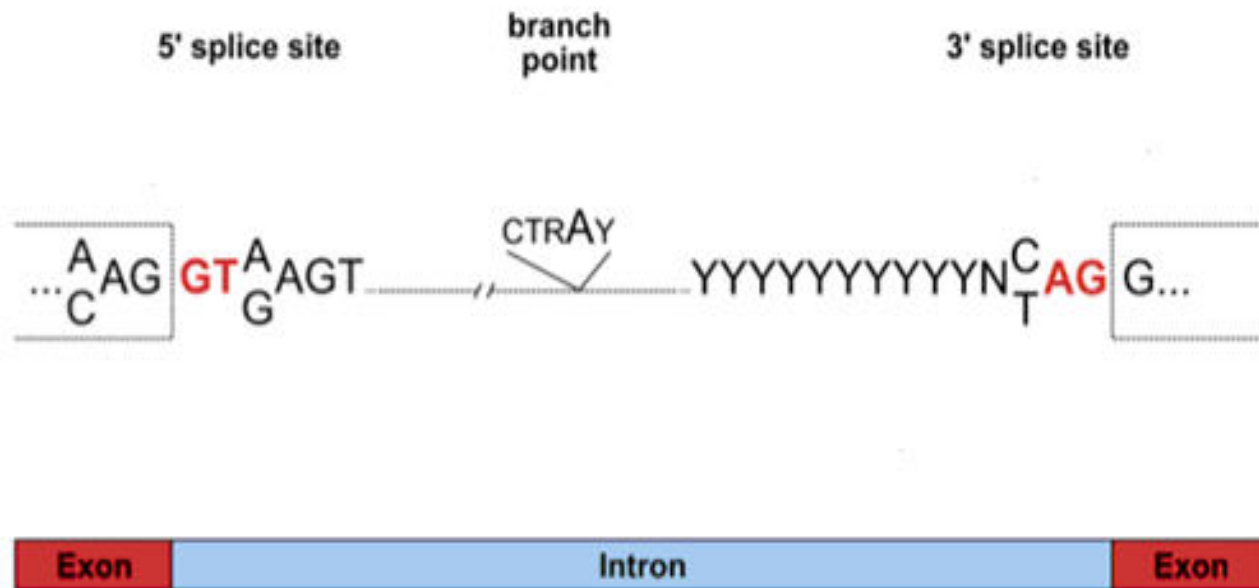
- The figure shows a third sequence necessary for splicing. This is called the **branchpoint site** (or branchpoint sequence).



# How splicing site is determined?

- It is found entirely within the intron, usually close to its 3' end, and is followed by a polypyrimidine tract (Py tract).

# How splicing site is determined?



# How splicing site is determined?

- The most highly conserved sequences are the GU in the 5' splice site, the AG in the 3' splice site, and the A at the branch site.

# How splicing site is determined?

- These highly conserved nucleotide sequences are all found within the intron itself.

# How splicing site is determined?

- This is perhaps because the sequence of most exons, in contrast to the introns, is constrained by the need to encode the specific amino acids of the protein product.

# Removal of Introns

- An intron is removed through two successive transesterification reactions in which phosphodiester linkages within the pre-mRNA are broken and new ones are formed.

# Removal of Introns

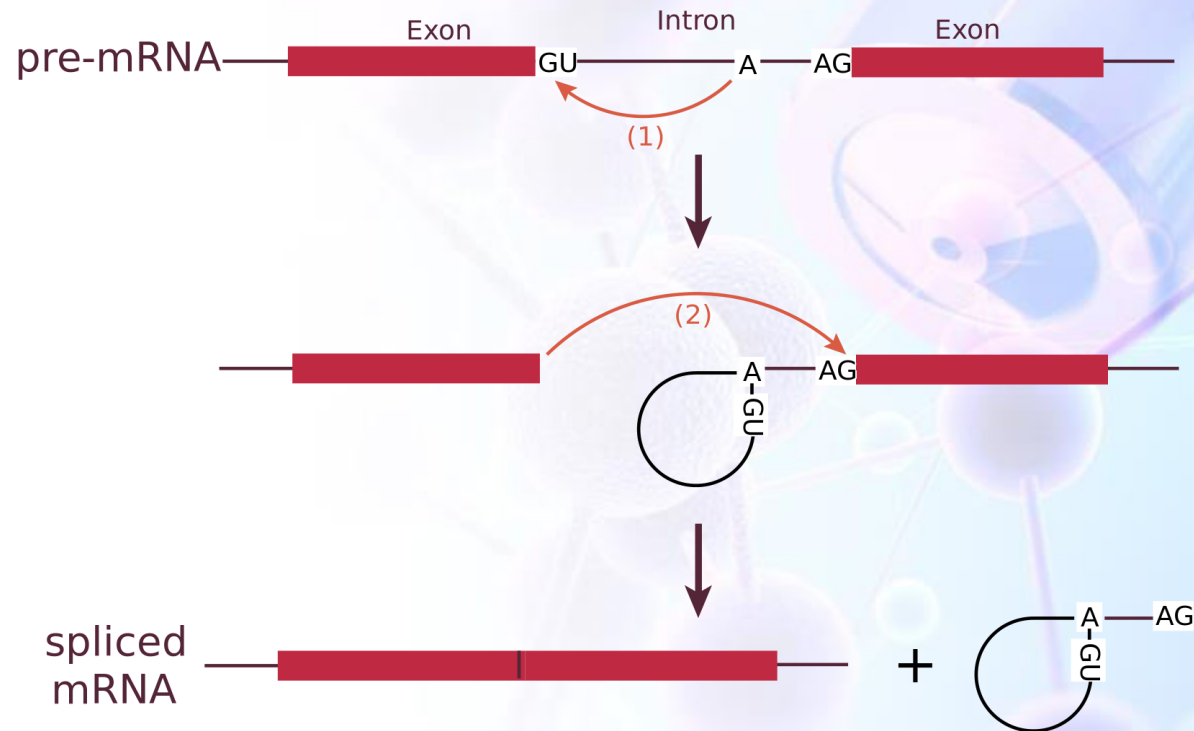
- The first reaction is triggered by the 2'-OH of the conserved A at the branch site. This group acts as a nucleophile to attack the phosphoryl group of the conserved G in the 5' splice site.



# Removal of Introns

- As a consequence of this first reaction, the phosphodiester bond between the sugar and the phosphate at the 5' junction between the intron and the exon is cleaved.
- The freed 5' end of the intron is joined to the A within the branch site.

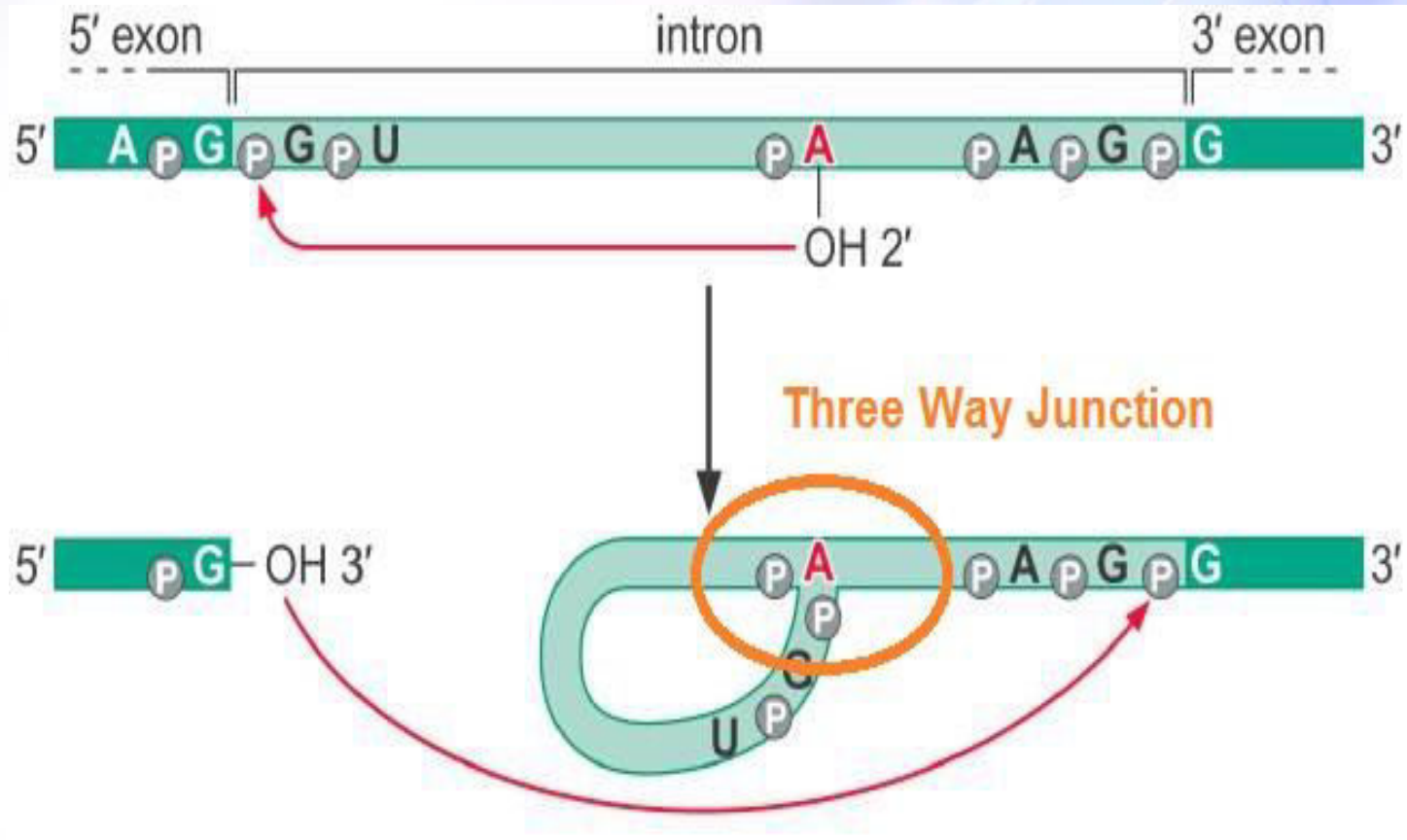
# Removal of Introns



# Removal of Introns

- Thus, in addition to the 5' and 3' backbone linkages, a third phosphodiester extends from the 2'-OH of that A to create a three-way junction (hence its description as a branchpoint).

# Removal of Introns



# Removal of Introns

- Note that the 5' exon is a leaving group in the first transesterification reaction.
- In the second reaction, the 5' exon reverses its role and becomes a nucleophile that attacks the phosphoryl group at the 3' splice site.

# Removal of Introns

- This second reaction has two consequences.
- First, and most importantly, it joins the 5' and 3' exons; thus, this is the step in which the two coding sequences are actually “spliced” together.



# Removal of Introns

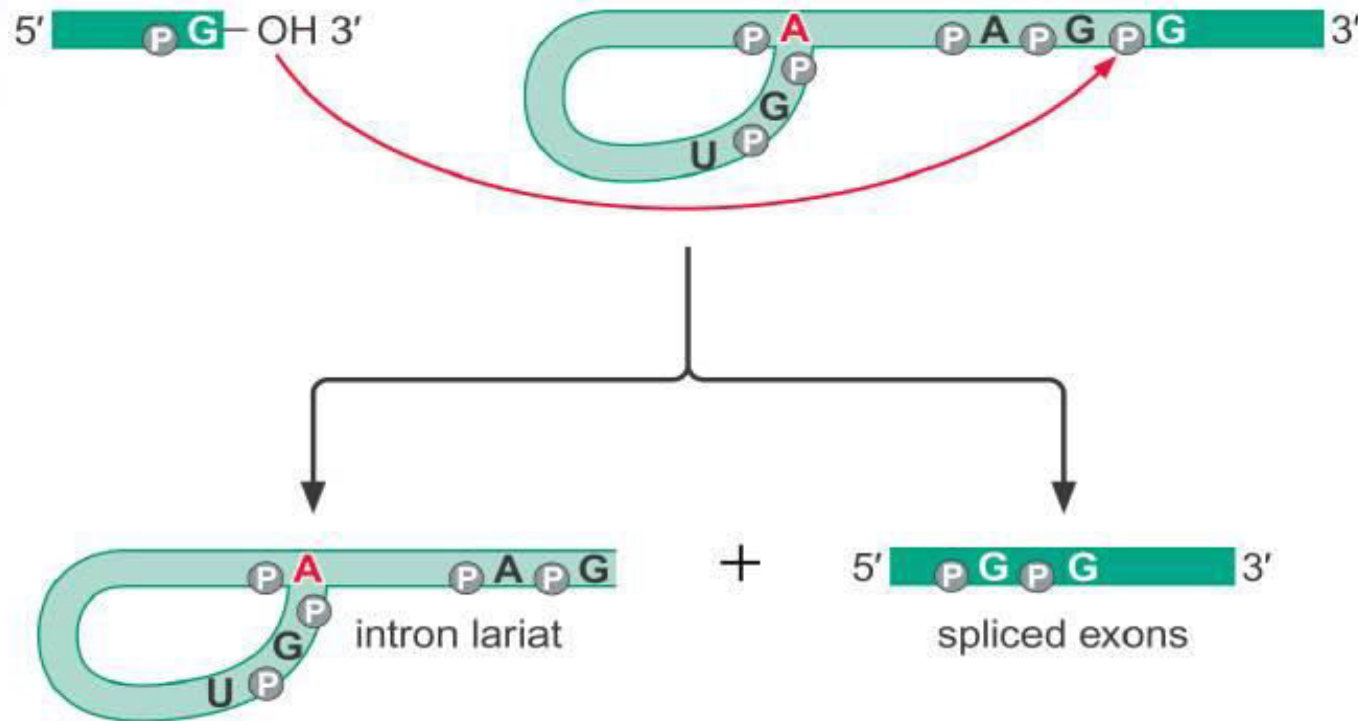
- Second, this same reaction liberates the intron, which serves as a leaving group.



# Removal of Introns

- Because the 5' end of the intron had been joined to branchpoint A in the first transesterification reaction, the newly liberated intron has the shape of a **Lariat**.

# Removal of Introns



Copyright © 2004 Pearson Education, Inc., publishing as Benjamin Cummings

# Removal of Introns

- In the two reaction steps, there is no net gain in the number of chemical bonds—two phosphodiester bonds are broken, and two new ones are made.

# Removal of Introns

- Because it is just a question of shuffling bonds, no energy input is demanded by the chemistry of this process.
- But a large amount of ATP is consumed during the splicing reaction.

# Removal of Introns

- This energy is required, to properly assemble and operate the splicing machinery.
- Another point regarding the splicing reaction is direction: what ensures that splicing only goes forward—that is, toward the products.

# Removal of Introns

- In principle, the reactions could go in the other direction, and indeed this can be forced to happen under special circumstances.
- But in practice, this does not happen in the cell.

# The Spliceosome

- The transesterification reactions are mediated by a huge molecular “machine” called the **Spliceosome**.
- This complex comprises about 150 proteins and five RNAs and is similar in size to a ribosome.



# The Spliceosome

- In performing even a single splicing reaction, the spliceosome hydrolyzes several molecules of ATP.

# The Spliceosome

- Strikingly, it is believed that many of the functions of the spliceosome are performed by its RNA components rather than the proteins.

# The Spliceosome

- Thus, RNAs locate the sequence elements at the intron – exon borders and likely participate in catalysis of the splicing reaction itself.

# The Spliceosome

- The five RNAs (U1, U2, U4, U5, and U6) are collectively called **small nuclear RNAs (snRNAs)**. Each of these RNAs is between 100 and 300 nucleotides long in most eukaryotes and is complexed with several proteins.

# The Spliceosome

- These RNA – protein complexes are called **small nuclear ribonuclear proteins** (snRNPs - pronounced “snurps”).

# The Spliceosome

- The spliceosome is the large complex made up of these snRNPs, but the exact makeup differs at different stages of the splicing reaction.

# The Spliceosome

- Different snRNPs come and go at different times, each performing particular functions in the reaction.



# The Spliceosome

- There are also many proteins within the spliceosome that are not part of the snRNPs, and others besides that are only loosely bound to the spliceosome.
- The snRNPs have three roles in splicing:-

# The Spliceosome

- They recognize the 5' splice site and the branch site; they bring those sites together as required; and they catalyze the RNA cleavage and joining reactions.

# The Spliceosome

- To perform these functions, RNA–RNA, RNA–protein, and protein–protein interactions are all important.
- Let us consider some of the RNA– RNA interactions.

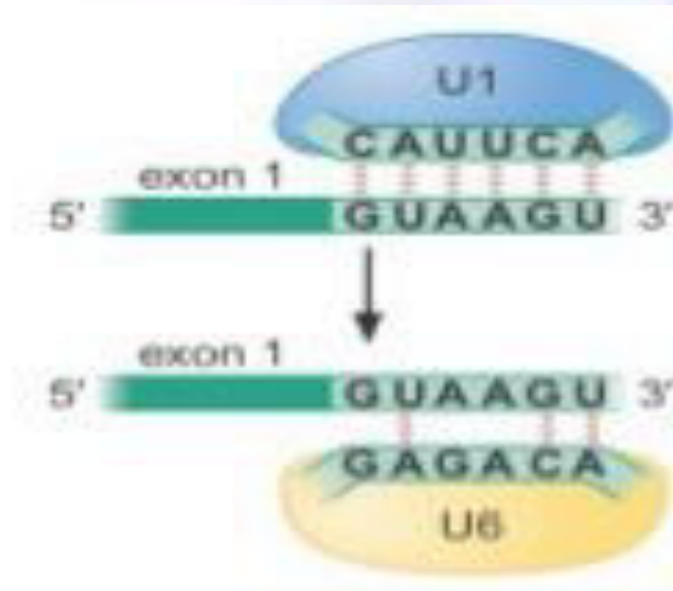
# The Spliceosome

- These interactions operate within individual snRNPs, between different snRNPs, and between snRNPs and the pre-mRNA.

# The Spliceosome

- The figure below shows the interaction, through complementary base pairing, of the U1 snRNA and the 5' splice site in the pre-mRNA.
- Subsequently in the reaction, that splice site is recognized by the U6 snRNA.

# The Spliceosome

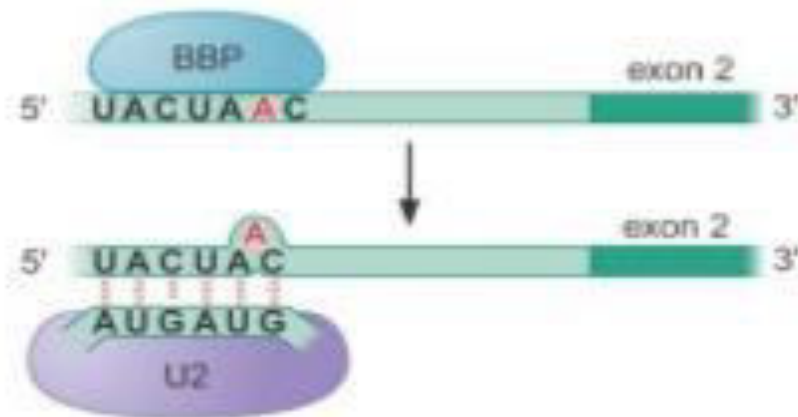
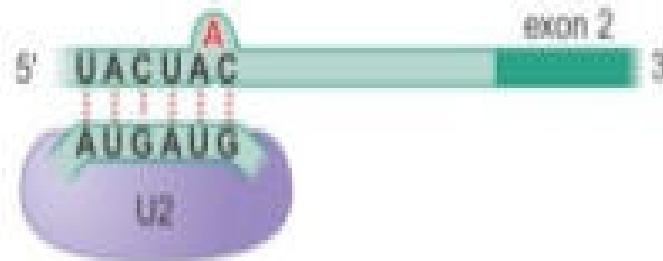
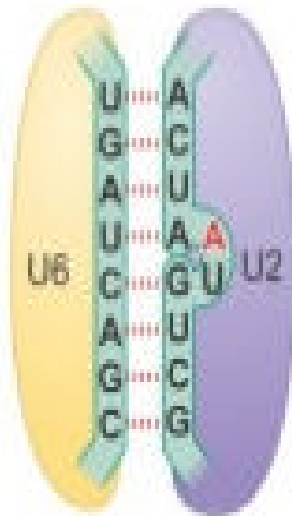


# The Spliceosome

- In another example, shown below, the branch site is recognized by the U2 snRNA.
- A third example shows an interaction between U2 and U6 snRNAs.
- This brings the 5' splice site and the branch site together.



# The Spliceosome



# The Spliceosome

- It is these and other similar interactions, and the rearrangements they lead to, that drive the splicing reaction and contribute to its precision.

# The Spliceosome

- Some non-snRNPs are also involved in splicing. One example, U2AF (U2 auxiliary factor), recognizes the polypyrimidine (Py) tract/3' splice site.

# The Spliceosome

- And then in the initial step of the splicing reaction, it helps another protein, branchpoint-binding protein (BBP), bind to the branch site.
- BBP (also called SF1) is then displaced by the U2 snRNP.

# The Spliceosome

- Other proteins involved in the splicing reaction include RNA-annealing factors, which help load snRNPs onto the mRNA, and DEAD-box helicase proteins.

# The Spliceosome

- The latter use their ATPase activity to dissociate given RNA - RNA interactions, allowing alternative pairs to form and thereby driving the rearrangements that occur through the splicing reaction.

# The Spliceosome

- They are also required to remove spliced mRNA from the spliceosome and trigger spliceosome disassembly.



# The Splicing Pathway

- In a splicing pathway, a number of the steps may differ slightly in their order or might even reverse.
- But the pathway reveals the series of events undertaken by the spliceosome to drive the splicing reaction in the cell.

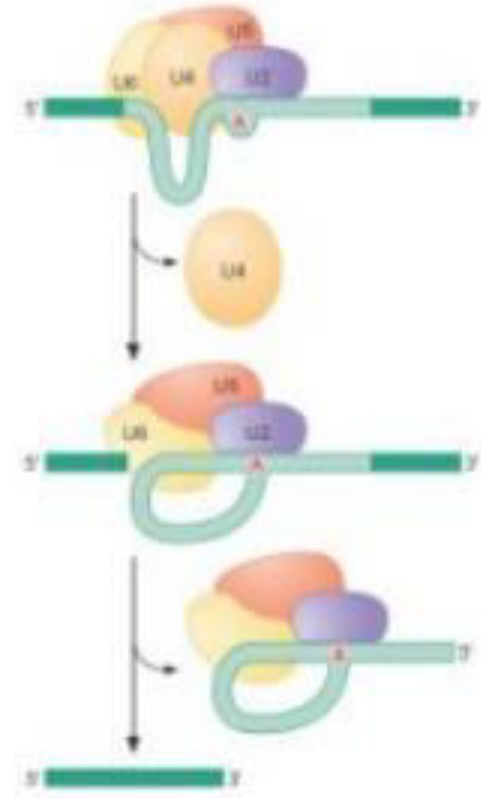
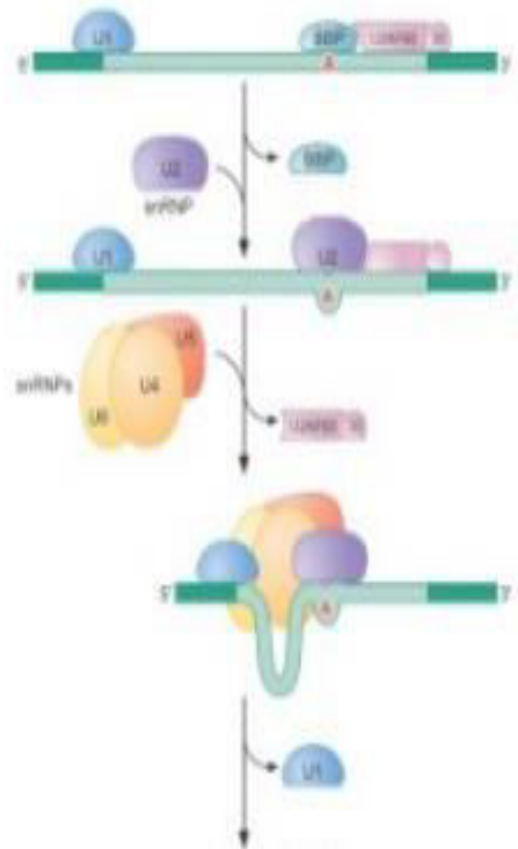
# The Splicing Pathway

- Initially, the 5' splice site is recognized by the U1 snRNP.
- U2AF is made up of two subunits, the larger of which (65) binds to the Py tract and the smaller (35) binds to the 3' splice site.

# The Splicing Pathway

- The former subunit interacts with BBP (SF1) and helps that protein bind to the branch site. This arrangement of proteins and RNA is called the early (E) complex.

# The Splicing Pathway



# The Splicing Pathway

- The former subunit interacts with BBP (SF1) and helps that protein bind to the branch site. This arrangement of proteins and RNA is called the **early (E) complex**.

# The Splicing Pathway

- U2 snRNP then binds to the branch site, aided by U2AF and displacing BBP (SF1). This arrangement is called the **A complex**.

# The Splicing Pathway

- The base pairing between the U2 snRNA and the branch site is such that the branch site A residue is extruded from the resulting stretch of double-helical RNA as a single nucleotide bulge.



# The Splicing Pathway

- This A residue is thus unpaired and available to react with the 5' splice site.
- The next step is a rearrangement of the A complex to bring together all three splice sites.

# The Splicing Pathway

- This is achieved as follows: the U4 and U6 snRNPs, along with the U5 snRNP, join the complex.
- Together, these three snRNPs are called the **tri-snRNP particle**.

# The Splicing Pathway

- Within this particle, the U4 and U6 snRNPs are held together by complementary base pairing between their RNA components, and the U5 snRNP is more loosely associated through protein-protein interactions.

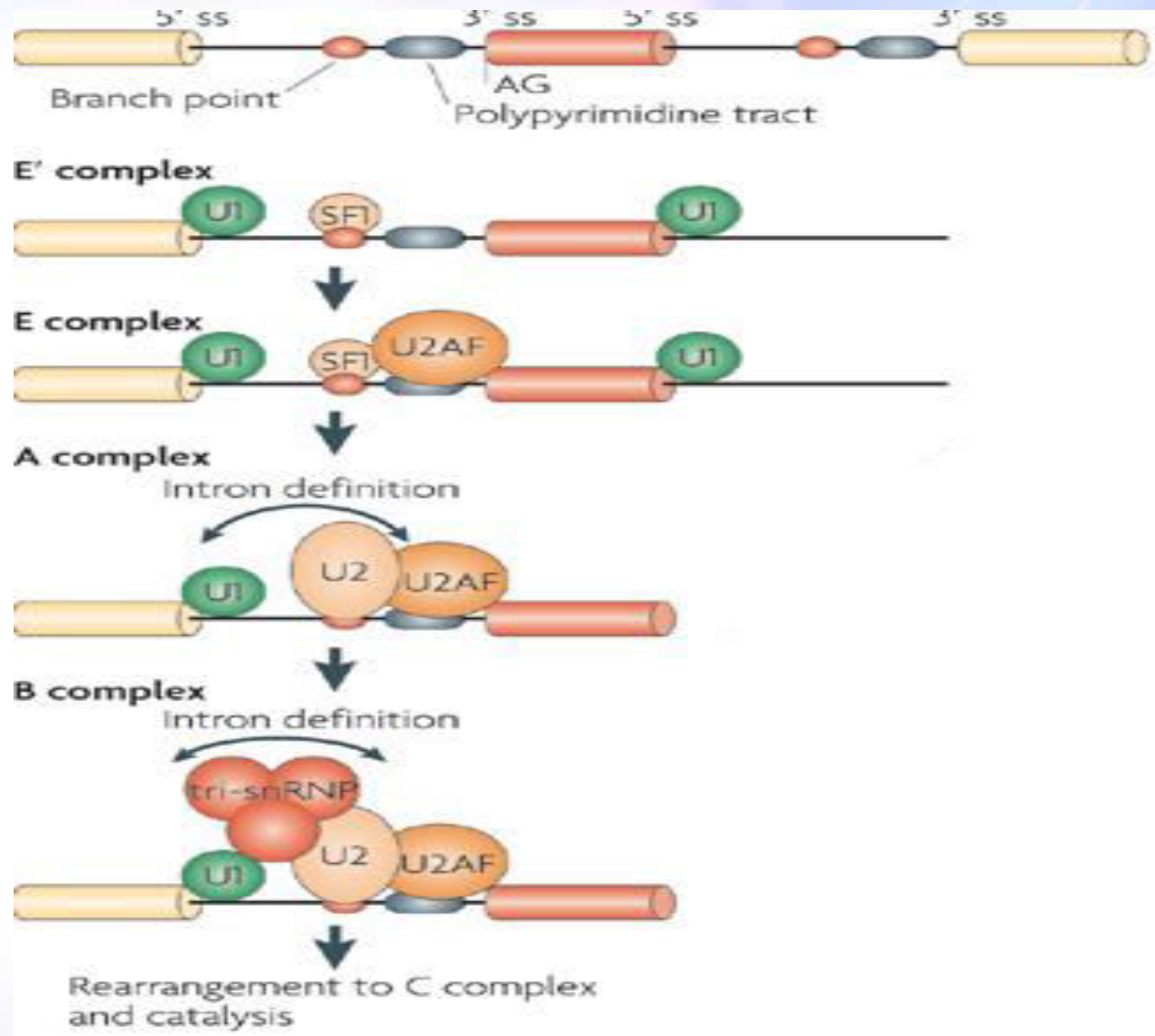
# The Splicing Pathway

- With the entry of the tri-snRNP, the A complex is converted into the **B complex**.
- In the next step, U1 leaves the complex, and U6 replaces it at the 5' splice site.

# The Splicing Pathway

- This requires that the base pairing between the U1 snRNA and the pre-mRNA be broken, allowing the U6 RNA to anneal with the same region.
- Those steps complete the assembly pathway.

# The Splicing Pathway





# The Splicing Pathway

- The next rearrangement triggers catalysis and occurs as follows:-
- U4 is released from the complex, allowing U6 to interact with U2 (through the RNA:RNA base pairing).



# The Splicing Pathway

- This arrangement, called the **C complex**, produces the active site. That is, the rearrangement brings together within the spliceosome those components that together form the active site.

# The Splicing Pathway

- The same rearrangement also ensures that the substrate RNA is properly positioned to be acted upon.

# The Splicing Pathway

- It is striking not only that the active site is primarily formed of RNA, but also that it is only formed at this stage of spliceosome assembly.
- Presumably, this strategy lessens the chance of aberrant splicing.

# The Splicing Pathway

- Linking the formation of the active site to the successful completion of earlier steps in spliceosome assembly makes it highly likely that the active site is available only at legitimate splice sites.

# The Splicing Pathway

- Formation of the active site juxtaposes the 5' splice site of the pre-mRNA and the branch site, facilitating the first transesterification reaction.

# The Splicing Pathway

- The second reaction, between the 5' and 3' splice sites, is aided by the U5 snRNP, which helps to bring the two exons together.

# The Splicing Pathway

- The final step involves release of the mRNA product and the snRNPs. The snRNPs are initially still bound to the lariat, but they get recycled after rapid degradation of that piece of RNA.



# Self-Splicing Introns

- Self-Splicing Introns Reveal That RNA Can Catalyze RNA Splicing.
- There are total three classes of splicing found in the cells:-
  - Nuclear pre-mRNA
  - Group II introns
  - Group I introns

# Self-Splicing Introns

- Thus far, we have dealt only with nuclear pre-mRNA splicing, that mediated by the spliceosome found in all eukaryotes.
- The so-called Group I and Group II are self-splicing introns.

# Self-Splicing Introns

- By “self- splicing” we mean that the intron itself folds into a specific conformation within the precursor RNA and catalyzes the chemistry of its own release.

# Self-Splicing Introns

- In terms of a practical definition, “self-splicing” refers to introns that can remove themselves from RNAs in the test tube in the absence of any proteins or other RNA molecules.

# Self-Splicing Introns

- The self splicing introns are grouped into two classes on the basis of their structure and splicing mechanism.
- Strictly speaking, self-splicing introns are not enzymes because they mediate only one round of RNA processing.

# Self-Splicing Introns

- In the case of group II introns, the chemistry of splicing and the RNA intermediates produced are the same as those for nuclear pre-mRNAs.

# Self-Splicing Introns

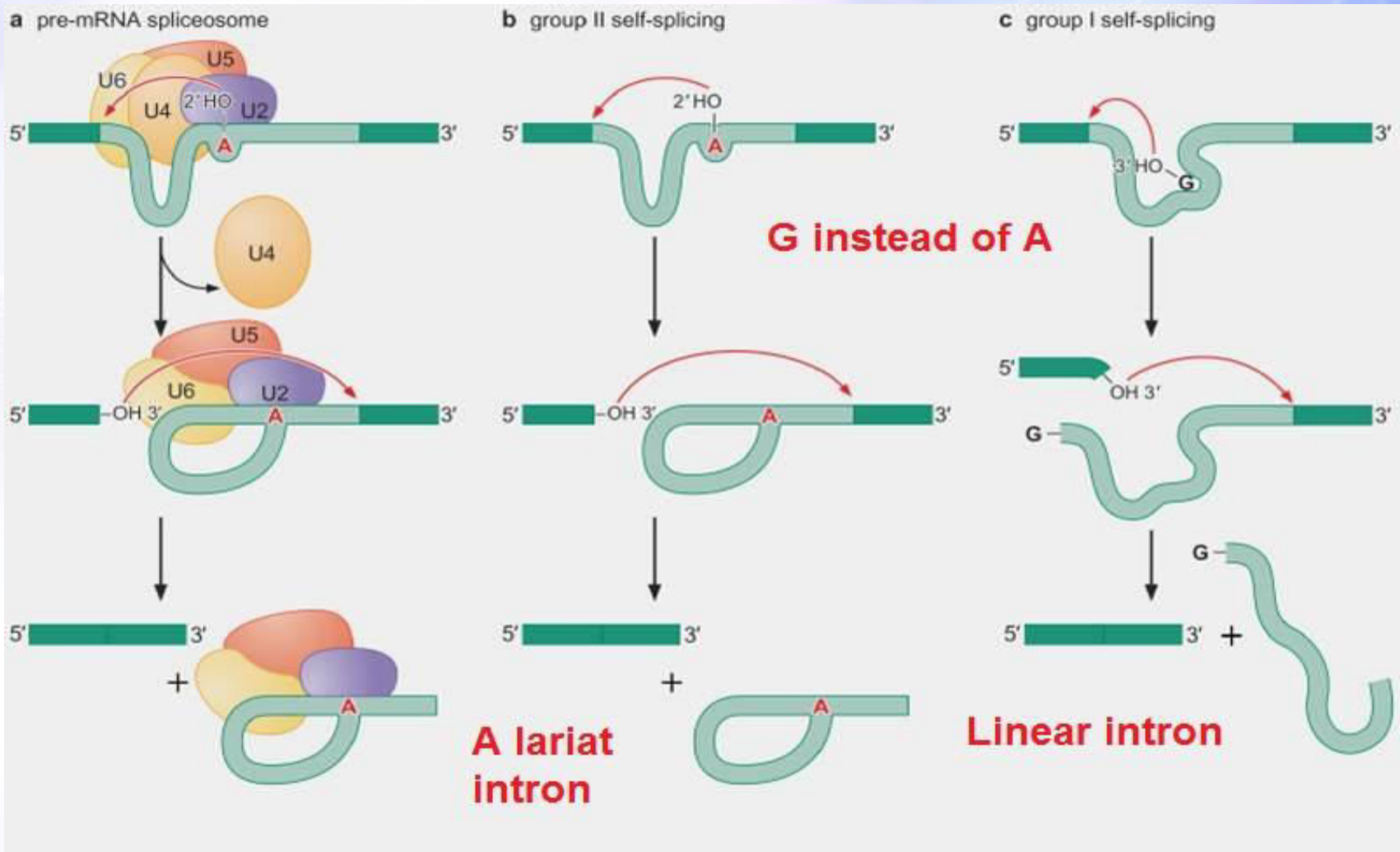
- For example, the intron uses an A residue within the branch site to attack the phosphodiester bond at the boundary between its 5' end and the end of the 5' exon - that is, at the 5' splice site.



# Self-Splicing Introns

- This reaction produces the branched lariat and is followed by a second reaction in which the newly freed 3'-OH of the exon attacks the 3' splice site, releasing the intron as a lariat and fusing the 3' and 5' exons.

# Self-Splicing Introns



# Group I Introns

- Group I introns splice by a different pathway.
- Instead of a branchpoint A residue, they use a free G nucleotide or nucleoside.

# Group I Introns

- This G species is bound by the RNA, and its 3'-OH group is presented to the 5' splice site.
- The same type of transesterification reaction that leads to the lariat formation fuses the G to the 5' end of the intron.

# Group I Introns

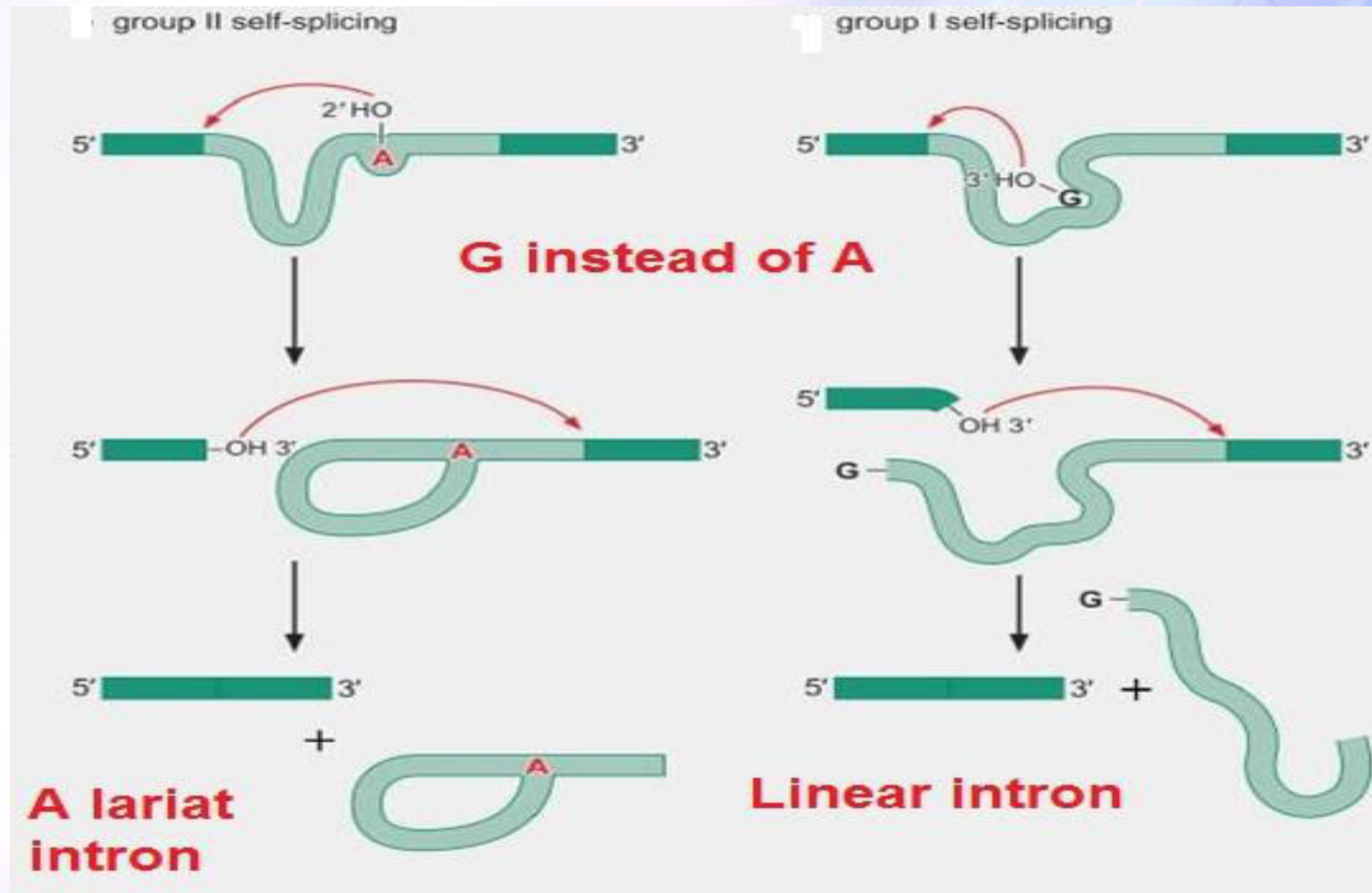
- The second reaction now proceeds just as it does in the earlier examples: the freed 3' end of the exon attacks the 3' splice site.

# Group I Introns

- This fuses the two exons and releases the intron, although, in this case, the intron is linear rather than a lariat structure.



# Group I Introns





# Group I Introns

- Group I introns, which are smaller than group II introns, share a conserved secondary structure.
- The structure of group I introns includes a binding pocket that will accommodate any guanine nucleotide or nucleoside.

# Group I Introns

- In addition, group I introns contain an “internal guide sequence” that base-pairs with the 5' splice site sequence and thereby determines the precise site at which nucleophilic attack by the G nucleotide takes place

# Group I Introns

- A typical self-splicing intron is between 400 and 1000 nucleotides long, and, in contrast to introns removed by spliceosomes, much of the sequence of a self-splicing intron is critical for the splicing reaction.

# Group I Introns

- This sequence requirement holds because the intron must fold into a precise structure to perform the reaction chemistry.

# Group I Introns

- In addition, in vivo, the intron is complexed with several proteins that help stabilize the correct structure — partly by shielding regions of the backbone from each other.

# Group I Introns

- Thus, the folding requires certain sections of the RNA backbone to be in close proximity to other sections, and the negative charges provided by the phosphates in those backbone regions would repel each other if not shielded.

# Group I Introns

- The similar chemistry seen in self- and spliceosome-mediated splicing is believed to reflect an evolutionary relationship.



# Group I Introns

- Perhaps ancestral group II – like self-splicing introns were the starting point for the evolution of modern pre-mRNA splicing.

# Group I Introns

- The catalytic functions provided by the RNA were retained, but the requirement for extensive sequence specificity within the intron itself was relieved.

# Group I Introns

- In this way, introns had only to retain the minimum of sequence elements required to target splicing to the correct places.

# Group I Introns

- The structure of the catalytic region that performs the first transesterification reaction is very similar in the group II intron and the pre-mRNA/snRNP complex.

# Group I Introns

- This observation, supports the speculation that early in the evolution of modern organisms, many catalytic functions in the cell were performed by RNAs, and that these functions have been replaced by proteins.

# Finding the Splice Sites Correctly

- We have already seen one mechanism that guards against inappropriate splicing: the active site of the spliceosome is only formed on RNA sequences that pass the test of being recognized by multiple elements.

# Finding the Splice Sites Correctly

- Thus, for example, the 5' splice site must be recognized initially by the U1 snRNP and then by the U6 snRNP.
- It is unlikely that both would recognize an incorrect sequence.



# Finding the Splice Sites Correctly

- Splice-site recognition is prone to two kinds of errors.
- First, splice sites can be skipped, with components bound at, for example, a given 5' splice site pairing with those at a 3' site beyond the correct one.

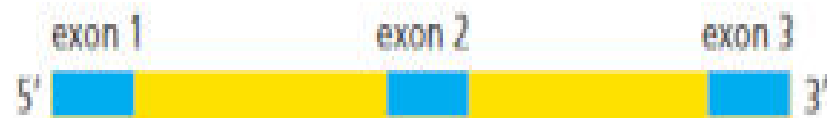
# Finding the Splice Sites Correctly

- Second, other sites, close in sequence but not legitimate splice sites, could be mistakenly recognized because the splice site consensus sequences are rather loose.

# Finding the Splice Sites Correctly

- Therefore, for example, components at a given 5' splice site might pair with components bound incorrectly at such a “pseudo” 3' splice site.

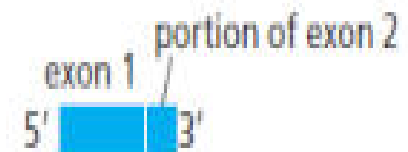
# Finding the Splice Sites Correctly



exon skipping



cryptic splice-site selection



# Finding the Splice Sites Correctly

- Two ways in which the accuracy of splice-site selection can be enhanced are as follows:-

# Finding the Splice Sites Correctly

- First, while transcribing a gene to produce the RNA, RNA polymerase II carries with it various proteins with roles in RNA processing.
- These include proteins involved in splicing.

# Finding the Splice Sites Correctly

- When a 5' splice site is encountered in the newly synthesized RNA, the factors that recognize that site are transferred from the polymerase carboxy-terminal "tail" onto the RNA.



# Finding the Splice Sites Correctly

- Once in place, the 5' splice site components are poised to interact with those other factors that bind to the next 3' splice site to be synthesized.

# Finding the Splice Sites Correctly

- Thus, the correct 3' splice site can be recognized before any competing sites further downstream have been transcribed. This co-transcriptional loading process greatly diminishes the likelihood of exon skipping.

# Finding the Splice Sites Correctly

- A second mechanism guards against the use of incorrect sites by ensuring that splice sites close to exons are recognized preferentially.

# Finding the Splice Sites Correctly

- So-called **SR** (Serine Arginine Rich) proteins bind to the sequences called **exonic splicing enhancers (ESEs)** within the exons.

# Finding the Splice Sites Correctly

- **SR proteins** bound to these sites interact with the components of the splicing machinery, recruiting them to the nearby splice sites.

# Finding the Splice Sites Correctly

- In this way, the machinery binds more efficiently to those splice sites than to incorrect sites not close to exons.

# Finding the Splice Sites Correctly

- Specifically, the SR proteins recruit the U2AF proteins to the 3' splice site and U1 snRNP to the 5' splice site.
- These factors demarcate the splice sites for the rest of the machinery to assemble correctly.



# Finding the Splice Sites Correctly

- SR proteins are essential for splicing.
- They not only ensure the accuracy and efficiency of constitutive splicing but also regulate alternative splicing.

# Alternative Spliceosome

- Higher eukaryotes (mammals, plants, etc.) use the major splicing machinery to direct splicing of the majority of their pre-mRNAs.

# Alternative Spliceosome

- But in these organisms (unlike in yeast), some pre-mRNAs are spliced by an alternative, low - abundance form of the spliceosome.

# Alternative Spliceosome

- This rare form of spliceosome contains some components common to the major spliceosome, but it contains other unique components as well.

# Alternative Spliceosome

- Thus, U11 and U12 components of the alternative spliceosome have the same roles in the splicing reaction as U1 and U2 of the major form, but they recognize distinct sequences.

# Alternative Spliceosome

- U4 and U6 have equivalent counterparts in both spliceosome forms — although these snRNPs are distinct, they share the same names.

# Alternative Spliceosome

- Finally, the identical U5 component is found in both the major and the alternative or so called **minor spliceosome**.



# Alternative Spliceosome

- The minor spliceosome recognizes rarely occurring introns having consensus sequences distinct from the sequences of most pre-mRNA introns.

# Alternative Spliceosome

- It should be emphasized that although these introns are rare, they are widely distributed; approximately 800 human genes contain at least one minor intron.

# Alternative Spliceosome

- Furthermore, mutations in minor snRNAs have recently been found to underlie some rare human genetic diseases.

# Alternative Spliceosome

- The minor form of the spliceosome is also known as the AT-AC spliceosome, because the termini of the originally identified rare introns contain AU at the 5' splice site and AC at the 3' site (or AT and AC in DNA).

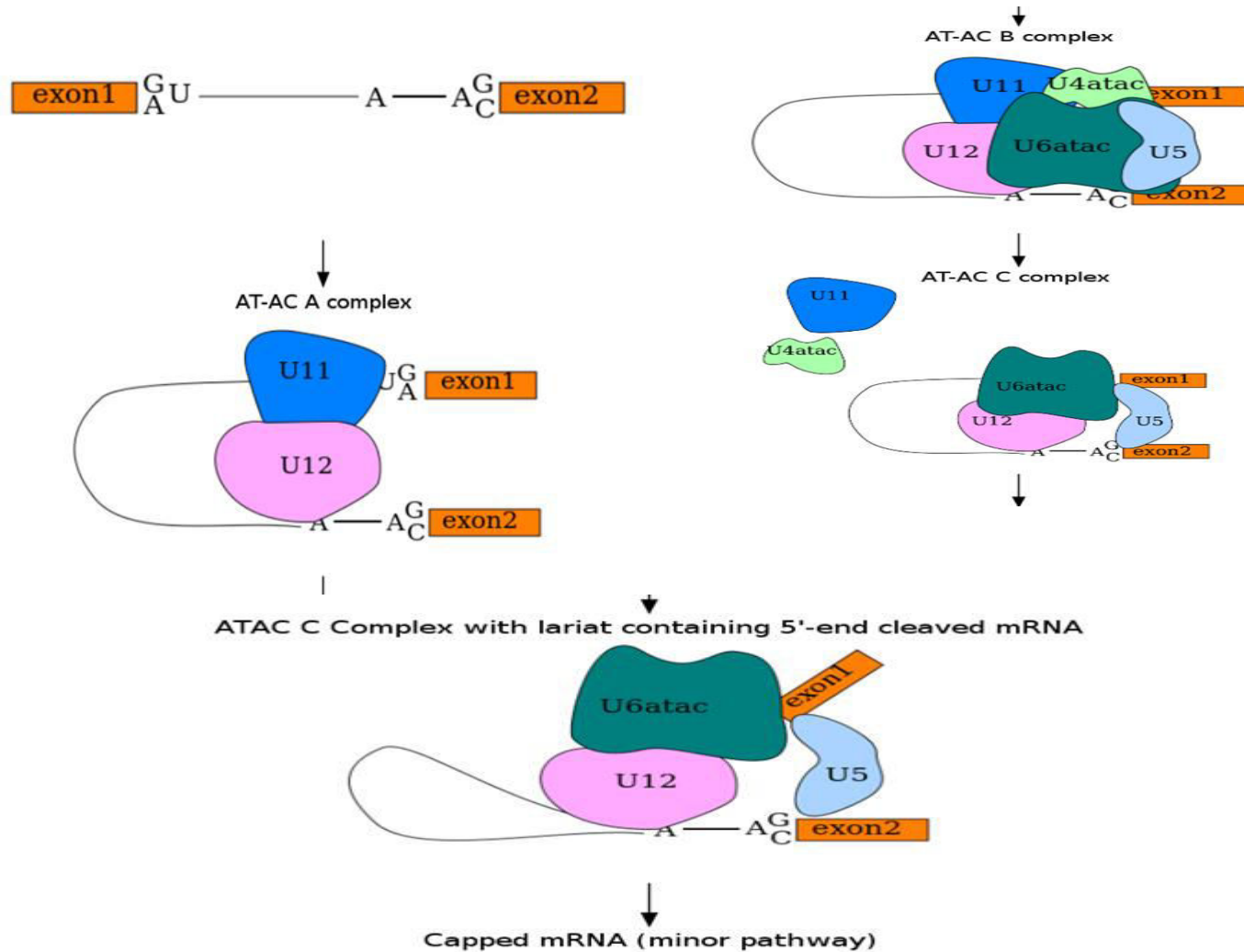
# Alternative Spliceosome

- Later it was found that many introns spliced by this pathway have GT-AG termini (like mainstream introns), but otherwise their consensus sequences are distinct from those of the major pathway.

# Alternative Spliceosome

- Despite the different splice site and branchsite sequences recognized by the two systems, these major and minor forms of spliceosomes both remove introns using the same chemical pathway.

# Alternative Spliceosome





# Alternative Spliceosome

- Consistent with this conserved mechanism, the differences in splice-site sequences recognized by these snRNPs are mirrored by complementary differences in the sequences of their snRNAs.

# Alternative Spliceosome

- Thus, it is the ability of the snRNAs and splice-site sequences to base-pair that is conserved, not any particular sequence within either.

END

# Alternative Splicing

- Many genes in higher eukaryotes encode RNAs that can be spliced in alternative ways to generate two or more different mRNAs and thus different protein products (isoforms).

# Alternative Splicing

- It is now believed that at least 40% of *Drosophila* genes and as many as 90% of human genes undergo alternative splicing.

# Alternative Splicing

- Many alternatively spliced genes generate only two alternative products, but in some cases,

# Alternative Splicing

- the number of potential alternatives that can be generated from a single gene is breathtaking - hundreds (e.g., in the human *Slo* gene) or even many thousands (for the *Drosophila Dscam* gene).

# Alternative Splicing

- Alternative splicing is sometimes used as a way of generating diversity, with alternative forms being generated stochastically.



# Alternative Splicing

- But in many cases, the process is regulated to ensure that different protein products are made in different cell types or in different conditions.

# Alternative Splicing

- For a simple case of alternative splicing, consider the gene for the mammalian muscle protein troponin T.
- The pre mRNA made from this gene that contains five exons.

# Alternative Splicing

- This pre-RNA is spliced to form two alternative mature mRNAs, each containing four exons.
- A different exon is eliminated from each of the two mRNAs, thus the two messages have three exons in common.

# Alternative Splicing



# Alternative Splicing

- Alternative splicing can occur in a number of ways. Thus, in addition to alternative exons, exons can be extended.
- In other cases, exons can be skipped (deliberately), or introns can be retained in the mature message.

# Alternative Splicing

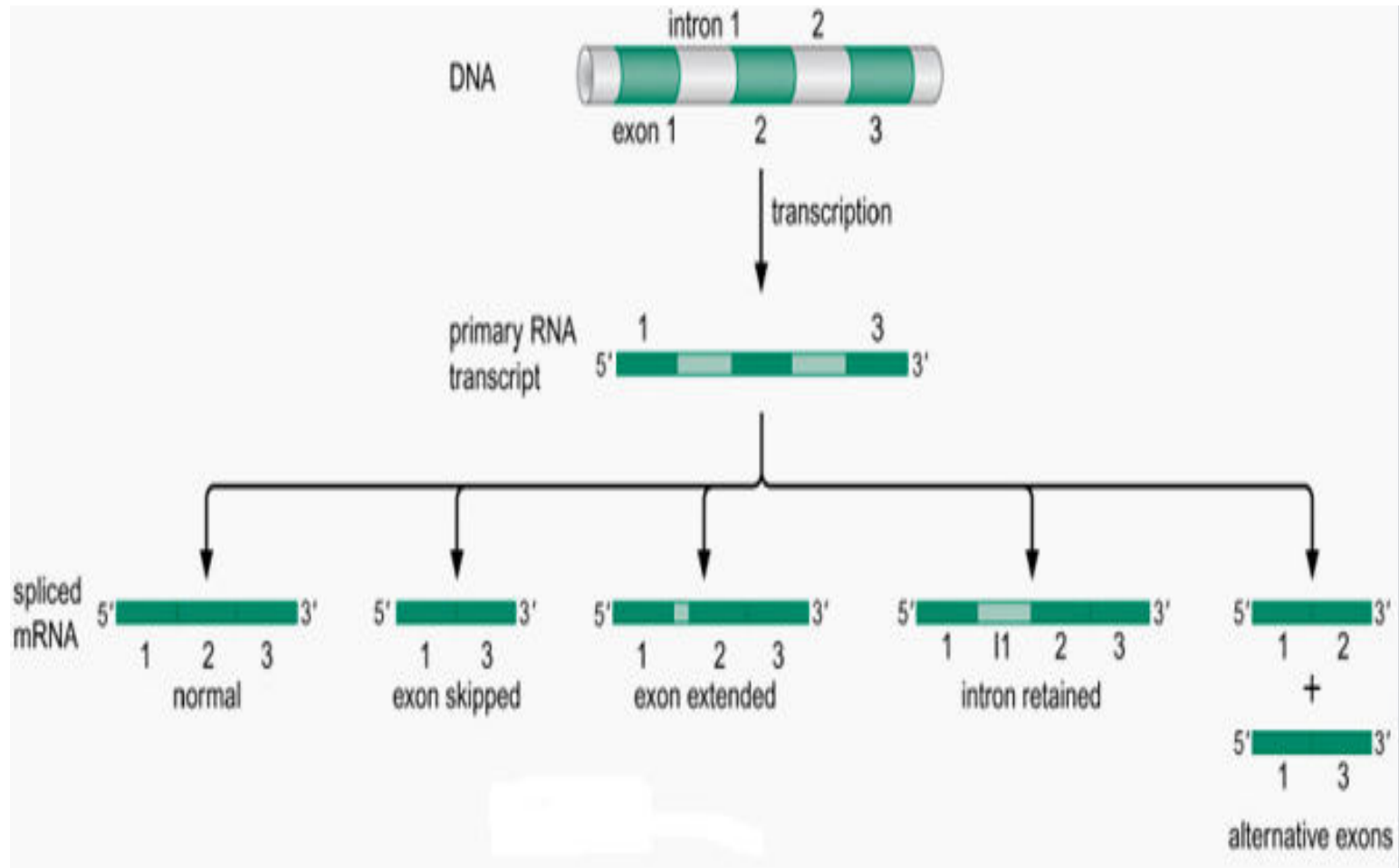
- Some alternative splicing results from transcription of a gene from alternative promoters, allowing one transcript to include a 5' exon not present in the other.

# Alternative Splicing

- Similarly, alternative poly-A sites allow 3' terminal exons to be extended or alternative 3' terminal exons to be used in some transcripts of a given gene. There are even cases of alternative *trans* splicing.



# Alternative Splicing



# Alternative Splicing

- The case of T antigen of the monkey virus SV40 shows an example of an extended exon.
- The T-antigen gene encodes two protein products: the large T antigen (T-ag) and the small t antigen (t-ag).

# Alternative Splicing

- The two proteins result from alternative splicing of the pre-mRNAs from the same gene.
- Thus the gene has two exons, and different mature mRNAs result from the use of two different 5' splice sites.

# Alternative Splicing

- In the mRNA encoding T-ag, exon 1 is spliced directly to exon 2, deleting the intron that lies between.
- The mRNA for t-ag, on the other hand, is formed using the alternative 5' splice site within the intron.

# Alternative Splicing

- Thus, in this case, the mRNA includes some of the intron as well.
- The reason this larger message encodes the smaller protein is because there is an in-frame stop codon within the region of the intron retained in this mRNA.

# Alternative Splicing

- Both forms of T antigens are made in a cell infected by SV40 but have different functions.
- Large T induces transformation and cell cycle re-entry, whereas small t blocks the apoptotic response of cells forced down that path.

# Alternative Splicing

- The ratio of the two forms produced differs depending on the level of the splicing regulator SF2/ASF.



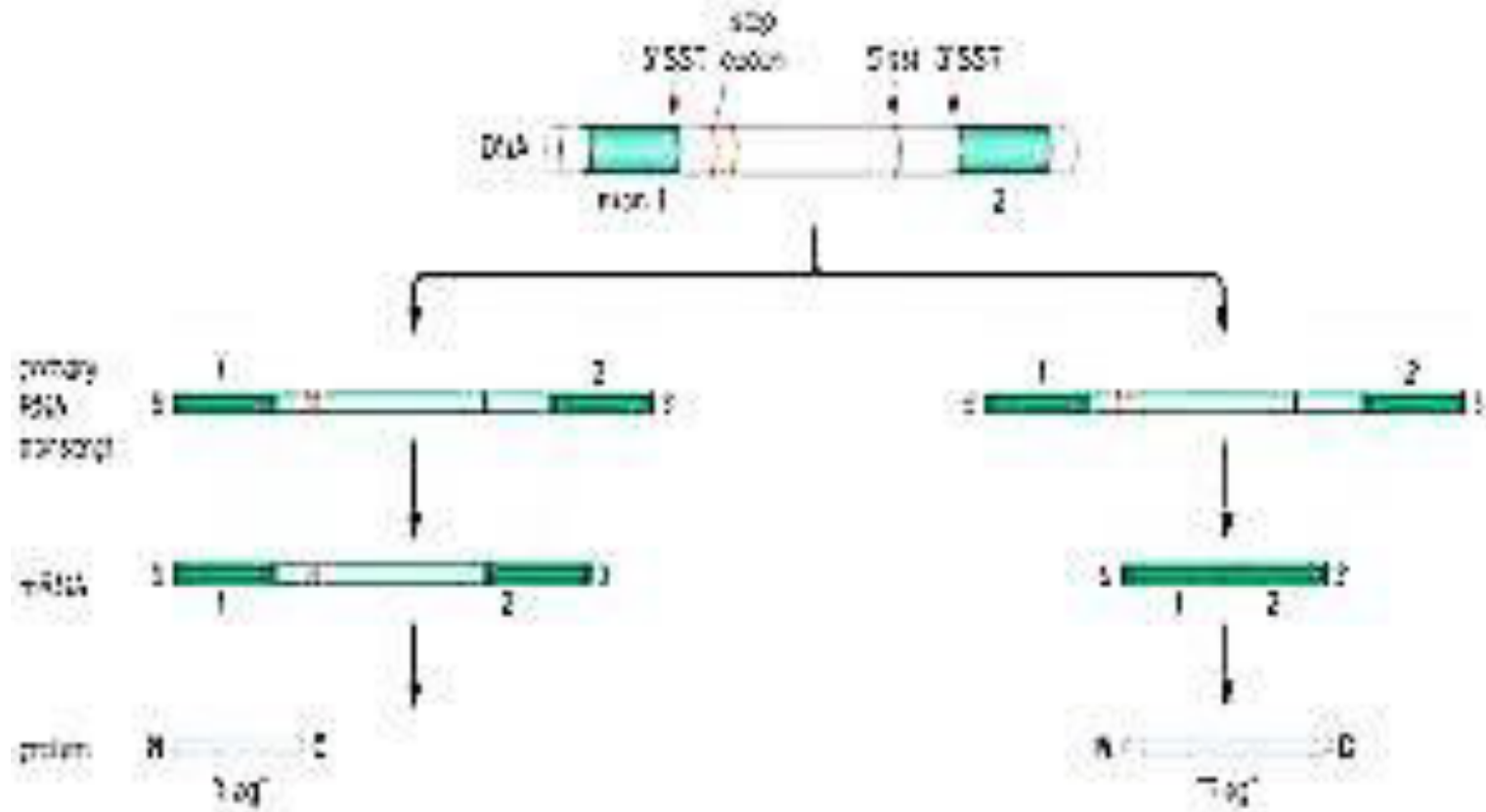
# Alternative Splicing

- When present at high levels, this protein directs the machinery to favor use of the 5' splice site that generates more of the t-ag mRNA.

# Alternative Splicing

- SF2/ASF is an SR protein and, when abundant, presumably binds sites within exon 2 and helps the spliceosome assemble there.

# Alternative Splicing



# Alternative Splicing

- In genome-wide studies, the most commonly seen forms of alternative splicing are cases in which complete exons are included or excluded from the mature message. Such exons are often called **cassette exons**.

# Alternative Splicing

- In ~10% of cases, cassette exons come in pairs, only one of which is included in the spliced message.
- In these cases, there must be mechanisms that ensure that the exons are spliced in a mutually exclusive fashion.

END

# Alternative Splicing

- Many genes in higher eukaryotes encode RNAs that can be spliced in alternative ways to generate two or more different mRNAs and thus different protein products (isoforms).

# Alternative Splicing

- It is now believed that at least 40% of *Drosophila* genes and as many as 90% of human genes undergo alternative splicing.



# Alternative Splicing

- Many alternatively spliced genes generate only two alternative products, but in some cases,

# Alternative Splicing

- the number of potential alternatives that can be generated from a single gene is breathtaking - hundreds (e.g., in the human *Slo* gene) or even many thousands (for the *Drosophila Dscam* gene).

# Alternative Splicing

- Alternative splicing is sometimes used as a way of generating diversity, with alternative forms being generated stochastically.

# Alternative Splicing

- But in many cases, the process is regulated to ensure that different protein products are made in different cell types or in different conditions.

# Alternative Splicing

- For a simple case of alternative splicing, consider the gene for the mammalian muscle protein troponin T.
- The pre mRNA made from this gene that contains five exons.

# Alternative Splicing

- This pre-RNA is spliced to form two alternative mature mRNAs, each containing four exons.
- A different exon is eliminated from each of the two mRNAs, thus the two messages have three exons in common.

# Alternative Splicing





# Alternative Splicing

- Alternative splicing can occur in a number of ways. Thus, in addition to alternative exons, exons can be extended.
- In other cases, exons can be skipped (deliberately), or introns can be retained in the mature message.

# Alternative Splicing

- Alternative splicing can occur in a number of ways. Thus, in addition to alternative exons, exons can be extended.
- In other cases, exons can be skipped (deliberately), or introns can be retained in the mature message.

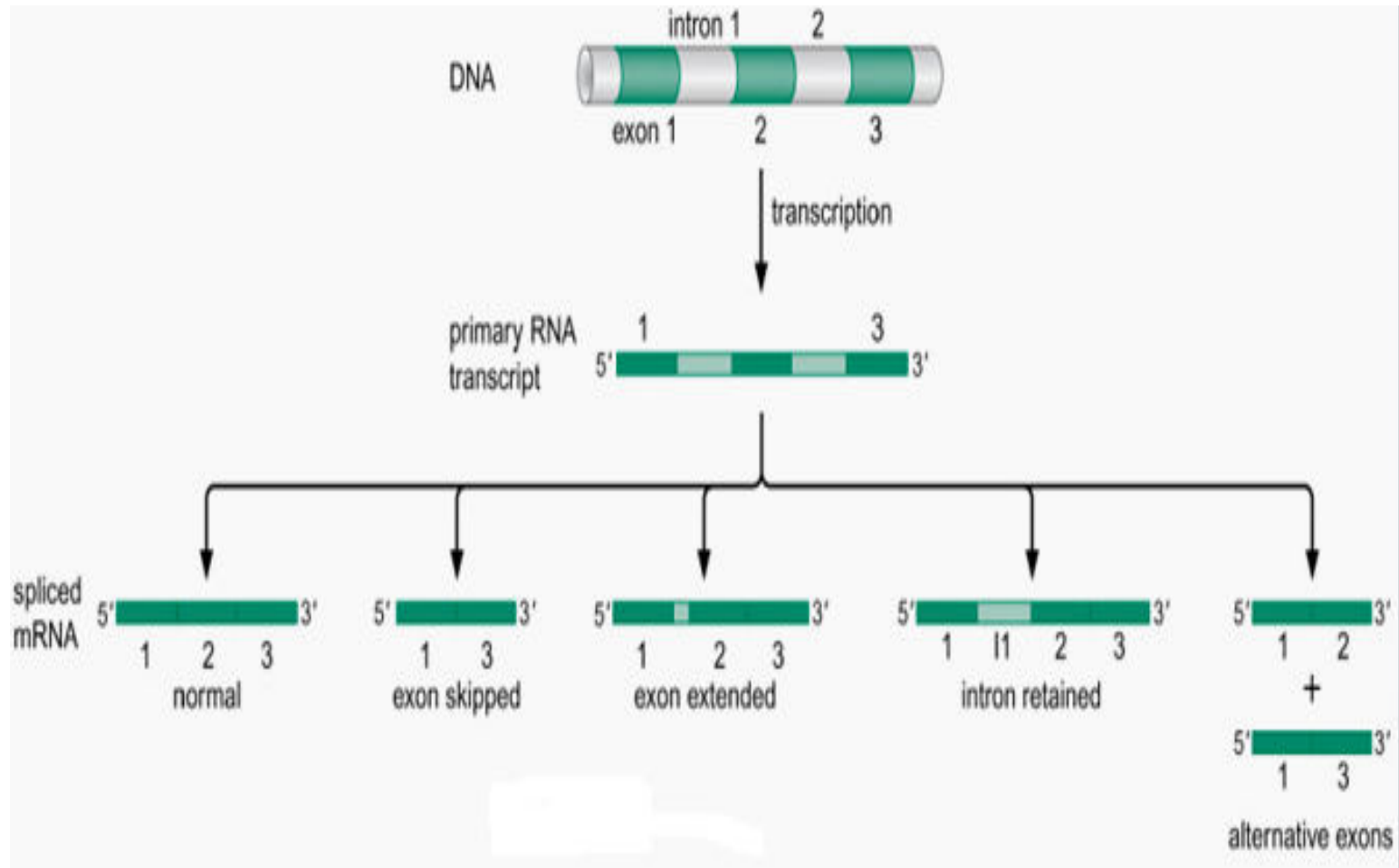
# Alternative Splicing

- Some alternative splicing results from transcription of a gene from alternative promoters, allowing one transcript to include a 5' exon not present in the other.

# Alternative Splicing

- Similarly, alternative poly-A sites allow 3' terminal exons to be extended or alternative 3' terminal exons to be used in some transcripts of a given gene. There are even cases of alternative *trans* splicing.

# Alternative Splicing



# Alternative Splicing

- The case of T antigen of the monkey virus SV40 shows an example of an extended exon.
- The T-antigen gene encodes two protein products: the large T antigen (T-ag) and the small t antigen (t-ag).

# Alternative Splicing

- The two proteins result from alternative splicing of the pre-mRNAs from the same gene.
- Thus the gene has two exons, and different mature mRNAs result from the use of two different 5' splice sites.



# Alternative Splicing

- In the mRNA encoding T-ag, exon 1 is spliced directly to exon 2, deleting the intron that lies between.
- The mRNA for t-ag, on the other hand, is formed using the alternative 5' splice site within the intron.

# Alternative Splicing

- Thus, in this case, the mRNA includes some of the intron as well.
- The reason this larger message encodes the smaller protein is because there is an in-frame stop codon within the region of the intron retained in this mRNA.

# Alternative Splicing

- Both forms of T antigens are made in a cell infected by SV40 but have different functions.
- Large T induces transformation and cell cycle re-entry, whereas small t blocks the apoptotic response of cells forced down that path.

# Alternative Splicing

- The ratio of the two forms produced differs depending on the level of the splicing regulator SF2/ASF.

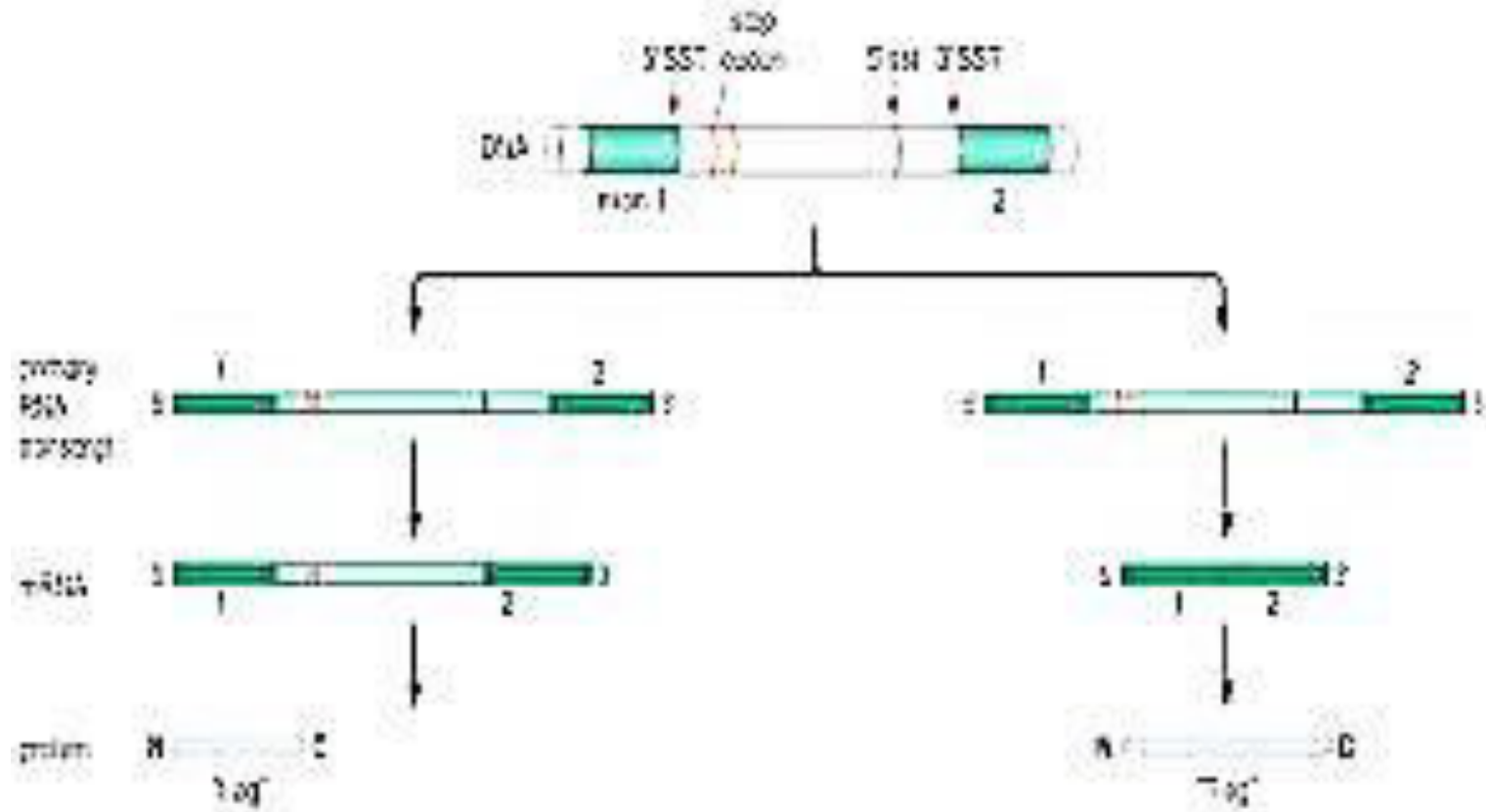
# Alternative Splicing

- When present at high levels, this protein directs the machinery to favor use of the 5' splice site that generates more of the t-ag mRNA.

# Alternative Splicing

- SF2/ASF is an SR protein and, when abundant, presumably binds sites within exon 2 and helps the spliceosome assemble there.

# Alternative Splicing





# Alternative Splicing

- In genome-wide studies, the most commonly seen forms of alternative splicing are cases in which complete exons are included or excluded from the mature message. Such exons are often called **cassette exons**.

# Alternative Splicing

- In ~10% of cases, cassette exons come in pairs, only one of which is included in the spliced message.
- In these cases, there must be mechanisms that ensure that the exons are spliced in a mutually exclusive fashion.

END

# Mutually Exclusive Splicing

- There are several mechanisms to ensure that selection of alternative exons is mutually exclusive — that is, that when one is chosen, the other is not.

# Mutually Exclusive Splicing

- ***Steric Hindrance***
- Consider two alternative exons separated by an intron. If the splice sites within the intron are too close together, splicing factors cannot bind to both sites at the same time.

# Mutually Exclusive Splicing

- Thus the binding of U1 snRNP to the 5' splice site of the intron between two alternative exons (exons 2 and 3) prevents the binding of U2 snRNP to the branchpoint within that same intron.

# Mutually Exclusive Splicing

- Alternatively, binding of U2 snRNP to the branchpoint excludes use of the 5' splice site. The splicing of exons 3 and 4 of  $\alpha$ -troponin is made mutually exclusive by this mechanism.

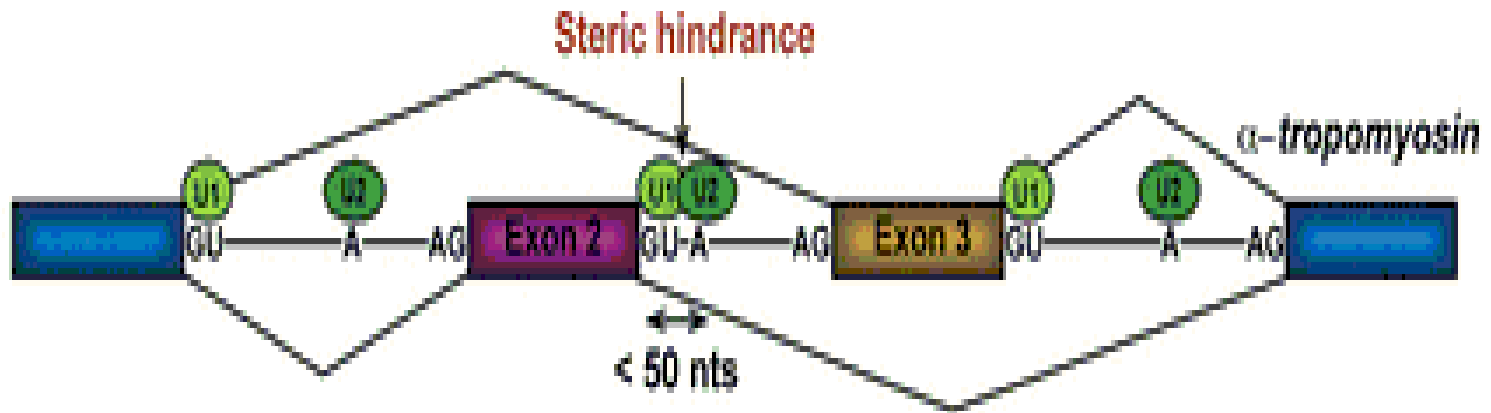


# Mutually Exclusive Splicing

- This arrangement can arise through the relative positions of the splice sites within an intron or because the intron is simply too small to work; in *Drosophila*, any intron under 59 nucleotides falls into that category.



# Mutually Exclusive Splicing



# Mutually Exclusive Splicing

- *Combinations of Major and Minor Splice Sites*
- The minor spliceosome recognizes splice sites distinct from those recognized by the major spliceosome.

# Mutually Exclusive Splicing

- Neither spliceosome can remove an intron that contains a combination of sites (i.e., a 5' splice site of one type and a 3' of the other).

# Mutually Exclusive Splicing

- Thus, by judicious arrangement of 5' and 3' splice sites recognized by these alternative spliceosomes, mutual exclusion can be achieved.
- The human *JNK1* gene is an example of this.

# Mutually Exclusive Splicing

- *Nonsense-Mediated Decay*
- Rather than forcing the splicing machinery to splice in a mutually exclusive fashion, this mechanism instead ensures that only messages that have one or another exon survive.

# Mutually Exclusive Splicing

- In other words, although not ensuring mutually exclusive splicing, the consequences of this mechanism amount to the same thing.

# Mutually Exclusive Splicing

- If the stop codon lies downstream (3') of all the exon-exon boundaries, the mRNA is spared from nonsense-mediated decay.



# Mutually Exclusive Splicing

- If, on the other hand, a stop codon is located upstream (5') to an exon-exon boundary, the mRNA is degraded.

# Mutually Exclusive Splicing

- Translating ribosomes, in conjunction with other surveillance proteins, assess this relationship for each individual mRNA.

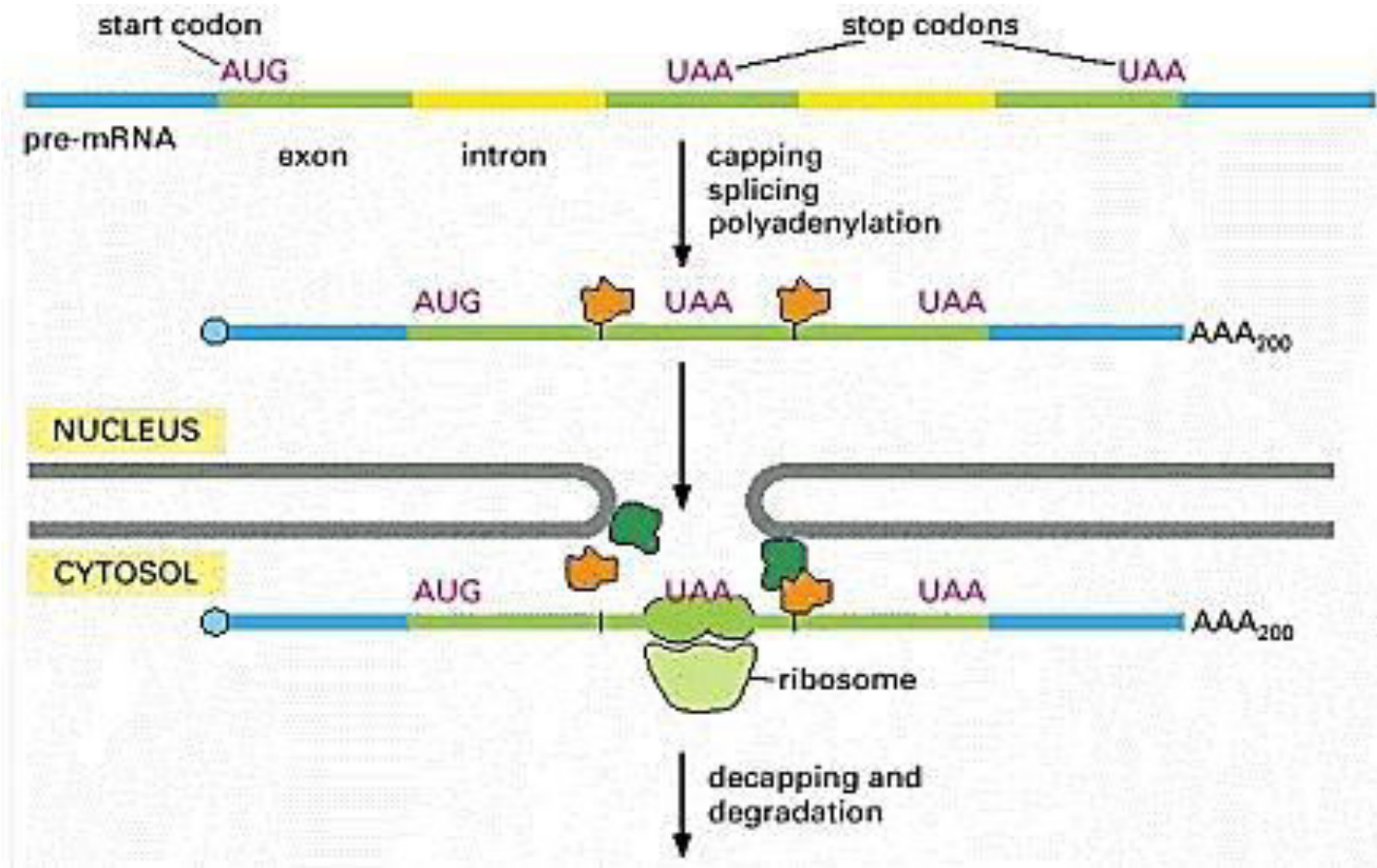
# Mutually Exclusive Splicing

- Exactly how this is accomplished is not understood in detail, but it is easy to understand why ribosomes must play a part.

# Mutually Exclusive Splicing

- Only in-frame termination codons trigger nonsense-mediated decay, and it is the relationship between the ribosome and the mRNA that defines the reading frame.

# Mutually Exclusive Splicing



# Regulation by Activators and Repressors

- Proteins that regulate splicing bind to specific sites called **exonic (or intronic) splicing enhancers** (ESE or ISE) or **silencers** (ESS or ISS).



# Regulation by Activators and Repressors

- The former enhance, and the latter repress, splicing at nearby splice sites.
- You have already encountered enhancers and the SR proteins that bind to them.



# Regulation by Activators and Repressors

- Indeed, these elements and proteins are important in directing the splicing machinery to many exons, even when alternative splicing is not involved.

# Regulation by Activators and Repressors

- In addition, in the example of T-antigen splicing described before, it was an SR protein that ensured that alternative splicing occurred.

# Regulation by Activators and Repressors

- But this protein family - which is large and diverse – has specific roles in regulated alternative splicing as well, directing the splicing machinery to different splice sites under different conditions.

# Regulation by Activators and Repressors

- Thus, the presence or activity of a given SR protein can determine whether a particular splice site is used in a particular cell type or at a particular stage of the development.

# Regulation by Activators and Repressors

- The SR proteins bind RNA using one domain—for example, the well characterized RNA-recognition motif (RRM).
- Each SR protein has another domain, rich in arginine and serine, called an **RS domain**.

# Regulation by Activators and Repressors

- The RS domain, found at the carboxy-terminal end of the protein, mediates interactions between the SR protein and proteins within the splicing machinery, recruiting that machinery to a nearby splice site.



# Regulation by Activators and Repressors

- This activator regulates the alternative splicing of a set of pre-mRNAs in the fly ovary.
- It works by binding to sites near the 3' splice site of specific exons in those pre-mRNAs and recruiting the U2AF splicing factor.



# Regulation by Activators and Repressors

- Most silencers are recognized by members of the heterogeneous nuclear ribonucleoprotein (hnRNP) family.
- These bind RNA but lack the RS domains and thus cannot recruit the splicing machinery.

# Regulation by Activators and Repressors

- Instead, by blocking specific splice sites, they repress the use of those sites.
- In case of *Dscam*, for example, Hrp36 inhibits inclusion of exon 6 variants in the mRNA.

# Regulation by Activators and Repressors

- Another example is hnRNPA1, which binds to an exonic silencer element within an exon of the HIV, *tat* pre- RNA and represses the inclusion of that exon in the final mRNA.

# Regulation by Activators and Repressors

- By binding to its site, the repressor blocks binding of the activator SC35 (an SR protein) to a nearby enhancer element.

# Regulation by Activators and Repressors

- In this case, blocking is not direct - the two binding sites do not overlap - but hnRNPA1 promotes cooperative binding of additional molecules of hnRNPA1 to adjacent sequences, spreading over the enhancer site.

# Regulation by Activators and Repressors

- When present, another SR protein (SF2/ASF) can overcome this repression because it has a higher affinity for the enhancer sequence than does SC35 and therefore displaces the repressors bound there.



# Regulation by Activators and Repressors

- Another mammalian splicing repressor is the hnRNPI protein.
- In some cases this protein blocks the binding of the basic splicing machinery by binding directly to the Py tract (hnRNPI is also called the polypyrimidine tract-binding protein).



# Regulation by Activators and Repressors

- In other cases it excludes a given exon from the mature mRNA by binding to sequences that flank that exon.

# Regulation by Activators and Repressors

- This exclusion occurs either because molecules of hnRNPI at each end of the exon interact and loop out the exon, which is then passed over by the spliceosome.

# Regulation by Activators and Repressors

- or because the molecules of hnRNPI at each end bind cooperatively with other molecules of hnRNPI, coating the RNA across the whole exon. This too would render the exon invisible to the splicing machinery.

# Regulation by Activators and Repressors

- We have emphasized alternative splicing as a way in which multiple protein products can be produced from a single gene. These different proteins are called **isoforms**.

# Regulation by Activators and Repressors

- They can have similar functions, distinct functions, or even antagonistic functions.
- But even some genes that encode only a single functional protein show alternative splicing.

# Regulation by Activators and Repressors

- In those cases, alternative splicing is used simply as a way of switching expression of the gene on and off.
- This is achieved in two ways.

# Regulation by Activators and Repressors

- Most straight forwardly, an exon contains a stop codon, and, when incorporated into mRNA, this prematurely terminates translation generating a truncated polypeptide.



# Regulation by Activators and Repressors

- Typically, such an incomplete protein is non-functional and rapidly degraded.
- Alternative splicing determines whether or not the exon with the stop codon is included in a given mRNA.

# Regulation by Activators and Repressors

- The second way alternative splicing can be used as an on/off switch is by regulating the use of an intron, which, when retained in the mRNA, ensures that species is not transported out of the nucleus and so is never translated.

END

# Regulation by Activators and Repressors

- Proteins that regulate splicing bind to specific sites called **exonic (or intronic) splicing enhancers** (ESE or ISE) or **silencers** (ESS or ISS).

# Regulation by Activators and Repressors

- The former enhance, and the latter repress, splicing at nearby splice sites.
- You have already encountered enhancers and the SR proteins that bind to them.

# Regulation by Activators and Repressors

- Indeed, these elements and proteins are important in directing the splicing machinery to many exons, even when alternative splicing is not involved.

# Regulation by Activators and Repressors

- In addition, in the example of T-antigen splicing described before, it was an SR protein that ensured that alternative splicing occurred.



# Regulation by Activators and Repressors

- But this protein family - which is large and diverse – has specific roles in regulated alternative splicing as well, directing the splicing machinery to different splice sites under different conditions.



# Regulation by Activators and Repressors

- Thus, the presence or activity of a given SR protein can determine whether a particular splice site is used in a particular cell type or at a particular stage of the development.

# Regulation by Activators and Repressors

- The SR proteins bind RNA using one domain—for example, the well characterized RNA-recognition motif (RRM).
- Each SR protein has another domain, rich in arginine and serine, called an **RS domain**.

# Regulation by Activators and Repressors

- The RS domain, found at the carboxy-terminal end of the protein, mediates interactions between the SR protein and proteins within the splicing machinery, recruiting that machinery to a nearby splice site.

# Regulation by Activators and Repressors

- This activator regulates the alternative splicing of a set of pre-mRNAs in the fly ovary.
- It works by binding to sites near the 3' splice site of specific exons in those pre-mRNAs and recruiting the U2AF splicing factor.

# Regulation by Activators and Repressors

- Most silencers are recognized by members of the heterogeneous nuclear ribonucleoprotein (hnRNP) family.
- These bind RNA but lack the RS domains and thus cannot recruit the splicing machinery.

# Regulation by Activators and Repressors

- Instead, by blocking specific splice sites, they repress the use of those sites.
- In case of *Dscam*, for example, Hrp36 inhibits inclusion of exon 6 variants in the mRNA.



# Regulation by Activators and Repressors

- Another example is hnRNPA1, which binds to an exonic silencer element within an exon of the HIV, *tat* pre- RNA and represses the inclusion of that exon in the final mRNA.



# Regulation by Activators and Repressors

- By binding to its site, the repressor blocks binding of the activator SC35 (an SR protein) to a nearby enhancer element.

# Regulation by Activators and Repressors

- In this case, blocking is not direct - the two binding sites do not overlap - but hnRNPA1 promotes cooperative binding of additional molecules of hnRNPA1 to adjacent sequences, spreading over the enhancer site.

# Regulation by Activators and Repressors

- When present, another SR protein (SF2/ASF) can overcome this repression because it has a higher affinity for the enhancer sequence than does SC35 and therefore displaces the repressors bound there.

# Regulation by Activators and Repressors

- Another mammalian splicing repressor is the hnRNPI protein.
- In some cases this protein blocks the binding of the basic splicing machinery by binding directly to the Py tract (hnRNPI is also called the polypyrimidine tract-binding protein).

# Regulation by Activators and Repressors

- In other cases it excludes a given exon from the mature mRNA by binding to sequences that flank that exon.

# Regulation by Activators and Repressors

- This exclusion occurs either because molecules of hnRNPI at each end of the exon interact and loop out the exon, which is then passed over by the spliceosome.



# Regulation by Activators and Repressors

- or because the molecules of hnRNPI at each end bind cooperatively with other molecules of hnRNPI, coating the RNA across the whole exon. This too would render the exon invisible to the splicing machinery.



# Regulation by Activators and Repressors

- We have emphasized alternative splicing as a way in which multiple protein products can be produced from a single gene. These different proteins are called **isoforms**.

# Regulation by Activators and Repressors

- They can have similar functions, distinct functions, or even antagonistic functions.
- But even some genes that encode only a single functional protein show alternative splicing.

# Regulation by Activators and Repressors

- In those cases, alternative splicing is used simply as a way of switching expression of the gene on and off.
- This is achieved in two ways.

# Regulation by Activators and Repressors

- Most straight forwardly, an exon contains a stop codon, and, when incorporated into mRNA, this prematurely terminates translation generating a truncated polypeptide.

# Regulation by Activators and Repressors

- Typically, such an incomplete protein is non-functional and rapidly degraded.
- Alternative splicing determines whether or not the exon with the stop codon is included in a given mRNA.

# Regulation by Activators and Repressors

- The second way alternative splicing can be used as an on/off switch is by regulating the use of an intron, which, when retained in the mRNA, ensures that species is not transported out of the nucleus and so is never translated.

END



# Mechanism of RNA Editing

- One important clue about the mechanism of RNA editing is that partially edited transcripts have been isolated, and these are always edited at their 3'-ends but not at their 5'-ends.



# Mechanism of RNA Editing

- This suggests strongly that editing proceeds in a 3'→5' direction.
- Kenneth Stuart and colleagues first reported this phenomenon in 1988.

# Mechanism of RNA Editing

- Their experimental tool was RT-PCR, starting with reverse transcriptase to make the first DNA strand from an RNA template, followed by standard PCR.

# Mechanism of RNA Editing

- In one experiment, Stuart and coworkers used pairs of PCR primers in which both were edited primers, both unedited primers, or one of each.

# Mechanism of RNA Editing

- A completely edited RNA will hybridize only to edited primers and give a PCR signal, whereas it will not hybridize to unedited primers, so any PCR protocol including at least one unedited primer will not give a signal from this RNA.

# Mechanism of RNA Editing

- By contrast, a completely unedited RNA will react only with unedited primers.

# Mechanism of RNA Editing

- But the real test is to use an unedited 5'-primer and an edited 3'-primer to detect 3'-edited transcripts, or an edited 5'-primer and an unedited 3'-primer to detect 5'-edited transcripts.

# Mechanism of RNA Editing

- If editing goes from 3' to 5' in the transcript, then 3'-edited transcripts, but not 5'-edited transcripts, should be detected.
- The signals were present only when both primers were edited, or the 3'-primer was edited.



# Mechanism of RNA Editing

- No signal was observed when only the 5'-primer was edited. Thus, 3'-editing occurred in the absence of 5'-editing, but 5'-editing did not occur without 3'-editing.
- This is consistent with editing in the 3'→5' direction.

# Mechanism of RNA Editing

- This experiment is valuable, but it has a flaw: None of the lanes involving the unedited 3'-primer shows a signal.
- What determines where the editing system should add or delete UMPs?

# Mechanism of RNA Editing

- Larry Simpson and colleagues found the answer in 1990 when they discovered **guide RNAs (gRNAs)** encoded in *Leishmania* maxicircles.

# Mechanism of RNA Editing

- They found in maxicircle DNA, seven short sequences that could produce short RNAs (gRNAs) complementary to parts of five different edited mitochondrial mRNAs.

# Mechanism of RNA Editing

- In principle, such gRNAs could direct the insertion and deletion of UMPs over a stretch of several dozen nucleotides in the mRNA.

# Mechanism of RNA Editing

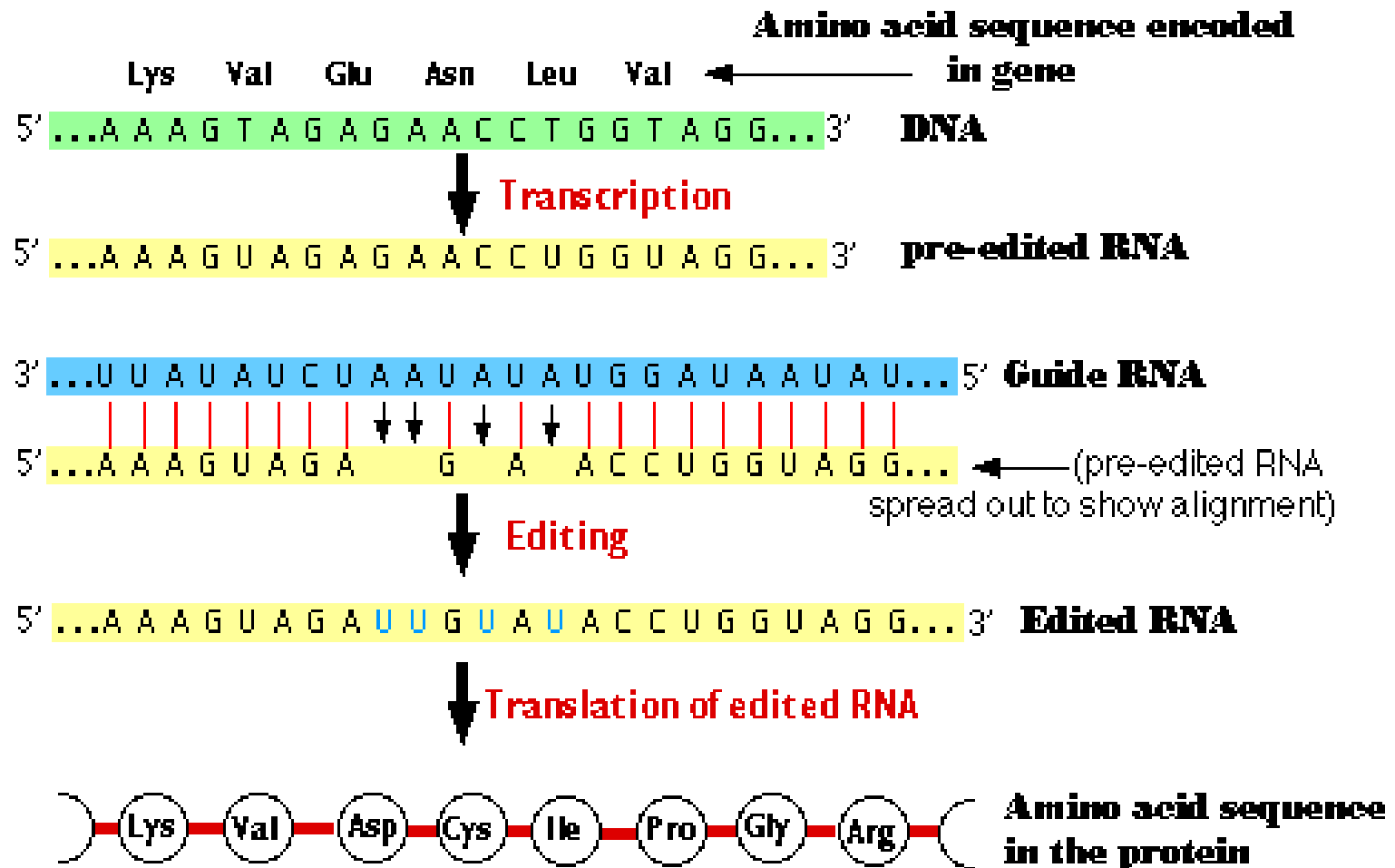
- Once that editing is done, another gRNA could hybridize near the 5'-end of the newly edited region and direct editing of a new segment.

# Mechanism of RNA Editing

- Working in this way from the 3'-end of the mRNA toward the 5'-end, successive gRNAs bind to regions edited by their predecessor gRNAs and direct further editing until they have finished the whole editing job.



# Mechanism of RNA Editing



# Mechanism of RNA Editing

- Later in 1990, Nancy Sturm and Larry Simpson found that minicircles also encode gRNAs.
- But besides the coding potential, they found direct evidence for the existence of gRNAs.

# Mechanism of RNA Editing

- The precise mechanism of editing, the cutting and pasting required to insert and delete UMPs, remained unclear for several years, but the enzyme activities found in kinetoplasts provided some hints.

# Mechanism of RNA Editing

- For example, kinetoplasts have a terminal **uridylyl transferase (TUTase)** that could add extra UMPs (uridylates) to the mRNA during editing.

# Mechanism of RNA Editing

- Because the mRNA has to be cut to accept these new UMPs, it must also be ligated together again, and kinetoplasts also contain an RNA ligase.

# Mechanism of RNA Editing

- Uridylates at the ends of gRNAs could be transferred to the pre-mRNA by transesterification. That is, the uridylates could be plucked off of the ends of gRNAs and transferred directly to the pre-mRNA.

# Mechanism of RNA Editing

- Then, in 1994, Scott Seiwert and Stuart used a mitochondrial extract and a gRNA to edit a synthetic pre-mRNA.
- They found that deletion of UMPs required three enzymatic activities.



# Mechanism of RNA Editing

- 1. An endonuclease that follows directions from the gRNA and cuts the pre-mRNA at the site where a UMP needs to be removed.
- 2. a 3'-exonuclease that is specific for terminal uridines; and
- 3. an RNA ligase.

# Mechanism of RNA Editing

- In 1996, using a similar in vitro system, Stuart and colleagues demonstrated that UMP insertion follows a similar three-step pathway.

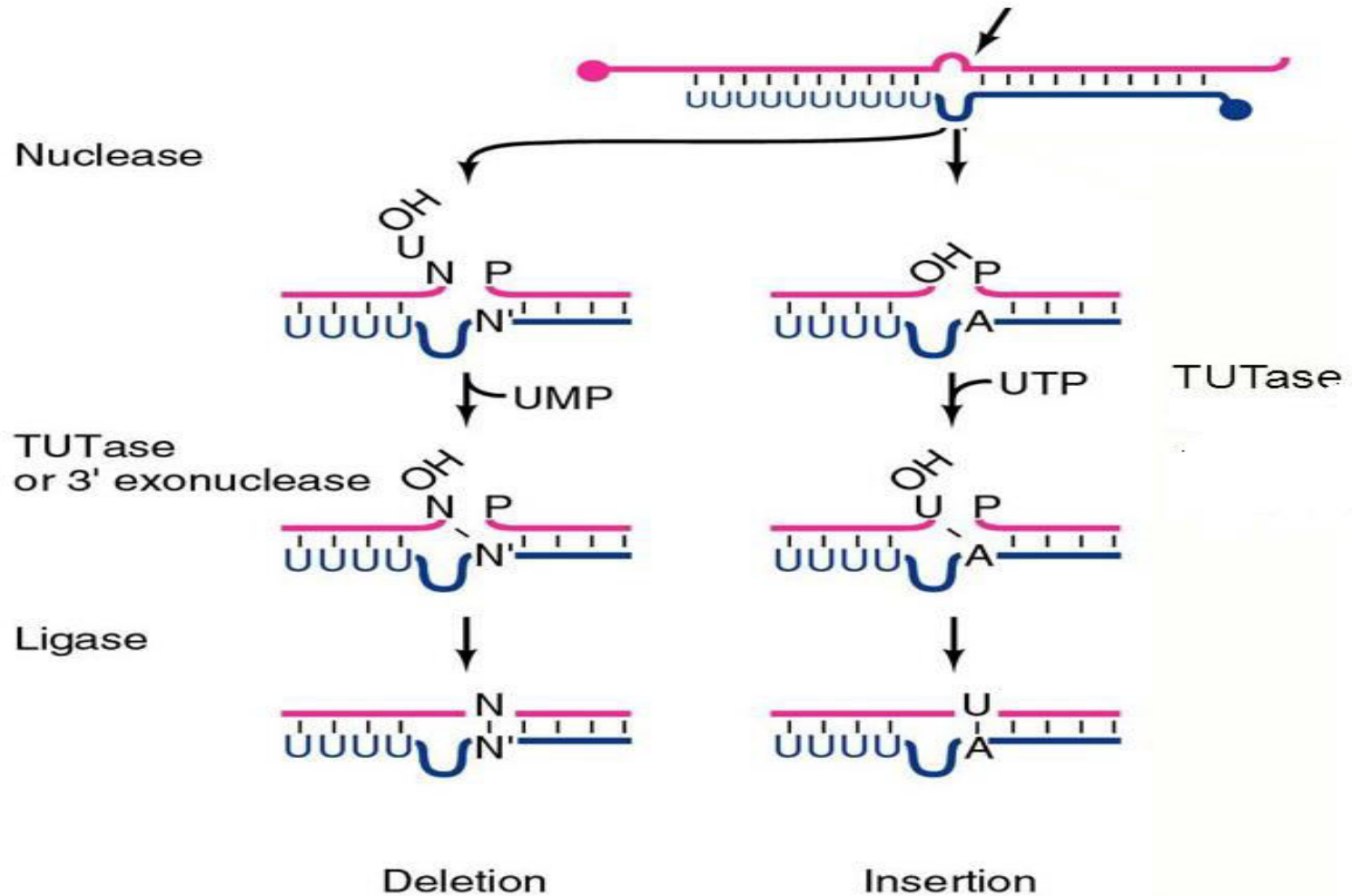
# Mechanism of RNA Editing

- 1. a gRNA-directed endonuclease cuts at the site where UMP insertion is required.
- 2. an enzyme (probably TUTase) transfers UMPs from UTP (not from gRNA), as directed by the gRNA.

# Mechanism of RNA Editing

- 3. an RNA ligase puts the two pieces of RNA back together.

# Mechanism of RNA Editing



# Mechanism of RNA Editing

- One important clue about the mechanism of RNA editing is that partially edited transcripts have been isolated, and these are always edited at their 3'-ends but not at their 5'-ends.

# Mechanism of RNA Editing

- This suggests strongly that editing proceeds in a 3'→5' direction.
- Kenneth Stuart and colleagues first reported this phenomenon in 1988.



# Mechanism of RNA Editing

- Their experimental tool was RT-PCR, starting with reverse transcriptase to make the first DNA strand from an RNA template, followed by standard PCR.

# Mechanism of RNA Editing

- In one experiment, Stuart and coworkers used pairs of PCR primers in which both were edited primers, both unedited primers, or one of each.

# Mechanism of RNA Editing

- A completely edited RNA will hybridize only to edited primers and give a PCR signal, whereas it will not hybridize to unedited primers, so any PCR protocol including at least one unedited primer will not give a signal from this RNA.

# Mechanism of RNA Editing

- By contrast, a completely unedited RNA will react only with unedited primers.

# Mechanism of RNA Editing

- But the real test is to use an unedited 5'-primer and an edited 3'-primer to detect 3'-edited transcripts, or an edited 5'-primer and an unedited 3'-primer to detect 5'-edited transcripts.

# Mechanism of RNA Editing

- If editing goes from 3' to 5' in the transcript, then 3'-edited transcripts, but not 5'-edited transcripts, should be detected.
- The signals were present only when both primers were edited, or the 3'-primer was edited.

# Mechanism of RNA Editing

- No signal was observed when only the 5'-primer was edited. Thus, 3'-editing occurred in the absence of 5'-editing, but 5'-editing did not occur without 3'-editing.
- This is consistent with editing in the 3'→5' direction.



# Mechanism of RNA Editing

- This experiment is valuable, but it has a flaw: None of the lanes involving the unedited 3'-primer shows a signal.
- What determines where the editing system should add or delete UMPs?

# Mechanism of RNA Editing

- Larry Simpson and colleagues found the answer in 1990 when they discovered **guide RNAs (gRNAs)** encoded in *Leishmania* maxicircles.

# Mechanism of RNA Editing

- They found in maxicircle DNA, seven short sequences that could produce short RNAs (gRNAs) complementary to parts of five different edited mitochondrial mRNAs.

# Mechanism of RNA Editing

- In principle, such gRNAs could direct the insertion and deletion of UMPs over a stretch of several dozen nucleotides in the mRNA.

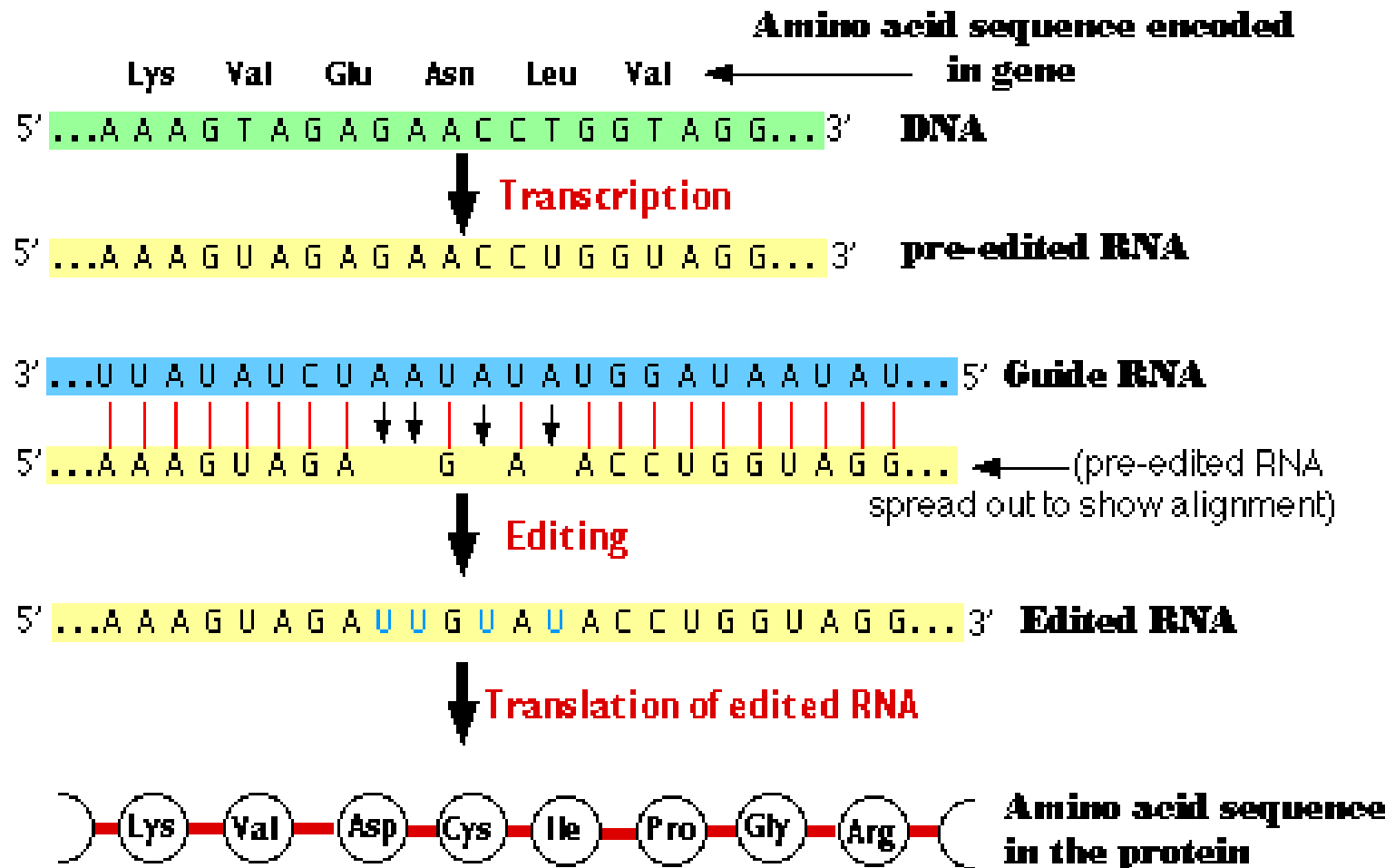
# Mechanism of RNA Editing

- Once that editing is done, another gRNA could hybridize near the 5'-end of the newly edited region and direct editing of a new segment.

# Mechanism of RNA Editing

- Working in this way from the 3'-end of the mRNA toward the 5'-end, successive gRNAs bind to regions edited by their predecessor gRNAs and direct further editing until they have finished the whole editing job.

# Mechanism of RNA Editing





# Mechanism of RNA Editing

- Later in 1990, Nancy Sturm and Larry Simpson found that minicircles also encode gRNAs.
- But besides the coding potential, they found direct evidence for the existence of gRNAs.

# Mechanism of RNA Editing

- The precise mechanism of editing, the cutting and pasting required to insert and delete UMPs, remained unclear for several years, but the enzyme activities found in kinetoplasts provided some hints.

# Mechanism of RNA Editing

- For example, kinetoplasts have a terminal **uridylyl transferase (TUTase)** that could add extra UMPs (uridylates) to the mRNA during editing.

# Mechanism of RNA Editing

- Because the mRNA has to be cut to accept these new UMPs, it must also be ligated together again, and kinetoplasts also contain an RNA ligase.

# Mechanism of RNA Editing

- Uridylates at the ends of gRNAs could be transferred to the pre-mRNA by transesterification. That is, the uridylates could be plucked off of the ends of gRNAs and transferred directly to the pre-mRNA.

# Mechanism of RNA Editing

- Then, in 1994, Scott Seiwert and Stuart used a mitochondrial extract and a gRNA to edit a synthetic pre-mRNA.
- They found that deletion of UMPs required three enzymatic activities.

# Mechanism of RNA Editing

- 1. An endonuclease that follows directions from the gRNA and cuts the pre-mRNA at the site where a UMP needs to be removed.
- 2. a 3'-exonuclease that is specific for terminal uridines; and
- 3. an RNA ligase.



# Mechanism of RNA Editing

- In 1996, using a similar in vitro system, Stuart and colleagues demonstrated that UMP insertion follows a similar three-step pathway.

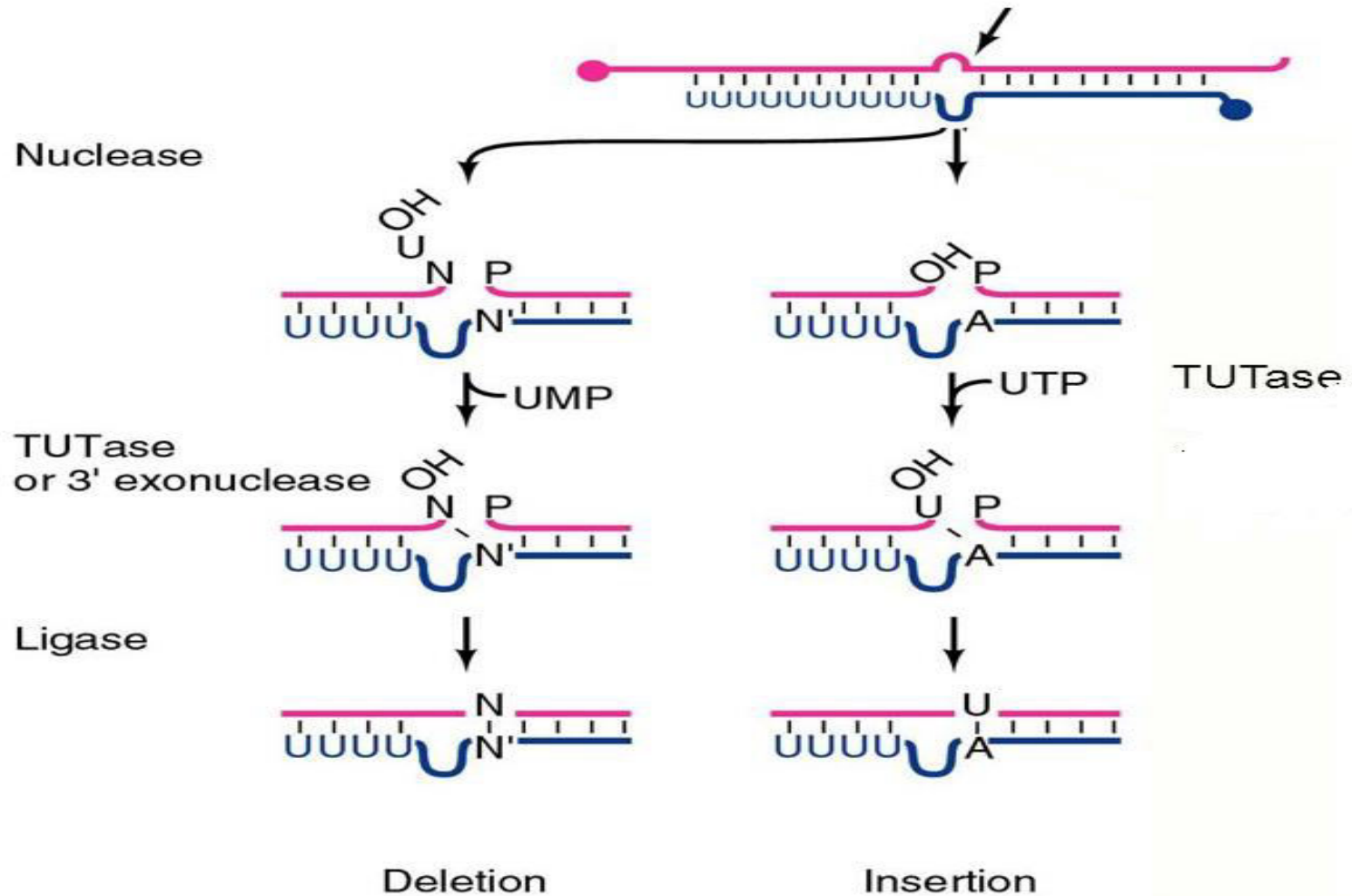
# Mechanism of RNA Editing

- 1. a gRNA-directed endonuclease cuts at the site where UMP insertion is required.
- 2. an enzyme (probably TUTase) transfers UMPs from UTP (not from gRNA), as directed by the gRNA.

# Mechanism of RNA Editing

- 3. an RNA ligase puts the two pieces of RNA back together.

# Mechanism of RNA Editing



# RNA Editing by Nucleotide Deamination

- RNA editing also plays a vital role in higher organisms even mammals.

# RNA Editing by Nucleotide Deamination

- Another kind of RNA editing takes place through the deamination of adenosine, which converts adenosine to inosine, which has an oxygen in place of adenine's amino group.

# RNA Editing by Nucleotide Deamination

- Because inosine forms base pairs with cytidine in the same way as guanosine, the deamination of adenosine changes the meaning of a codon.



# RNA Editing by Nucleotide Deamination

- For example, an ACG (threonine) codon becomes an ICG codon, which would be read by the ribosome as GCG (alanine).

# RNA Editing by Nucleotide Deamination

- This kind of RNA editing is directed by an enzyme called adenosine deaminase acting on RNA (ADAR).
- Humans and mice contain three ADAR genes: ADAR1, ADAR2, and ADAR3.

# RNA Editing by Nucleotide Deamination

- The products of the first two are ubiquitous in the body, but the third gene product is found only in the brain.

# RNA Editing by Nucleotide Deamination

- These enzymes are very specific. It would be disastrous if they deaminated every adenosine in an mRNA, so they select only certain adenosines in certain mRNAs.

# RNA Editing by Nucleotide Deamination

- For example, ADAR2 deaminates one adenosine in the glutamate-sensitive ion-channel receptor subunit B (GluR-B) mRNA, with greater than 99% efficiency.
- This alteration in the mRNA changes a glutamine codon to an arginine codon.

# RNA Editing by Nucleotide Deamination

- We know that an ion channel containing the GluR-B protein with a glutamine instead of an arginine is too permeable to calcium ions.
- Therefore, mice with a defective ADAR2 gene would have serious problems.

# RNA Editing by Nucleotide Deamination

- Peter Seeburg and colleagues demonstrated that the only critical target of ADAR2 is the GluR-B transcript.



# RNA Editing by Nucleotide Deamination

- The *Drosophila* genome contains only one ADAR gene.
- When this gene is mutated so the flies lack all ADAR activity, they do not carry out any mRNA editing at known editing sites.

# RNA Editing by Nucleotide Deamination

- These mutant flies are viable, but they have difficulty in walking, cannot fly, and suffer progressive neural degeneration, particularly in the brain.

# RNA Editing by Nucleotide Deamination

- Thus, the phenotype of this mutation is similar to the phenotype of mutations in the gene for ADAR2 in mammals.

# RNA Editing by Nucleotide Deamination

- ADAR1 also appears to be essential for mammalian life.
- Kazuko Nishikura and coworkers mutated mouse stem cells to heterozygous mutant ( $\text{ADAR}^{+/-}$ ).

# RNA Editing by Nucleotide Deamination

- They found that no embryo containing this mutated gene survived to birth.
- Thus, even heterozygous mutations in ADAR1 appear to be embryonic lethal.

# RNA Editing by Nucleotide Deamination

- Interestingly, certain tumors lose ADAR activity. In particular, a very malignant human brain tumor called glioblastoma multiforme (GBM) has very low ADAR2 activity, and a corresponding underediting in the GluR-B mRNA.

# RNA Editing by Nucleotide Deamination

- Another kind of editing is carried out by cytidine deaminase acting on RNA (CDAR), which converts cytidine to uridine.



# RNA Editing by Nucleotide Deamination

- This C→U editing is defective in about 25% of the benign peripheral nerve sheath tumors found in neurofibromatosis type I patients.
- C→U editing also appears to occur in HIV transcripts in human cells.

# RNA Editing by Nucleotide Deamination

- Still another kind of editing that occurs in HIV-infected human cells is G→A editing.
- But this kind of editing cannot be explained by a single-step deamination, and it is unclear how it is accomplished.

END

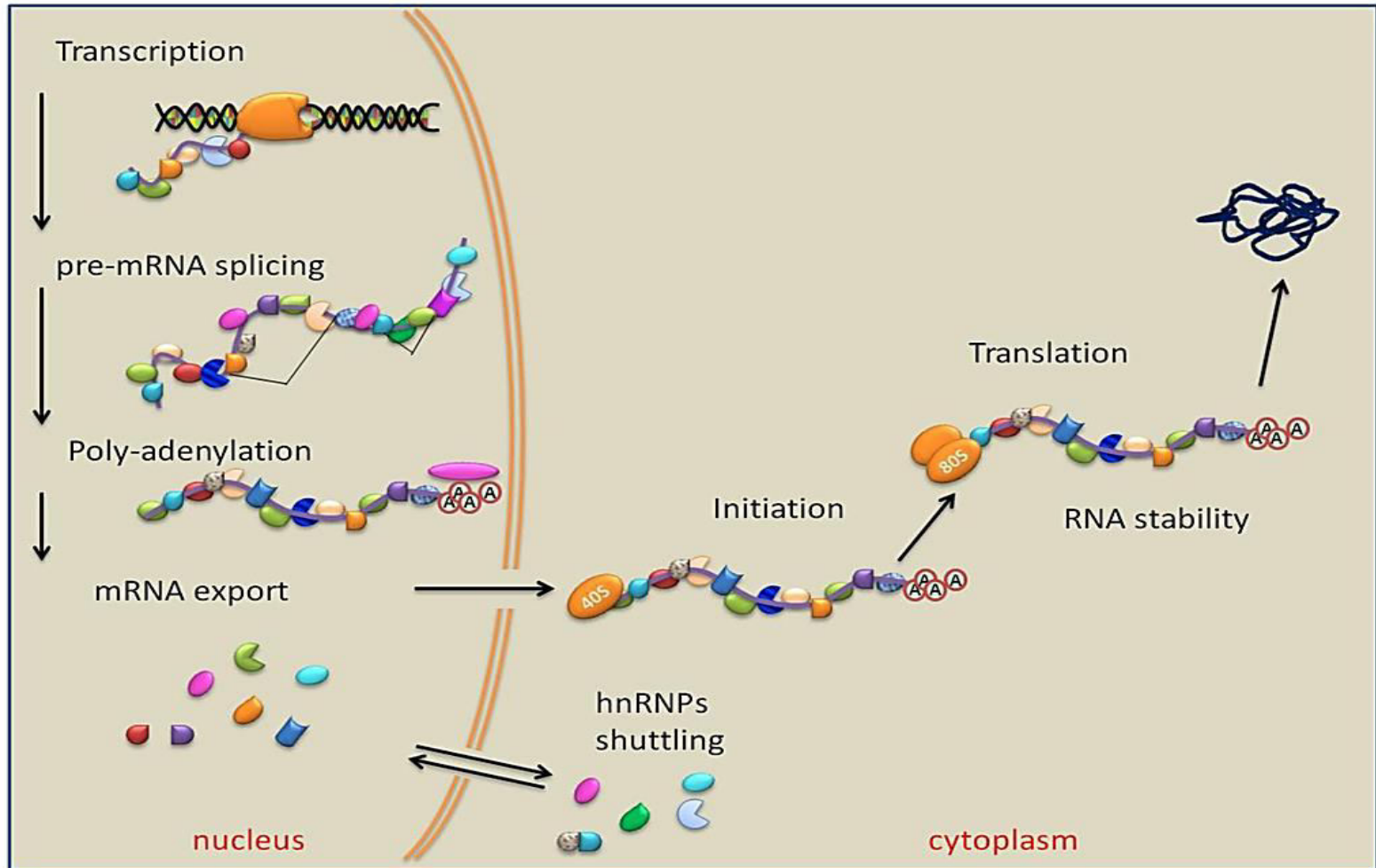
# mRNA Transport

- Once fully processed-capped, intron-free, and polyadenylated mRNA is transported out of the nucleus and into the cytoplasm where it is translated to give its protein product.

# mRNA Transport

- Movement from the nucleus to the cytoplasm is not a passive process.
- Indeed, it must be carefully regulated: the fully processed mRNAs represent only a small proportion of the RNA found in the nucleus.

# mRNA Transport



# mRNA Transport

- Many of the other RNAs would be detrimental to the cell if exported.
- These include, for example, damaged or misprocessed RNAs, and liberated introns.



# mRNA Transport

- How are RNA selection and transport achieved?
- From the moment an RNA molecule starts to be transcribed, it becomes associated with proteins of various sorts.



# mRNA Transport

- Initially proteins involved in capping, then splicing factors, and finally the proteins that mediate polyadenylation.

# mRNA Transport

- Some of these proteins are replaced at various steps along the processing path, but others (including some SR proteins) are not; and, moreover, additional proteins join.

# mRNA Transport

- As a result, a typical mature mRNA carries a collection of proteins that identifies it as being mRNA destined for transport.

# mRNA Transport

- Other RNAs not only lack the particular signature collection required for transport, but have their own alternative set of proteins that actively blocks export.

# mRNA Transport

- Thus, for example, excised introns will often carry hnRNPs. and these probably mark such an RNA for nuclear retention and destruction.

# mRNA Transport

- Mature mRNAs carry residual SR proteins, and even another group of proteins that bind specifically to exon-exon junctions (which are only found in spliced species of course).

# mRNA Transport

- The mRNAs do also contain some hnRNPs, but fewer than are typically bound to introns, and in a different context as well.



# mRNA Transport

- This emphasizes the fact that it is the set of proteins, not any individual kind of protein, that marks RNAs for either export or retention in the nucleus.

# mRNA Transport

- Max Birnstiel and colleagues in 1991 demonstrated that transcripts of a bacterial neomycin gene transplanted into monkey COS1 cells remained in the nucleus.

# mRNA Transport

- They reasoned that the lack of a polyadenylation signal in the bacterial gene would have left the transcripts without a mature 3'-end, and that might be the reason for defective transport to the cytoplasm.

# mRNA Transport

- To test this hypothesis, they provided the neomycin gene with the strong polyadenylation signal from a mammalian  $\beta$ -globin gene.

# mRNA Transport

- This allowed for polyadenylation of the neomycin transcripts, which were then efficiently transported out of the nucleus into the cytoplasm.

# mRNA Transport

- This allowed for polyadenylation of the neomycin transcripts, which were then efficiently transported out of the nucleus into the cytoplasm.

# mRNA Transport

- Export takes place through a special structure in the nuclear membrane called the nuclear pore complex.



# mRNA Transport

- Small molecules - those under about 50 Kd -can pass through these pores unaided; but larger molecules and complexes, including mRNAs and their associated proteins, require active transport.

# mRNA Transport

- Some of the proteins associated with the RNA carry nuclear export signals that are recognized by export receptors that guide the RNA out through the pore.

# mRNA Transport

- Once in the cytoplasm, the proteins are discarded, and are then recognized for import back into the nucleus where they associate with another mRNA and repeat the cycle.

END

# Translation

- The genetic information contained within the order of nucleotides in messenger RNA (mRNA) is interpreted to generate the linear sequences of amino acids in proteins. This process is known as **translation**.

# Translation

- Translation is among the most highly conserved across all organisms and among the most energetically costly for the cell.

# Translation

- In rapidly growing bacterial cells, up to 80% of the cell's energy and 50% of the cell's dry weight are dedicated to protein synthesis.

# Translation

- Indeed, the synthesis of a single protein requires the coordinated action of well over 100 proteins and RNAs.



# Translation

- Unlike the complementarity between the DNA template and the ribonucleotides of the mRNA, the side chains of amino acids have little or no specific affinity for the purine and pyrimidine bases found in RNA.

# Translation

- For example, the hydrophobic side chains of the amino acids alanine, valine, leucine, and isoleucine cannot form hydrogen bonds with the amino and keto groups of the nucleotide bases.

# Translation

- Similarly, it is hard to imagine that how the bases of RNA can have unique affinities for the aromatic amino acids phenylalanine, tyrosine, and tryptophan.

# Translation

- So Francis H. Crick in 1955 proposed that before their incorporation into polypeptides, amino acids must attach to a special adaptor molecule that is capable of directly interacting with and recognizing the coding units of the mRNA.

# Translation

- Paul C. Zamecnik and Mahlon B. Hoagland (1957) showed that before their incorporation into proteins, amino acids are attached to a class of RNA molecules (representing 15% of all cellular RNA).

# Translation

- These RNAs are called **transfer RNAs (tRNAs)** because their attached amino acid is subsequently transferred to the growing polypeptide chain.

# Translation

- The machinery responsible for translating the language of mRNAs into the language of proteins is composed of four primary components:-



# Translation

- mRNAs, tRNAs, aminoacyl-tRNA synthetases and the ribosomes.

# Translation

- Together, these components accomplish the extraordinary task of translating a code written in a four-base alphabet into a second code written in the language of the 20 amino acids.

# Translation

- The mRNA provides the information that must be interpreted by the translation machinery and is the template for translation.

# Translation

- The protein-coding region of the mRNA consists of an ordered series of three nucleotide- long units called **codons** that specify the order of amino acids.

# Translation

- The tRNAs provide the physical interface between the amino acids being added to the growing polypeptide chain and the codons in the mRNA.

# Translation

- Enzymes called aminoacyl-tRNA synthetases couple amino acids to specific tRNAs that recognize the appropriate codon(s).

# Translation

- The final major player in translation is the ribosome, a remarkable, multimegadalton machine composed of both RNA and protein.



# Translation

- The ribosome coordinates the correct recognition of the mRNA by each tRNA and catalyzes peptide-bond formation between the growing polypeptide chain and the amino acid attached to the selected tRNA.

END

# Messenger RNA

- The protein-coding region(s) of each mRNA is composed of a contiguous, non-overlapping string of codons called an **open reading frame** (commonly known as an **ORF**).

# Messenger RNA

- Each ORF specifies a single protein and starts and ends at internal sites within the mRNA.
- That is, the ends of an ORF are distinct from the ends of the mRNA.

# Messenger RNA

- Translation starts at the 5' end of the ORF and proceeds one codon at a time to the 3' end. The first and last codons of an ORF are known as the **start** and **stop codons**.

# Messenger RNA

- In bacteria, the start codon is usually 5'-AUG-3', but 5'-GUG-3' and sometimes even 5'-UUG-3' are also used.
- Eukaryotic cells always use 5'-AUG-3' as the start codon.

# Messenger RNA

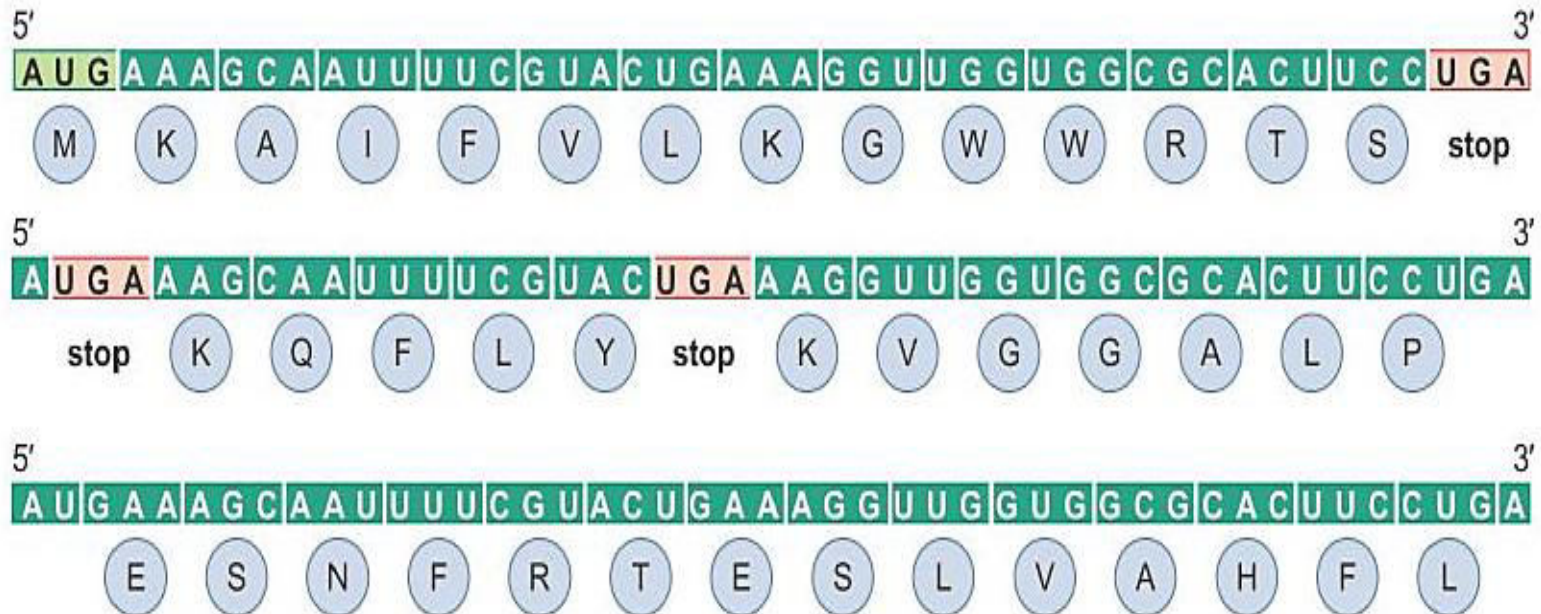
- The start codon has two important functions.
- First, it specifies the first amino acid to be incorporated into the growing polypeptide chain.
- Second, it defines the reading frame for all subsequent codons.

# Messenger RNA

- Because each codon is immediately adjacent to the next codon, and because codons are three nucleotides long, any stretch of mRNA could be translated in three different reading frames.



# Messenger RNA



Three possible reading frames of the *E. coli trp* leader sequence

# Messenger RNA

- Once translation starts, however, the reading frame is determined.
- Thus, by setting the location of the first codon, the start codon determines the location of all following codons.

# Messenger RNA

- Stop codons, of which there are three (5'-UAG-3', 5'-UGA-3', and 5'-UAA-3'), define the end of the ORF and signal termination of polypeptide synthesis.

# Messenger RNA

- You can now understand the origin of the term open reading frame. It is a contiguous stretch of codons “read” in a particular frame (as set by the first codon) that is “open” to translation because it lacks a stop codon.

# Messenger RNA

- mRNAs contain at least one ORF. The number of ORFs per mRNA is different between eukaryotes and prokaryotes.
- Eukaryotic mRNAs almost always contain a single ORF.

# Messenger RNA

- In contrast, prokaryotic mRNAs frequently contain two or more ORFs.
- mRNAs containing multiple ORFs are known as polycistronic RNAs and those encoding a single ORF are known as monocistronic RNAs.



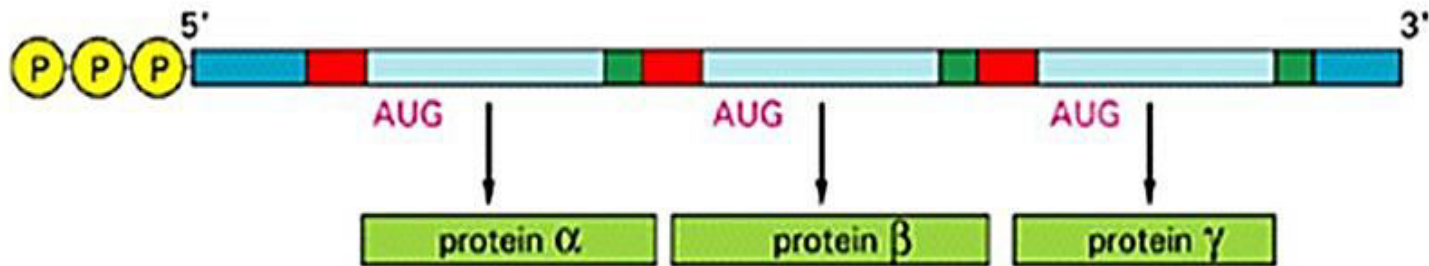
# Messenger RNA

- The polycistronic mRNAs found in bacteria often encode proteins that perform related functions, such as different steps in the biosynthesis of an amino acid or nucleotide.

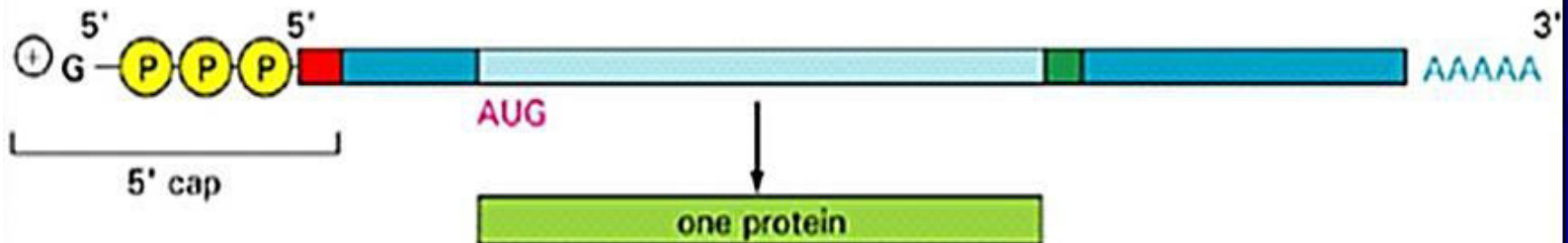


# Messenger RNA

procaryotic mRNA



eucaryotic mRNA



# Prokaryotic mRNAs

- For translation to occur, the ribosome must be recruited to the mRNA.
- Prokaryotic mRNAs have a ribosome-binding site that recruits the translational machinery.

# Prokaryotic mRNAs

- To facilitate binding by a ribosome, many prokaryotic ORFs contain a short sequence upstream (on the 5' side) of the start codon called the **ribosome-binding site (RBS)**.

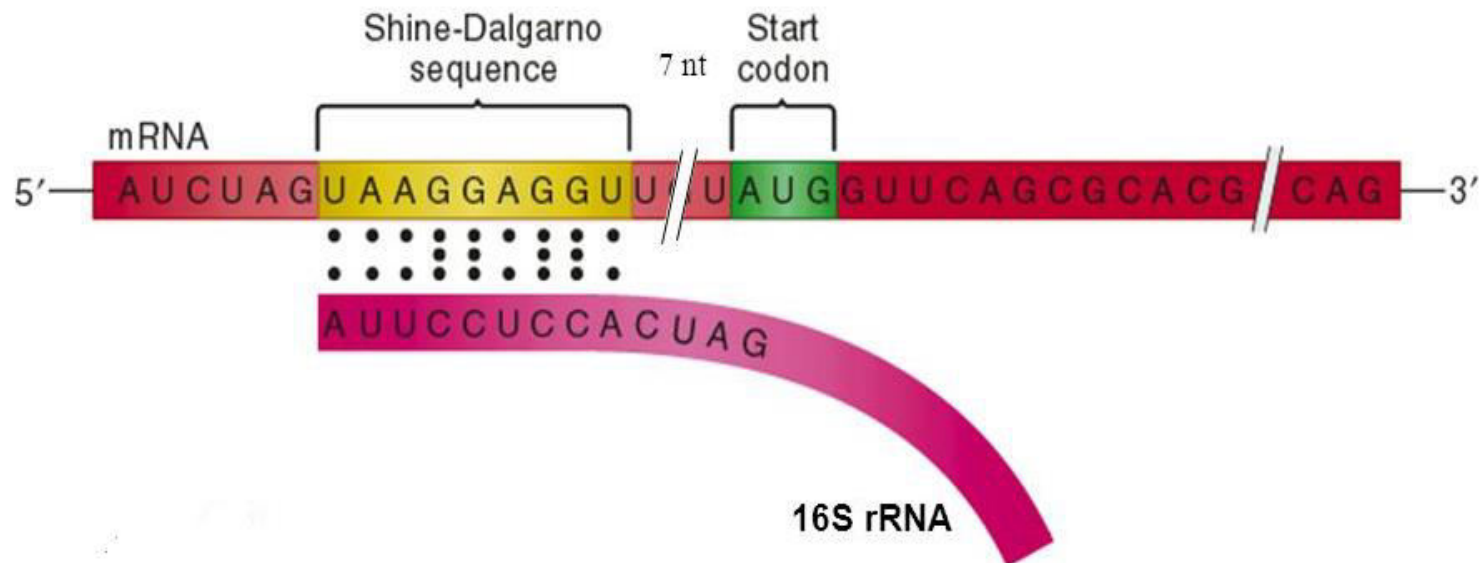
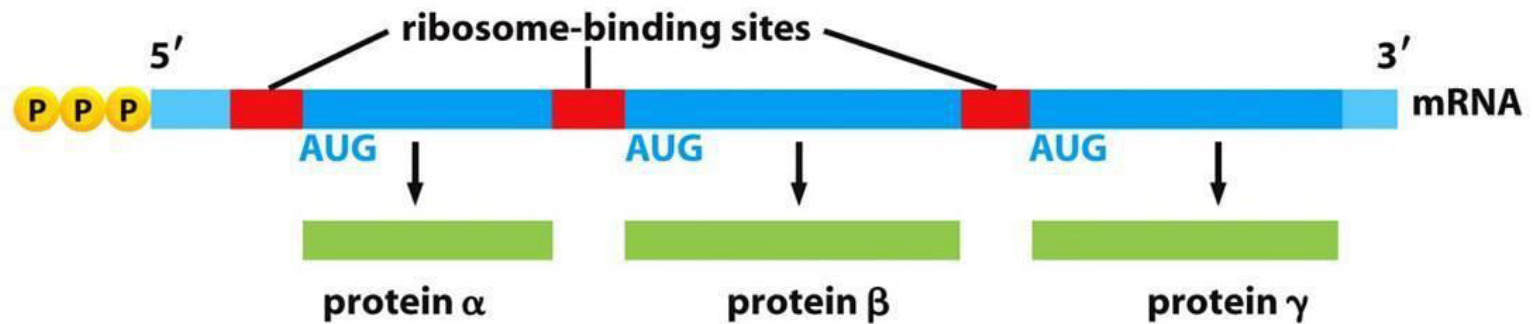
# Prokaryotic mRNAs

- This element is also referred to as a **Shine Öalgarno sequence** after the scientists who discovered it by comparing the sequences of multiple mRNAs.

# Prokaryotic mRNAs

- The RBS, typically located 3–9 bp on the 5' side of the start codon, is complementary to a sequence located near the 3' end of one of the ribosomal RNA components, the 16S ribosomal RNA (rRNA).

# Prokaryotic mRNAs



# Prokaryotic mRNAs

- The RBS base-pairs with this RNA, thereby aligning the ribosome with the beginning of the ORF. The core of this region of the 16S rRNA has the sequence 5'-CCUCCU-3'.



# Prokaryotic mRNAs

- Not surprisingly, prokaryotic RBS are most often a subset of the sequence 5'-AGGAGG-3'.

# Prokaryotic mRNAs

- The extent of complementarity and the spacing between the RBS and the start codon has a strong influence on how actively a particular ORF is translated.

# Prokaryotic mRNAs

- High complementarity and proper spacing promote active translation, whereas limited complementarity and/or poor spacing generally support lower levels of translation.

# Prokaryotic mRNAs

- Some prokaryotic ORFs lack a strong RBS but are nonetheless actively translated.
- These ORFs are not the first ORF in an mRNA but instead are located just after another ORF in a polycistronic message.

# Prokaryotic mRNAs

- In these cases, the start codon of the downstream ORF often overlaps the 3' end of the upstream ORF.

# Prokaryotic mRNAs

- Thus, a ribosome that has just completed translating the upstream ORF is positioned to begin translating from the start codon for the downstream ORF.

# Prokaryotic mRNAs

- This phenomenon of linked translation between overlapping ORFs is known as **translational coupling**.
- So in this situation translation of the downstream ORF requires translation of the upstream ORF.



# Prokaryotic mRNAs

- Indeed, with two translationally coupled genes, a mutation that leads to a premature stop codon in the upstream ORF also prevents translation of the downstream ORF.

END

# Eukaryotic mRNAs

- Unlike their prokaryotic counterparts, eukaryotic mRNAs recruit ribosomes using a specific chemical modification called the **5' cap**, which is located at the extreme 5' end of the mRNA.

# Eukaryotic mRNAs

- The 5' cap is a methylated guanine nucleotide that is joined to the 5' end of the mRNA via an unusual 5'-to-5' linkage.

# Eukaryotic mRNAs

- Created in three steps, the guanine nucleotide of the 5' cap is connected to the 5' end of the mRNA through three phosphate groups.

# Eukaryotic mRNAs

- The resulting 5' cap is required to recruit the ribosome to the mRNA. Once bound to the mRNA, the ribosome moves in a 5' → 3' direction until it encounters a 5'-AUG-3' start codon, a process called **scanning**.

# Eukaryotic mRNAs

- Two other features of eukaryotic mRNAs stimulate translation. One feature is the presence, in some mRNAs, of a purine three bases upstream of the start codon and a guanine immediately downstream (5'-G/ANNAUGG-3').

# Eukaryotic mRNAs

- This sequence was originally identified by Marilyn Kozak and is referred to as the Kozak sequence. Many eukaryotic mRNA lack these bases, but their presence increases the efficiency of translation.



# Eukaryotic mRNAs

- In contrast to the situation in prokaryotes, these bases are thought to interact with the initiator tRNA, not with an RNA component of the ribosome.

# Eukaryotic mRNAs

- A second feature that contributes to efficient translation is the presence of a poly-A tail at the extreme 3' end of the mRNA.
- This tail is added enzymatically by the enzyme poly-A polymerase.

# Eukaryotic mRNAs

- Despite its location at the 3' end of the mRNA, the poly-A tail enhances the level of translation of them RNA by enhancing the recruitment of key translation initiation factors.

# Eukaryotic mRNAs

- Importantly, in addition to their roles in translation, these 5'- and 3'-end modifications also protect eukaryotic mRNAs from rapid degradation.

END

# Transfer RNA

- The heart of protein synthesis is the “translation” of nucleotide sequence information (in the form of codons) into amino acids.

# Transfer RNA

- This is accomplished by tRNA molecules, which act as adaptors between codons and the amino acids they specify.

# Transfer RNA

- There are many types of tRNA molecules, but each is attached to a specific amino acid, and each recognizes a particular codon, or codons, in the mRNA (most tRNAs recognize more than one codon).



# Transfer RNA

- tRNA molecules are between 75 and 95 ribonucleotides in length.
- Although the exact sequence varies, all tRNAs have certain features in common.

# Transfer RNA

- First, all tRNAs end at the 3' terminus with the sequence 5'-CCA-3'.
- Consistent with this absolute conservation, the 3' end of this sequence is the site that is attached to the cognate amino acid.

# Transfer RNA

- A second striking aspect of tRNAs is the presence of several unusual bases in their primary structure.
- These unusual features are created post-transcriptionally by enzymatic modification of normal bases in the polynucleotide chain.

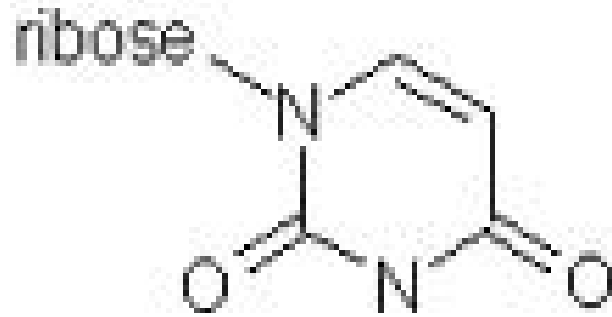
# Transfer RNA

- For example, pseudouridine ( $\psi$ U) is derived from uridine by an isomerization in which the site of attachment of the uracil base to the ribose is switched from the nitrogen at ring position 1 to the carbon at ring position 5.

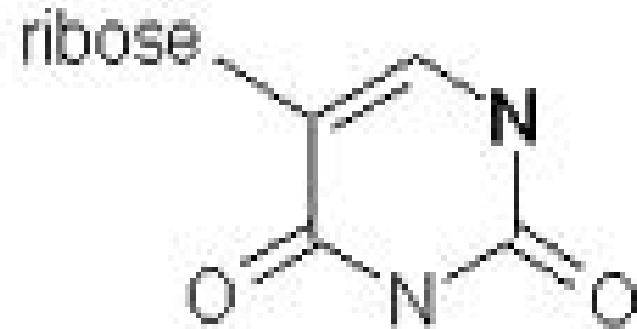
# Transfer RNA

- Likewise, dihydrouridine (D) is derived from uridine by enzymatic reduction of the double bond between the carbons at positions 5 and 6.

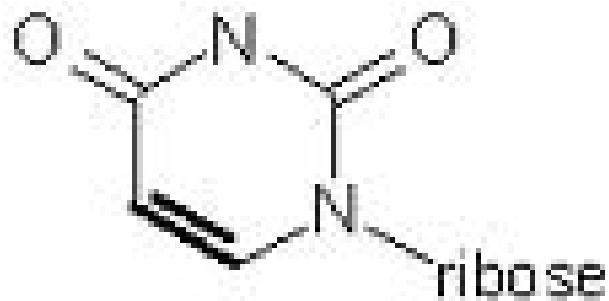
# Transfer RNA



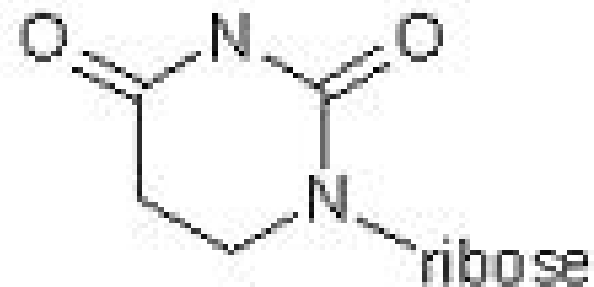
**uridine**



**Pseudouridine**



**uridine**



**dihydrouridine**

# Transfer RNA

- Other unusual bases found in tRNA include hypoxanthine, thymine, and methylguanine.
- These modified bases are not essential for tRNA function, but cells lacking these modified bases show reduced rates of growth.



# Transfer RNA

- This observation suggests that the modified bases lead to improved tRNA function.
- For example, hypoxanthine plays an important role in the process of codon recognition by certain tRNAs.

END

# Secondary Structure of tRNA

- RNA molecules typically contain regions of self complementarity that enable them to form limited stretches of double helix that are held together by base pairing.

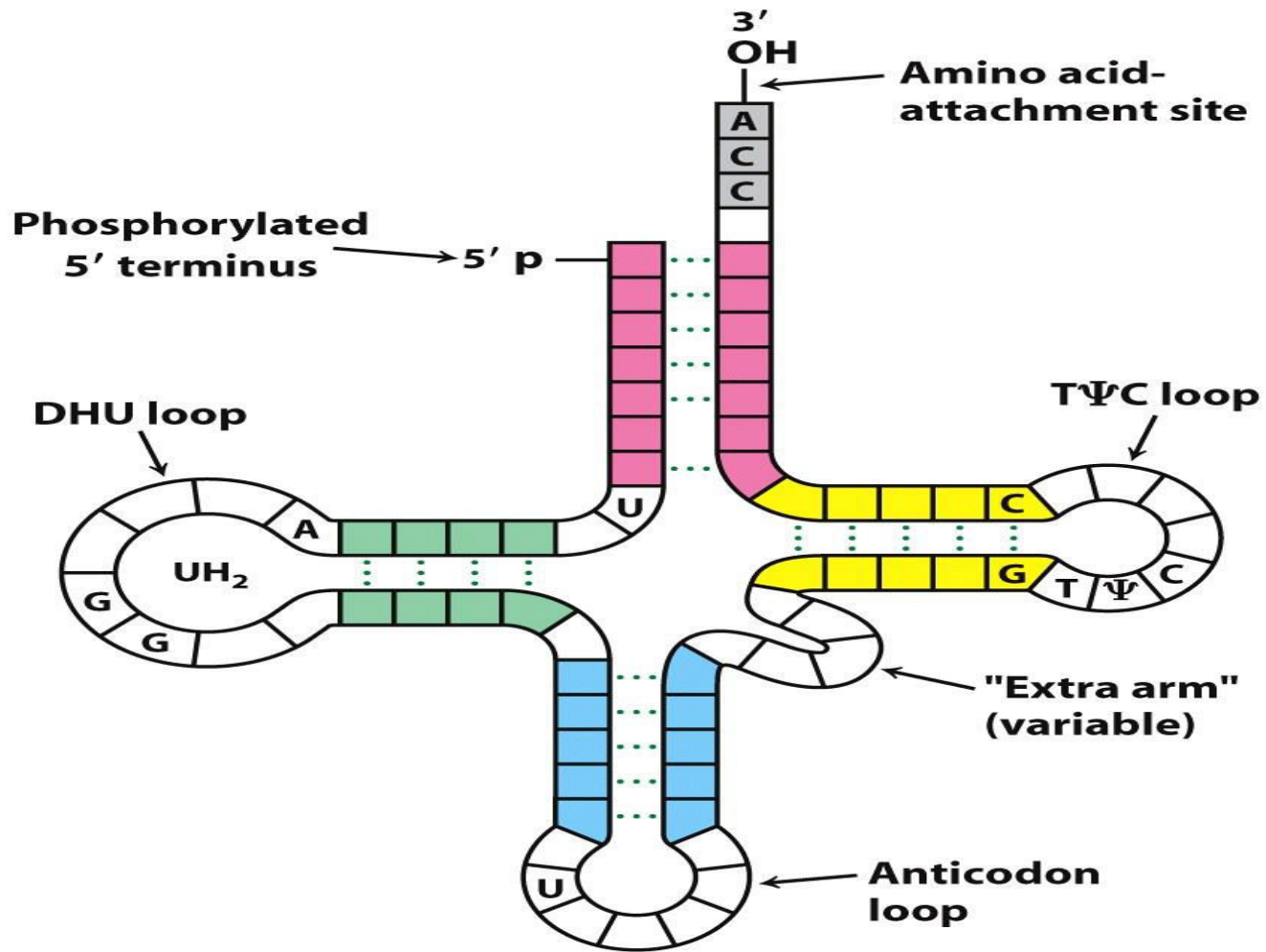
# Secondary Structure of tRNA

- tRNA molecules show a characteristic and highly conserved pattern of single-stranded and double stranded regions (secondary structure) that can be illustrated as a cloverleaf.

# Secondary Structure of tRNA

- The principal features of the tRNA cloverleaf are an acceptor stem, three stem-loops (referred to as the  $\psi$ U loop, the D loop, and the anticodon loop), and a fourth variable loop.

# Secondary Structure of tRNA



**Figure 30.3**  
*Biochemistry, Seventh Edition*  
© 2012 W. H. Freeman and Company

# Secondary Structure of tRNA

- **The Acceptor Stem:**
- It is so-named because it is the site of attachment of the amino acid, is formed by pairing between the 5' and 3' ends of the tRNA molecule.

# Secondary Structure of tRNA

- **The  $\psi$ U Loop:**
- It is so-named because of the characteristic presence of the unusual base  $\psi$ U in the loop. The modified base is often found within the sequence 5'-TCUCG-3'.



# Secondary Structure of tRNA

- **The D Loop:**
- It takes its name from the characteristic presence of dihydrouridines in the loop.

# Secondary Structure of tRNA

- **The Anticodon Loop:**
- As its name implies, contains the anticodon, a three-nucleotide-long sequence that is responsible for recognizing the codon by base pairing with the mRNA.

# Secondary Structure of tRNA

- **The Variable Loop:**
- It sits between the anticodon loop and the  $\psi$ U loop and, as its name implies, varies in size from 3 to 21 bases.

END

# Attachment of Amino Acids to tRNA

- tRNA molecules to which an amino acid is attached are said to be **charged**, and tRNAs that lack an amino acid are said to be **uncharged**.

# Attachment of Amino Acids to tRNA

- Charging requires an acyl linkage between the carboxyl group of the amino acid and the 2'- or 3'-hydroxyl group of the adenosine nucleotide that protrudes from the acceptor stem at the 3' end of the tRNA.

# Attachment of Amino Acids to tRNA

- This acyl linkage is a high-energy bond because its hydrolysis results in a large change in free energy.

# Attachment of Amino Acids to tRNA

- This is significant for protein synthesis: the energy released when this acyl bond is broken is coupled to the formation of the peptide bonds that link amino acids to each other in polypeptide chains.



# Attachment of Amino Acids to tRNA

- All aminoacyl-tRNA synthetases attach an amino acid to a tRNA in two enzymatic steps:
  - Adenylation
  - tRNA charging

# Attachment of Amino Acids to tRNA

- Step one is **adenylation** in which the amino acid reacts with ATP to become adenylylated with the concomitant release of pyrophosphate.

# Attachment of Amino Acids to tRNA

- Adenylylation refers to transfer of AMP, as opposed to adenylation, which would indicate the transfer of adenine.

# Attachment of Amino Acids to tRNA

- The principal driving force for the adenylation reaction is the subsequent hydrolysis of pyrophosphate by pyrophosphatase.

# Attachment of Amino Acids to tRNA

- As a result of adenylylation, the amino acid is attached to adenylic acid via a high-energy ester bond in which the carbonyl group of the amino acid is joined to the phosphoryl group of AMP.

# Attachment of Amino Acids to tRNA

- Step two is tRNA Charging in which the adenylylated amino acid, which remains tightly bound to the synthetase, reacts with tRNA.

# Attachment of Amino Acids to tRNA

- This reaction results in the transfer of the amino acid to the 3' end of the tRNA via the 2'- or 3'-hydroxyl and the release of AMP.



# Attachment of Amino Acids to tRNA

- There are two classes of tRNA synthetases:
- Class I enzymes attach the amino acid to the 2'-OH of the tRNA and are generally monomeric.

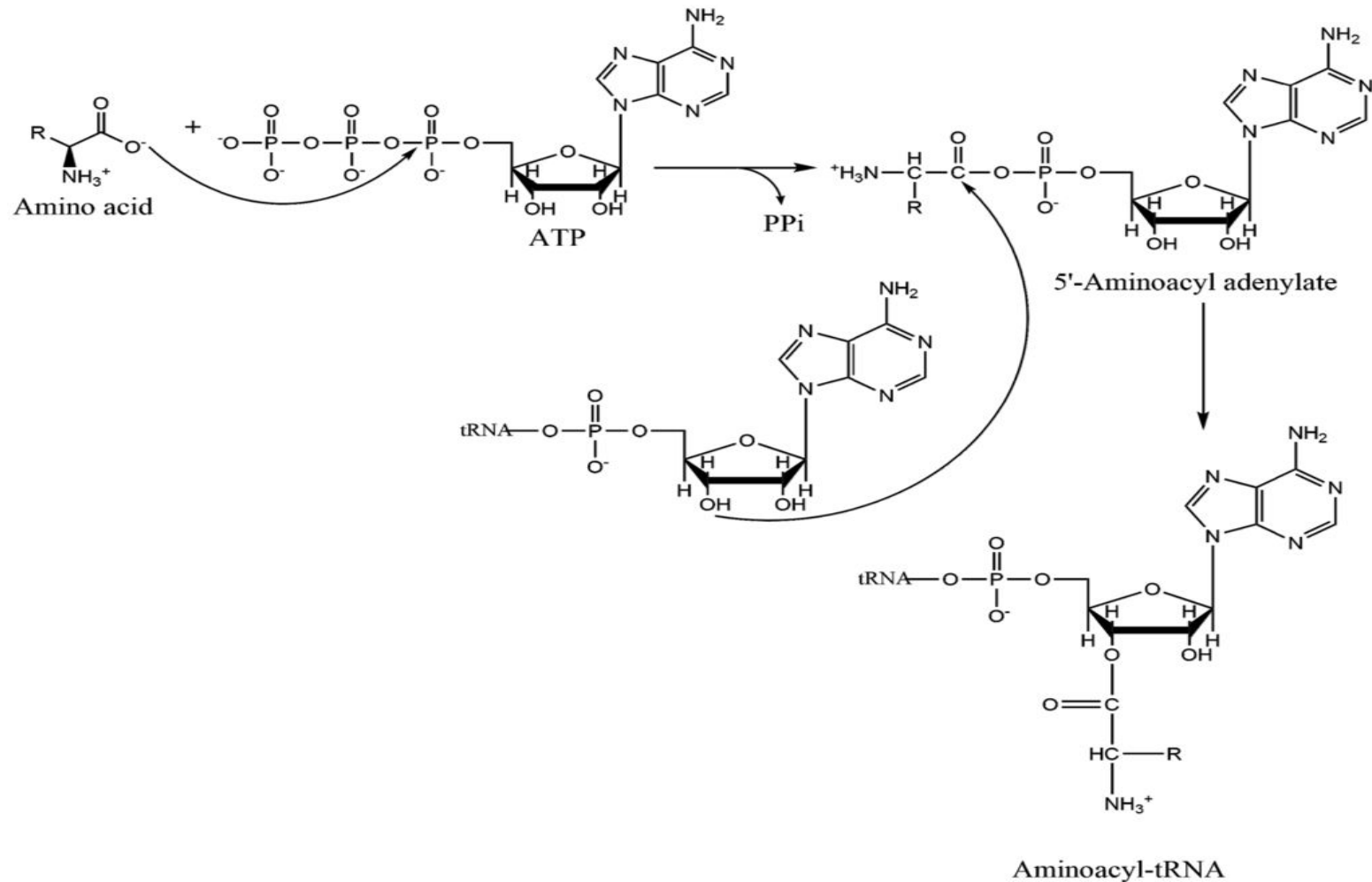
# Attachment of Amino Acids to tRNA

- Class II enzymes attach the amino acid to the 3'-OH of the tRNA and are typically dimeric or tetrameric.

# Attachment of Amino Acids to tRNA

- Although the initial coupling between the tRNA and the amino acid is different, once released from the synthetase, the amino acid rapidly equilibrates between attachment at the 3'-OH and the 2'-OH.

# Attachment of Amino Acids to tRNA



# Attachment of Amino Acids to tRNA

- Each of the 20 amino acids is attached to the appropriate tRNA by a single, dedicated tRNA synthetase.

# Attachment of Amino Acids to tRNA

- Because most amino acids are specified by more than one codon, it is not uncommon for one synthetase to recognize and charge more than one tRNA (known as **isoaccepting tRNAs**).

# Attachment of Amino Acids to tRNA

- Nevertheless, the same tRNA synthetase is responsible for charging all tRNAs for a particular amino acid.
- Thus, one and only one tRNA synthetase attaches each amino acid to all of the appropriate tRNAs.

END



# The Ribosomes

- The ribosome is the macromolecular machine that directs the synthesis of proteins.
- The ribosome is larger and more complex than the minimal machinery required for DNA or RNA synthesis.

# The Ribosomes

- The machinery for polymerizing amino acids is composed of at least three RNA molecules and more than 50 different proteins, with an overall molecular mass of  $>2.5$  MDa.

# The Ribosomes

- Compared with the speed of DNA replication i.e., 200 – 1000 nucleotides per second; translation takes place at a rate of only two to 20 amino acids per second.

# The Ribosomes

- In prokaryotes, the transcription machinery and the translation machinery are located in the same compartment. Thus, the ribosome can commence translation of the mRNA as it emerges from the RNA polymerase.

# The Ribosomes

- This situation allows the ribosome to proceed in tandem with the RNA polymerase as it elongates the transcript.

# The Ribosomes

- Recall that the 5' end of an RNA is synthesized first, and thus the ribosome, which begins translation at the 5' end of the mRNA, can start translating a nascent transcript as soon as it emerges from the RNA polymerase.

END

# Formation of Peptide Bonds

- Each new amino acid is added to the carboxyl terminus of the growing polypeptide chain (often referred to as synthesis in the amino- to carboxy-terminal direction).



# Formation of Peptide Bonds

- The ribosome catalyzes a single chemical reaction — the formation of a peptide bond.

# Formation of Peptide Bonds

- This reaction occurs between the amino acid residue at the carboxy-terminal end of the growing polypeptide and the incoming amino acid to be added to the chain.

# Formation of Peptide Bonds

- Both the growing chain and the incoming amino acid are attached to tRNAs; as a result, during peptide-bond formation, the growing polypeptide is continuously attached to a tRNA.

# Formation of Peptide Bonds

- The actual substrates for each round of amino acid addition are two charged species of tRNAs — an aminoacyl-tRNA and a **peptidyl-tRNA**.

# Formation of Peptide Bonds

- As you know the aminoacyl-tRNA is attached at its 3' end to the carboxyl group of the amino acid. The peptidyl-tRNA is attached in exactly the same manner (at its 3' end) to the carboxyl terminus of the growing polypeptide chain.

# Formation of Peptide Bonds

- The bond between the aminoacyl-tRNA and the amino acid is not broken during the formation of the next peptide bond.

# Formation of Peptide Bonds

- Instead, the bond between the peptidyl-tRNA and the growing polypeptide chain is broken as the growing chain is attached to the amino group of the amino acid attached to the aminoacyl-tRNA to form a new peptide bond.



# Formation of Peptide Bonds

- To catalyze peptide-bond formation, the 3' ends of these two tRNAs are brought into close proximity by the ribosome.

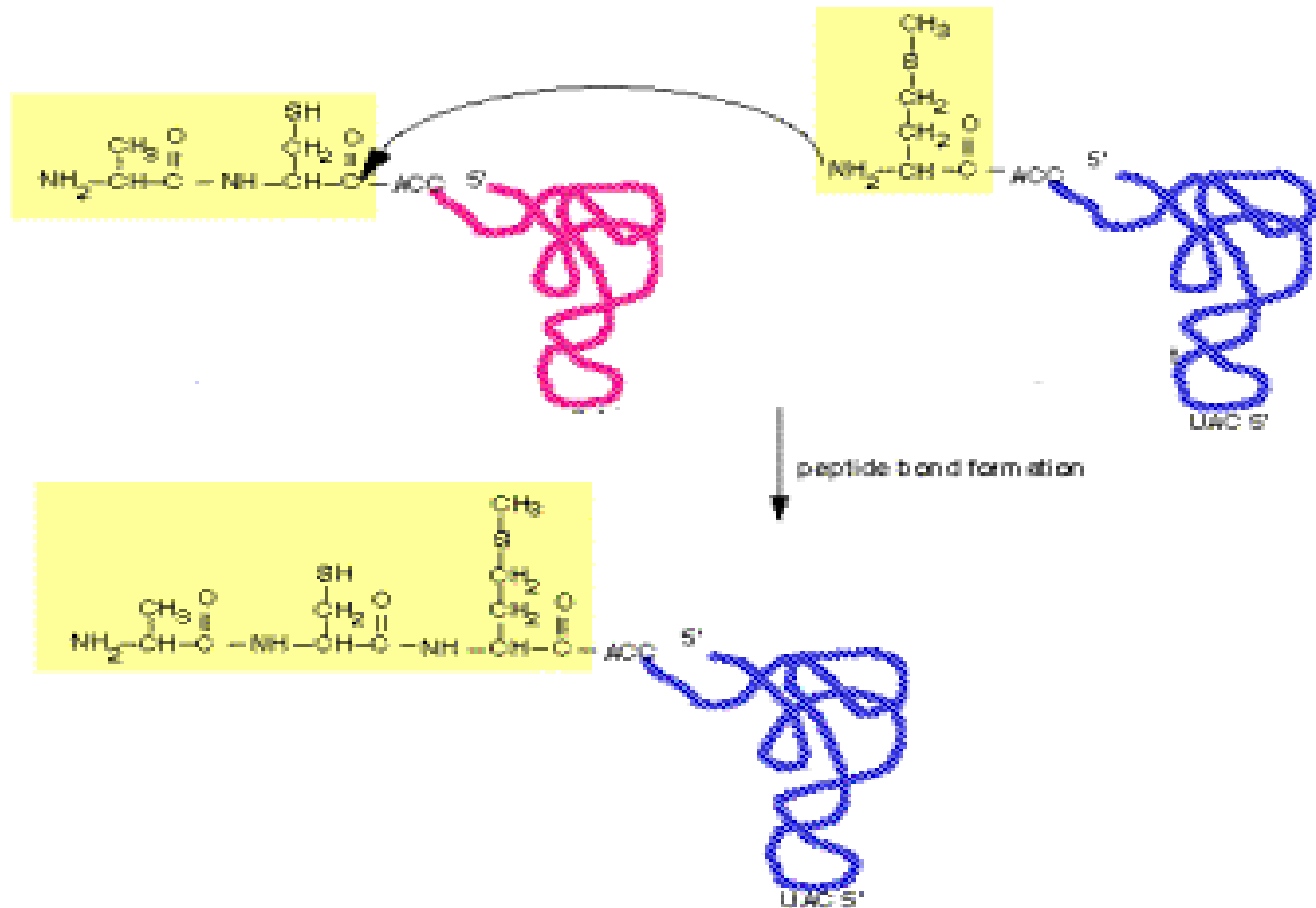
# Formation of Peptide Bonds

- The resulting tRNA positioning allows the amino group of the amino acid attached to aminoacyl-tRNA to attack the carbonyl group of the most carboxy-terminal amino acid attached to the peptidyl-tRNA.

# Formation of Peptide Bonds

- The result of this nucleophilic attack is the formation of a new peptide bond between the amino acids attached to the tRNAs and the release of the polypeptide chain from the peptidyl tRNA.

# Formation of Peptide Bonds



# Formation of Peptide Bonds

- There are two consequences of this method of polypeptide synthesis. First, this mechanism of peptide-bond formation requires that the amino terminus of the protein be synthesized before the carboxyl terminus.

# Formation of Peptide Bonds

- Second, the growing polypeptide chain is transferred from the peptidyl-tRNA to the aminoacyl-tRNA. For this reason, the reaction to form a new peptide bond is called the **peptidyl transferase reaction**.

# Formation of Peptide Bonds

- Interestingly, peptide-bond formation takes place without the simultaneous hydrolysis of a nucleoside triphosphate.



# Formation of Peptide Bonds

- This is because peptide-bond formation is driven by breaking the high-energy acyl bond that joins the growing polypeptide chain to the tRNA.

# Formation of Peptide Bonds

- Recall that this bond was created during the tRNA synthetase – catalyzed reaction that is responsible for charging tRNA.
- And the charging reaction involves the hydrolysis of a molecule of ATP.

# Formation of Peptide Bonds

- Thus, the energy for peptide-bond formation originates from the molecule of ATP that was hydrolyzed during the tRNA charging reaction.

END

# Binding Sites on Ribosomes for tRNA

- The ribosome is composed of two subassemblies of RNA and protein known as the large and small subunits.

# Binding Sites on Ribosomes for tRNA

- The large subunit contains the **peptidyl transferase center**, which is responsible for the formation of peptide bonds.

# Binding Sites on Ribosomes for tRNA

- The small subunit contains the **decoding center** in which charged tRNAs read or "decode" the codon units of the mRNA.

# Binding Sites on Ribosomes for tRNA

- Both the decoding center and the peptidyl transferase center are buried within the intact ribosome.



# Binding Sites on Ribosomes for tRNA

- Yet, mRNA must be threaded through the decoding center during translation, and the nascent polypeptide chain must escape from the peptidyl transferase center.

# Binding Sites on Ribosomes for tRNA

- How do these polymers enter and exit the ribosome?
- The answer is provided by the structure of the ribosome, which reveals that there are "tunnels" in and out of the ribosome.

# Binding Sites on Ribosomes for tRNA

- To perform the peptidyl transferase reaction, the ribosome must be able to bind at least two tRNAs simultaneously.
- In fact, the ribosome contains three tRNA-binding sites, called the A-, P-, and E-sites.

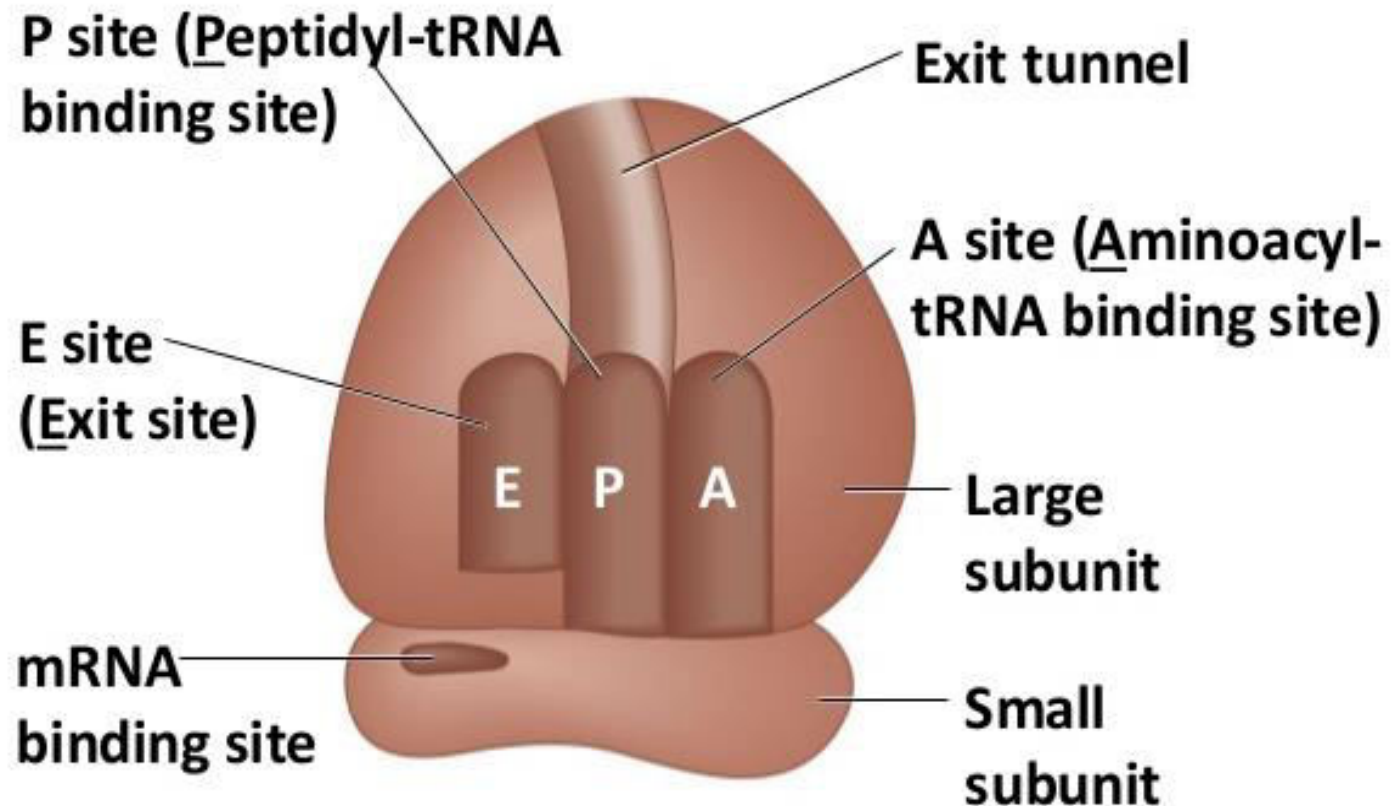
# Binding Sites on Ribosomes for tRNA

- The **A-site** is the binding site for the aminoacylated-tRNA, the **P-site** is the binding site for the peptidyl-tRNA, and

# Binding Sites on Ribosomes for tRNA

- The **E-site** is the binding site for the tRNA that is released after the growing polypeptide chain has been transferred to the aminoacyl-tRNA (E is for “exiting”).

# Binding Sites on Ribosomes for tRNA



# Binding Sites on Ribosomes for tRNA

- Each tRNA binding site is formed at the interface between the large and the small subunits of the ribosome.



# Binding Sites on Ribosomes for tRNA

- In this way, the bound tRNAs can span the distance between the peptidyl transferase center in the large subunit and the decoding center in the small subunit.

# Binding Sites on Ribosomes for tRNA

- The 3' ends of the tRNAs that are coupled to the amino acid or to the growing peptide chain are adjacent to the large subunit.
- The anticodon loops of the bound tRNAs are located adjacent to the small subunit.

END

# Initiation of Translation

- For translation to be successfully initiated, three events must occur:-
  - i) the ribosome must be recruited to the mRNA.
  - ii) a charged tRNA must be placed into the P-site of the ribosome.

# Initiation of Translation

- iii) the ribosome must be precisely positioned over the start codon.

# Initiation of Translation

- The correct positioning of the ribosome over the start codon is critical because this establishes the reading frame for the translation of the mRNA.

# Initiation of Translation

- In prokaryotes, the assembly of the ribosome on an mRNA occurs one subunit at a time. The small subunit associates with the mRNA first.

# Initiation of Translation

- In prokaryotes, the association of the small subunit with the mRNA is mediated by base-pairing interactions between the RBS and the 16S rRNA.



# Initiation of Translation

- For ideally positioned RBSs, the small subunit is positioned on the mRNA such that the start codon will be in the P-site when the large subunit joins the complex.

# Initiation of Translation

- The large subunit joins its partner only at the very end of the initiation process, just before the formation of the first peptide bond.
- Thus, many of the key events of translation initiation occur in the absence of the full ribosome.

# Initiation of Translation

- Translation initiation is the only time a tRNA binds to the P-site without previously occupying the A-site. This event requires a special tRNA known as the **initiator tRNA**.

# Initiation of Translation

- The initiator tRNA base-pairs with the start codon (AUG or GUG). AUG and GUG have a different meaning when they occur within an ORF, where they are read by tRNAs for methionine and valine, respectively.

# Initiation of Translation

- Although the initiator tRNA is first charged with a methionine, a formyl group is rapidly added to the methionine amino group by a separate enzyme (**Met-tRNA transformylase**).

# Initiation of Translation

- Thus rather than valine or methionine, the initiator tRNA is coupled to N-formyl methionine. The charged initiator tRNA is referred to as **fMet-tRNA<sup>fMet</sup>**.

# Initiation of Translation

- Because N-formyl methionine is the first amino acid to be incorporated into a polypeptide chain, one might think that all prokaryotic proteins have a formyl group at their amino termini.



# Initiation of Translation

- This is not the case, however, because an enzyme known as a **deformylase** removes the formyl group from the amino terminus during or after the synthesis of the polypeptide chain.

# Initiation of Translation

- In fact, many mature prokaryotic proteins do not even start with a methionine; aminopeptidases often remove the amino-terminal methionine as well as one or two additional amino acids.

END

# The Initiation Factors

- The initiation of prokaryotic translation commences with the small subunit and is catalyzed by three translation initiation factors called IF1, IF2, and IF3.
- Each factor facilitates a key step in the initiation process.

# The Initiation Factors

- **IF1:**
- It prevents tRNAs from binding to the portion of the small subunit that will become part of the A-site.

# The Initiation Factors

- **IF2:**
- It is a GTPase that interacts with three key components of the initiation machinery: the small subunit, IF1, and the charged initiator tRNA (fMet-tRNA<sup>fMet</sup>).

# The Initiation Factors

- By interacting with these components, IF2 facilitates the association of fMet-tRNA<sup>fMet</sup> with the small subunit and prevents other charged tRNAs from associating with the small subunit.

# The Initiation Factors

- **IF3:**
- It binds to the small subunit and blocks it from re-associating with a large subunit. Because initiation requires a free small subunit, the binding of IF3 is critical for a new cycle of translation.



# The Initiation Factors

- IF3 becomes associated with the small subunit at the end of a previous round of translation when it helps to dissociate the 70S ribosome into its large and small subunits.

# The Initiation Factors

- Each of the initiation factors binds at, or near, one of the three tRNA binding sites on the small subunit.

# The Initiation Factors

- Consistent with its role in blocking the binding of charged tRNAs to the A-site, IF1 binds directly to the portion of the small subunit that will become the A-site.

# The Initiation Factors

- IF2 binds to IF1 and reaches over the A-site into the P-site to contact the fMet - tRNA<sup>fMet</sup>.
- Finally, IF3 occupies the part of the small subunit that will become the E-site.

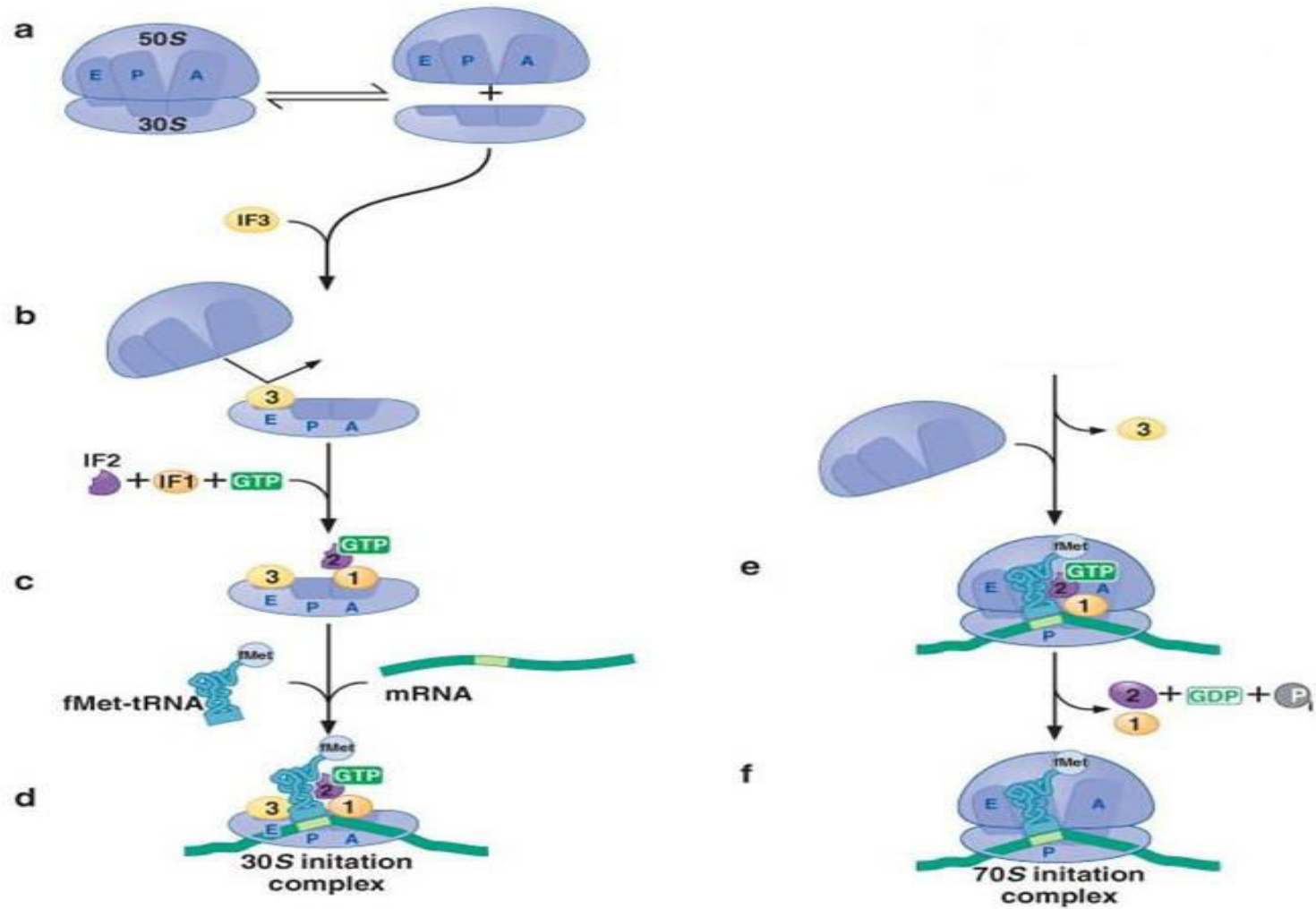
# The Initiation Factors

- Thus, of the three potential tRNA-binding sites on the small subunit, only the P-site is capable of binding a tRNA in the presence of the initiation factors.

# The Initiation Factors

- With all three initiation factors bound, the small subunit is prepared to bind to the mRNA and the initiator tRNA .
- These two RNAs can bind in either order and independently of each other.

# The Initiation Factors





# The Initiation Factors

- Binding fMet-tRNA<sup>fMet</sup> to the small subunit is facilitated by its interactions with IF2 bound to GTP and base pairing between the anticodon and the start codon of the mRNA.

# The Initiation Factors

- Similarly, base pairing between the fMet-tRNA<sup>fMet</sup> and the mRNA serves to position the start codon in the P-site.

# The Initiation Factors

- The last step of initiation involves the association of the large subunit to create the **70S initiation complex**.
- When the start codon and fMet-tRNA<sup>fMet</sup> base-pair, the small subunit undergoes a change in conformation.

# The Initiation Factors

- This altered conformation results in the release of IF3.
- In the absence of IF3, the large subunit is free to bind to the small subunit with its cargo of IF1, IF2, mRNA, and fMet-tRNA<sup>fMet</sup>.

# The Initiation Factors

- In particular, IF2 acts as an initial docking site of the large subunit, and this interaction subsequently stimulates the GTPase activity of IF2.GTP.

# The Initiation Factors

- IF2 bound to GDP has reduced affinity for the ribosome and the initiator tRNA, leading to the release of IF2.GDP as well as IF1 from the ribosome.

# The Initiation Factors

- Thus, the net result of initiation is the formation of an intact (70S) ribosome assembled at the start site of the mRNA with fMet-tRNA<sup>fMet</sup> in the P-site and an empty A-site.



# The Initiation Factors

- The ribosome – mRNA complex is now poised to accept a charged tRNA into the A-site and commence polypeptide synthesis.

END

# Translation Elongation

- Once the ribosome is assembled with the charged initiator tRNA in the P site, polypeptide synthesis can begin.
- There are three key events that must occur for the correct addition of each amino acid.

# Translation Elongation

- First, the correct aminoacyl-tRNA is loaded into the A site of the ribosome as dictated by the A-site codon.

# Translation Elongation

- Second, a peptide bond is formed between the aminoacyl-tRNA in the A site and the peptide chain that is attached to the peptidyl-tRNA in the P site.

# Translation Elongation

- This peptidyl transferase reaction results in the transfer of the growing polypeptide from the tRNA in the P site to the amino acid moiety of the charged tRNA in the A site.

# Translation Elongation

- Third, the resulting peptidyl-tRNA in the A site and its associated codon must be translocated to the P site so that the ribosome is poised for another cycle of codon recognition and peptide bond formation.

# Translation Elongation

- As with the original positioning of the mRNA, this shift must occur precisely to maintain the correct reading frame of the message.
- Two auxiliary proteins known as **elongation factors** control these events.



# Translation Elongation

- Both of these factors use the energy of GTP binding and hydrolysis to enhance the rate and accuracy of ribosome function.

# Translation Elongation

- Unlike the initiation of translation, the mechanism of elongation is highly conserved between prokaryotic and eukaryotic cells.

# Translation Elongation

- Aminoacyl-tRNAs do not bind to the ribosome on their own. Instead, they are "escorted " to the ribosome by the elongation factor EF-Tu.

# Translation Elongation

- Once a tRNA is aminocylated, EF-Tu binds to the tRNA's 3' end, masking the coupled amino acid. This interaction prevents the bound aminoacyl-tRNA from participating in peptide bond formation until it is released from EF-Tu.

# Translation Elongation

- Like the initiation factor IF2, the elongation factor EF-Tu binds and hydrolyzes GTP and the type of guanine nucleotide bound governs its function.

# Translation Elongation

- EF-Tu can only bind to an aminoacyl-tRNA when it is associated with GTP. EF-Tu bound to GDP, or lacking any bound nucleotide, shows little affinity for aminoacyl-tRNAs.

# Translation Elongation

- Thus, when EF-Tu hydrolyzes its bound GTP, any associated aminoacyl-tRNA is released.



# Translation Elongation

- The trigger that activates the EF-Tu GTPase is the same domain on the large subunit of the ribosome that activates the IF2 GTPase when the large subunit joins the initiation complex.

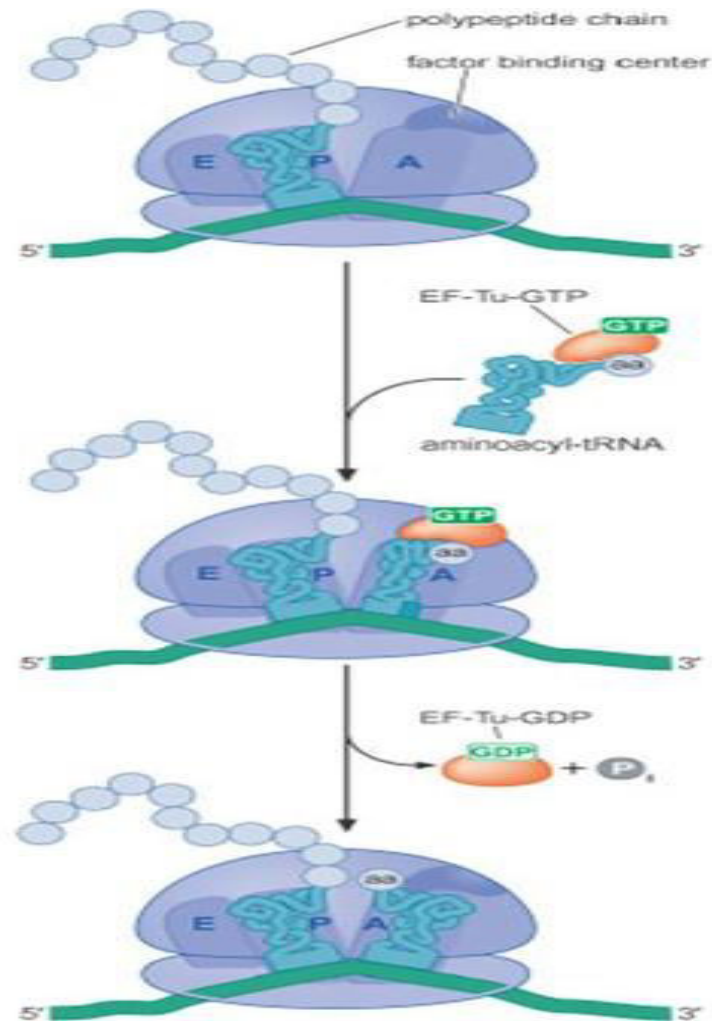
# Translation Elongation

- This domain is known as the **factor binding center**.
- EF-Tu only interacts with the factor binding center after the tRNA is loaded into the A site and a correct codon-anticodon match is made.

# Translation Elongation

- At this point, EF-Tu hydrolyzes its bound GTP and is released from the ribosome.
- The control of GTP hydrolysis by EF-Tu is critical to the specificity of translation.

# Translation Elongation



Pearson Education, Inc., publishing as Benjamin

# Translation Elongation

- The error rate of translation is between  $10^{-3}$  to  $10^{-4}$ .
- The ultimate basis for the selection of the correct aminoacyl-tRNA is the base pairing between the charged tRNA and the codon displayed in the A site of the ribosome.

# Translation Elongation

- However, in some cases, the base pairing in the anticodon-codon interaction may be mismatched, yet the ribosome rarely allows such mismatched aminoacyl-tRNAs to continue in the translation process.

END

# The Ribosome Is a Ribozyme

- Once the correctly charged tRNA has been placed in the A site and has rotated into the peptidyl transferase center, peptide bond formation takes place.



# The Ribosome Is a Ribozyme

- This reaction is catalyzed by RNA, specifically the 23 S rRNA component of the large subunit.

# The Ribosome Is a Ribozyme

- Early evidence for this came from experiments in which it was shown that a large subunit that had been largely stripped of its proteins was still able to carry out peptide bond formation.

# The Ribosome Is a Ribozyme

- Proof that the peptidyl transferase is entirely composed of RNA has come from the high-resolution, three-dimensional structure of the ribosome, which reveals that no amino acid is located closer than 18 Å from the active site.

# The Ribosome Is a Ribozyme

- Because catalysis requires distances in the 1 - 3 Å range, it is clear that the peptidyl transferase center is a ribozyme. That is an enzyme composed of RNA.

# The Ribosome Is a Ribozyme

- How does the 23 S rRNA catalyze peptide bond formation?
- The exact mechanism remains to be determined, but some answers to this question are beginning to emerge.

# The Ribosome Is a Ribozyme

- First, base-pairing between the 23 S rRNA and the CCA ends of the tRNAs in the A and the P sites help to position the alpha-amino group of the aminoacyl-tRNA to attack the carbonyl group of the growing polypeptide attached to the peptidyl-tRNA.

# The Ribosome Is a Ribozyme

- These interactions are also likely to stabilize the aminoacyl-tRNA after accommodation.



# The Ribosome Is a Ribozyme

- Because close proximity of substrates is rarely sufficient to generate high levels of catalysis, it is hypothesized that other elements of the ribosomal RNA change the chemical environment of the peptidyl transferase active site.

# The Ribosome Is a Ribozyme

- For example, it has been proposed that nucleotides in the peptidyl transferase center accept a hydrogen from the alpha amino group of the aminoacyl-tRNA, making the associated nitrogen a stronger nucleophile.

# The Ribosome Is a Ribozyme

- This is a common mechanism used by many proteins to stimulate nucleophilic attack of carbonyl groups.

END

# Translocation in the Large Subunit

- Once the peptidyl transferase reaction has occurred, the tRNA in the P-site is deacetylated (no longer attached to an amino acid), and the growing polypeptide chain is linked to the tRNA in the A-site.

# Translocation in the Large Subunit

- For a new round of peptide chain elongation to occur, the P-site tRNA must move to the E-site and the A-site tRNA must move to the P-site.
- At the same time, the mRNA must move by three nucleotides to expose the next codon.

# Translocation in the Large Subunit

- These movements are coordinated within the ribosome and are collectively referred to as **translocation**.
- The initial steps of translocation are coupled to the peptidyl transferase reaction.



# Translocation in the Large Subunit

- Once the growing peptide chain has been transferred to the A-site tRNA, the A- and P-site tRNAs have a preference to occupy new positions in the large subunit.



# Translocation in the Large Subunit

- The 3' end of the A-site tRNA is bound to the growing polypeptide chain and prefers to bind in the P-site of the large subunit.

# Translocation in the Large Subunit

- The now deacetylated P-site tRNA is no longer attached to the growing polypeptide chain and prefers to bind in the E-site of the large subunit.

# Translocation in the Large Subunit

- In contrast, at this time, the anticodons of these tRNAs remain in their initial location in the small subunit bound to the mRNA.

# Translocation in the Large Subunit

- Thus, translocation is initiated in the large subunit before the small subunit, and the tRNAs are said to be in “hybrid states.”

# Translocation in the Large Subunit

- Their 3' ends have shifted into a new location, but their anticodon ends are still in their pre-peptidyl transfer position.

# Translocation in the Large Subunit

- Importantly, this change is associated with a counter clockwise rotation of the small subunit relative to the large subunit facilitating interaction of the tRNAs with distinct tRNA-binding sites in the different subunits.

# Translocation in the Large Subunit

- The completion of translocation requires the action of a second elongation factor called **EF-G**.
- Initial binding of EF-G to the ribosome occurs when associated with GTP.



# Translocation in the Large Subunit

- After the peptidyl transferase reaction, EF-G-GTP binds to and stabilizes the ribosome in the rotated, hybrid state.

# Translocation in the Large Subunit

- When EF-G–GTP binds, it contacts the factor-binding center of the large subunit, which stimulates GTP hydrolysis.
- GTP hydrolysis changes the conformation of EF-G with two consequences.

# Translocation in the Large Subunit

- First, interactions between EF-G–GDP and the ribosome are thought to “unlock” the ribosome.

# Translocation in the Large Subunit

- Structural studies reveal that there are “gates” that separate the A-, P-, and E-sites and EF-G–GDP is said to unlock the ribosome by opening these gates.

# Translocation in the Large Subunit

- Second, the changed EF-G–GDP conformation binds to the A-site of the decoding center.
- This interaction competes with the tRNA for binding to the A-site of the decoding center.

# Translocation in the Large Subunit

- Because the ribosome is unlocked, the formerly A-site tRNA can move into the P-site, allowing EF-G-GDP to bind the A-site.

# Translocation in the Large Subunit

- Completion of translocation is accompanied by a clockwise rotation of the small subunit back to its starting position.
- The resulting ribosome structure has dramatically reduced affinity for EF-G–GDP.



# Translocation in the Large Subunit

- Release of EF-G results in the return of the ribosome to a “locked” state in which the tRNAs and mRNA are once again tightly associated with the small subunit decoding center and the gates between the A-, P- and E-sites are closed.

# Translocation in the Large Subunit

- Together, these events result in the translocation of the A-site tRNA into the P-site, the P-site tRNA into the E-site, and the movement of the mRNA by exactly 3 bp.
- The ribosome is now ready for a new cycle of amino acid addition to begin.

END

# Termination of Translation

- The ribosome's cycle of aminoacyl-tRNA binding, peptide-bond formation, and translocation continues until one of the three stop codons enters the A-site.

# Termination of Translation

- It was initially postulated that there would be one or more chain terminating tRNAs that would recognize these codons.
- However, this is not the case.

# Termination of Translation

- Instead, stop codons are recognized by proteins called **release factors (RFs)** that activate the hydrolysis of the polypeptide from the peptidyl-tRNA.

# Termination of Translation

- There are two classes of release factors.
- Class I release factors recognize the stop codons and trigger hydrolysis of the peptide chain from the tRNA in the P-site.



# Termination of Translation

- Prokaryotes have two class I release factors called RF1 and RF2.
- RF1 recognizes the stop codon UAG and RF2 recognizes the stop codon UGA.
- The third stop codon, UAA, is recognized by both RF1 and RF2.



# Termination of Translation

- In eukaryotic cells, there is a single class I release factor called **eRF1** that recognizes all three stop codons.
- Class II release factors stimulate the dissociation of the class I factors from the ribosome after release of the polypeptide chain.

# Termination of Translation

- Prokaryotes and eukaryotes have only one class II factor called **RF3** and **eRF3**, respectively.
- Like EF-G, IF2, and EF-Tu, class II release factors are regulated by GTP binding and hydrolysis.

# Termination of Translation

- How do release factors recognize stop codons?
- Because release factors are composed entirely of protein, protein–RNA interaction must mediate stop codon recognition.

# Termination of Translation

- Experiments in which short coding regions were genetically swapped between RF1 and RF2 (having different stop-codon specificity) identified a three-amino-acid sequence that is critical for release factor specificity.

# Termination of Translation

- Exchange of these three amino acids between RF1 and RF2 swaps the stop-codon specificity of the two complexes.

# Termination of Translation

- For this reason, this three-amino-acid sequence is called a peptide anticodon and must interact with and recognize stop codons.
- A 3D structure of RF1 bound to the ribosome confirms that RF1 binds to the A-site of the ribosome.



# Termination of Translation

- In this structure, the peptide anticodon is located very near the anticodon, but it is likely that there are additional protein regions that contribute to codon recognition.



# Termination of Translation

- A region of class I release factors that stimulates polypeptide release has also been identified.
- All class I factors share a conserved three-amino-acid sequence (glycine, glycine, glutamine) that is essential for polypeptide release.

# Termination of Translation

- Moreover, the structure of RF1 bound to the ribosome confirms that the GGQ motif is located in close proximity to the peptidyl transferase center.

# Termination of Translation

- It remains unclear whether the GGG motif is directly involved in the release of polypeptide from the peptidyl-tRNA or it induces a change in the peptidyl transferase center that allows the center itself to catalyze hydrolysis.

# Termination of Translation

- Studies of the conserved bases found adjacent to the CCA ends in the peptidyl transferase center indicate that several of these residues are required for peptide hydrolysis.

# Termination of Translation

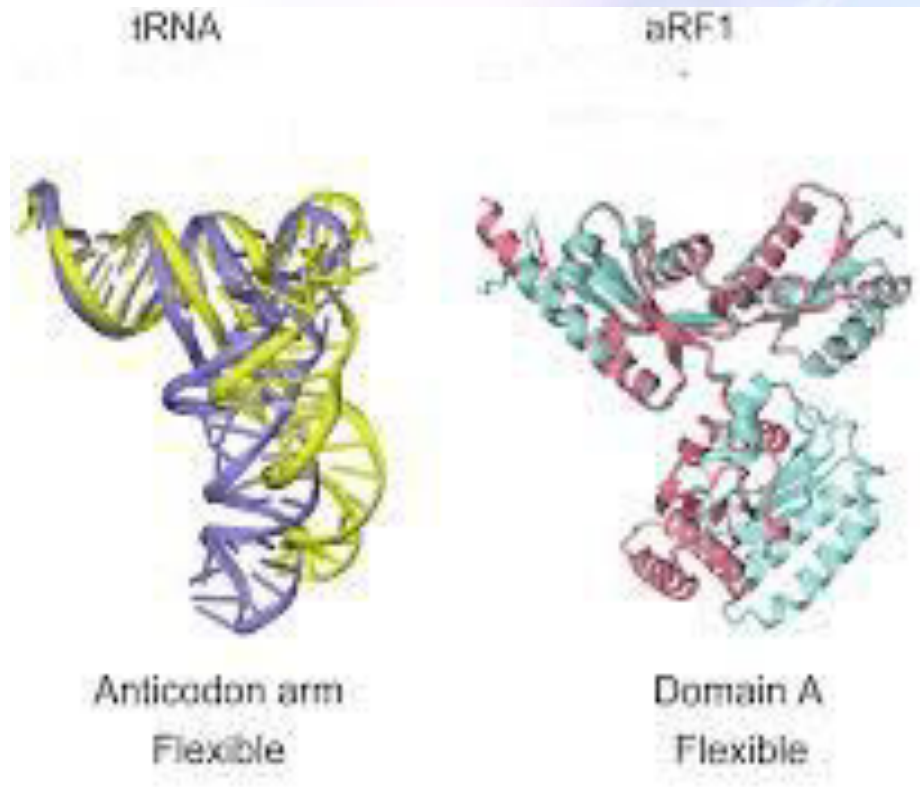
- Indeed, these bases appear to play a more important role in peptide release than they do in peptide-bond formation.

# Termination of Translation

- Together, these studies have led to the hypothesis that class I release factors functionally mimic a tRNA, having a peptide anticodon that interacts with the stop codon and a GGQ motif that reaches into the peptidyl transferase center.



# Termination of Translation





# Termination of Translation

- Once the class I release factor has triggered the hydrolysis of the peptidyl tRNA linkage, it must be removed from the ribosome.

# Termination of Translation

- This step is stimulated by the class II release factor, RF3.
- RF3 is a GTP-binding protein but, unlike the other GTP-binding proteins involved in translation, this factor has a higher affinity for GDP than GTP.

# Termination of Translation

- Thus, free RF3 is predominantly in the GDP-bound form.
- RF3-GDP binds to the ribosome in a manner that depends on the presence of a class I release factor.

# Termination of Translation

- After the class I release factor stimulates polypeptide release, a change in the conformation of the ribosome and the class I release factor stimulates RF3 to exchange its bound GDP for a GTP.

# Termination of Translation

- The binding of GTP to RF3 leads to the formation of a high-affinity interaction with the ribosome that favors the rotated hybrid state.
- This change in conformation displaces the class I factor from the ribosome.

# Termination of Translation

- These changes also allow RF3 to associate with the factor-binding center of the large subunit. As with other GTP-binding proteins involved in translation, this interaction stimulates the hydrolysis of GTP.

# Termination of Translation

- In the absence of a bound class I factor, the resulting RF3.GDP has a low affinity for the ribosome and is released.



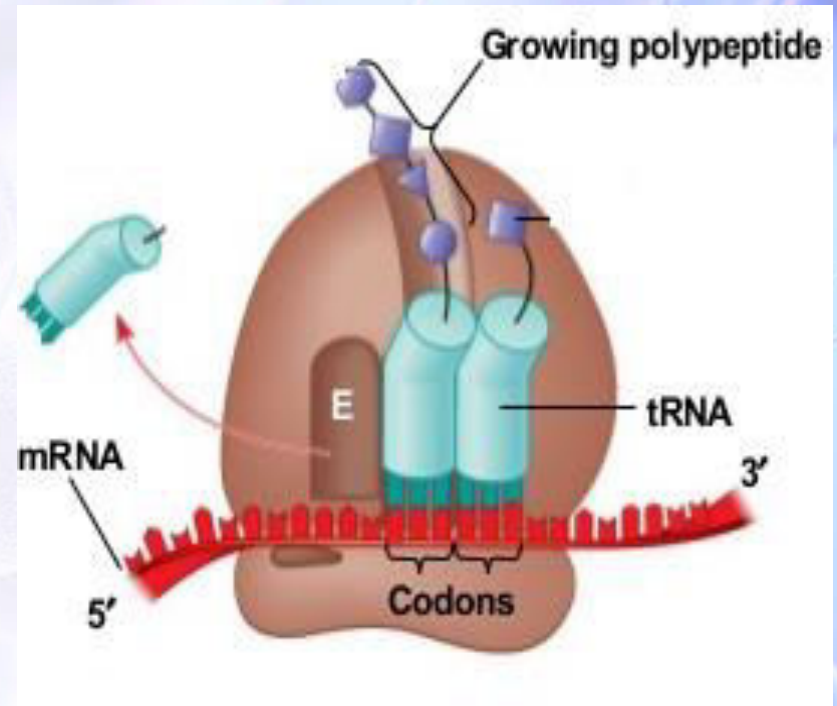
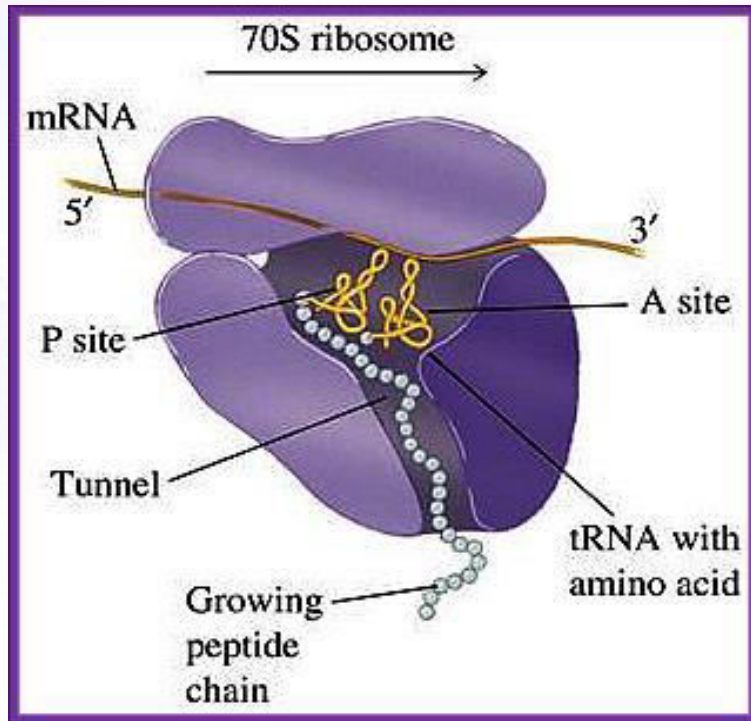
# Nascent polypeptide processing & folding

- The nascent protein passes through a peptide exit tunnel that extends from the peptide transferase centre to the ribosome surface.

# Nascent polypeptide processing & folding

- The exit tunnel in the bacterial ribosome is about 80-100 Å long and about 10 Å in its diameter at its narrowest point but widens to about twice that diameter at the rim of the exit pore.

# Nascent polypeptide processing & folding



# Nascent polypeptide processing & folding

- The exit tunnel can accommodate an  $\alpha$  helix with about 60 residues or an extended peptide with about half that number of residues.

# Nascent polypeptide processing & folding

- The space within the exit tunnel may permit the nascent polypeptide chain to assume an  $\alpha$ -helical conformation but is too narrow to permit more extensive folding.

# Nascent polypeptide processing & folding

- As the nascent peptide chain emerges from the tunnel, it interacts with enzymes that catalyze co-translational modifications, chaperones that assist in folding and prevent mis-folding,



# Nascent polypeptide processing & folding

- and the signal recognition particle that facilitates transport across the cell membrane.
- Peptide deformylase and methionine aminopeptidase bind at the rim of the bacterial ribosome's exit pore.



# Nascent polypeptide processing & folding

- The deformylase cleaves the N-terminal formyl group from the nascent polypeptide as it emerges from the exit tunnel.

# Nascent polypeptide processing & folding

- Then the aminopeptidase recognizes about 60% of the different nascent polypeptides and removes their N-terminal methionine.

# Nascent polypeptide processing & folding

- Nascent eukaryotic polypeptides do not have an N-terminal formyl group but do begin with methionine.
- Ribosome-bound methionine aminopeptidases remove the N-terminal methionine.

# Nascent polypeptide processing & folding

- In bacteria, co-translational protein folding is assisted by a 48 kDa chaperone called the **trigger factor**, which binds at the bacterial ribosome's exit pore.

# Nascent polypeptide processing & folding

- The trigger factor protein transiently associates with the L23 protein on the 50S subunit.
- Its residence time on the ribosome depends on whether the ribosome has a nascent protein in the exit tunnel.

# Nascent polypeptide processing & folding

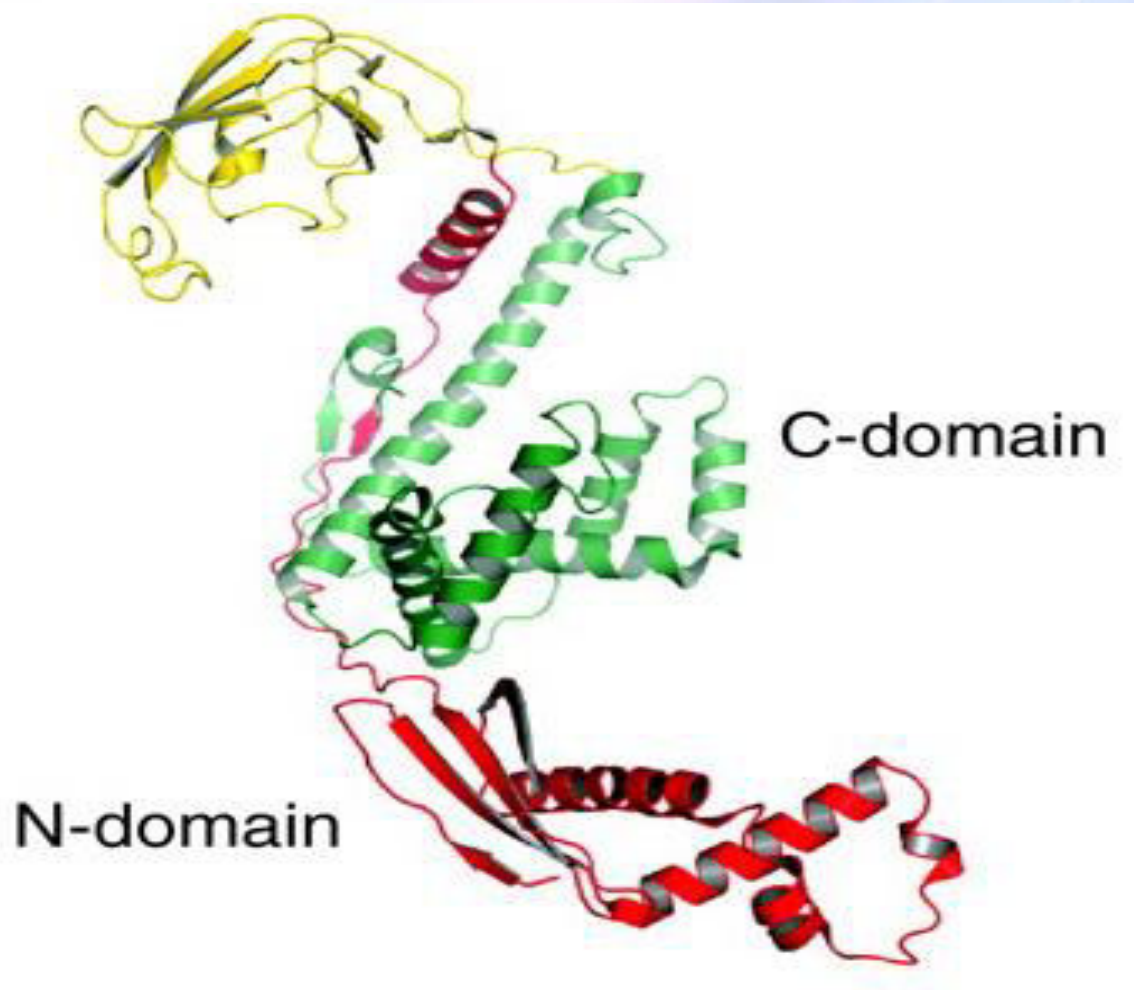
- If a nascent protein is not present, the average residence time is about 11 – 15 seconds but this time increases several – fold when a nascent protein is present.

# Nascent polypeptide processing & folding

- The trigger factor contains three domains that arrange to form a characteristic elongated dragon-shaped structure.
- The N-terminal domain forms the tail that binds to the 50S subunit.



# Nascent polypeptide processing & folding



# Nascent polypeptide processing & folding

- The middle domain which forms the dragon's head has peptidyl prolyl *cis/trans* isomerase activity that is not essential for the trigger factor's chaperone's function.

# Nascent polypeptide processing & folding

- The C-terminal domain the central body of the dragon and is responsible for the trigger factor's chaperone activity.

# Nascent polypeptide processing & folding

- The trigger factor binds to the hydrophobic patches as they emerge from the ribosome and sometimes remain associated with the segment even after polypeptide chain completion.

# Nascent polypeptide processing & folding

- Moreover, a single nascent polypeptide chain or free polypeptide may have two or more trigger factors associated with it.

# Nascent polypeptide processing & folding

- A chaperone also associates with the large subunit of the eukaryotic ribosome.
- This chaperone consists of three different subunits that differ in both sequence and structure from the trigger factor.



# Nascent polypeptide processing & folding

- In yeast, deletion of any single subunit results in slow growth and cold sensitivity.
- Protein folding is very important for the normal functioning of the cells.



# Nascent polypeptide processing & folding

- The accumulation of toxic proteins, which result from misfolding may lead to several neurological disorders including Alzheimer's, Huntington's and Parkinson's diseases.

# Signal Sequence

- The biologists observed that some ribosomes appear to exist free in the cytoplasm while others are bound to the endoplasmic reticulum in eukaryotes or cell membrane in prokaryotes.

# Signal Sequence

- The free ribosomes synthesize cytoplasmic and mitochondrial proteins whereas, the membrane-bound ribosomes synthesize integral membrane proteins, lysosomal proteins and secretory proteins.

# Signal Sequence

- The reason why some ribosomes bind to the endoplasmic reticulum was not clear at first.
- One possibility was that cells have two distinct kinds of ribosomes, free and bound.

# Signal Sequence

- Studies by Gunter Blobel and David D. Sabitini in 1971 indicated that free and bound ribosomes appeared to be the same.
- Therefore, there must be some other explanation for the same.

# Signal Sequence

- Blobel and Sabitini proposed the signal hypothesis to explain how cells determine whether a protein will be synthesized on a free or membrane-bound ribosome.
- According to this hypothesis:-

# Signal Sequence

- 1) Free and membrane-bound ribosomes are identical.
- 2) Protein synthesis always begins on free ribosomes.



# Signal Sequence

- 3) Nascent secretory, trans-membrane, or lysosomal proteins have sequences of 20-30 amino acids at their amino terminus that act as signals to bind the ribosomes to the endoplasmic reticulum.

# Signal Sequence

- Ribosomes only bind to the endoplasmic reticulum when synthesizing proteins with signal sequences.
- Although, signal sequences vary from one protein to another, certain common features can be recognized.

# Signal Sequence

- Signal sequences can be divided into three parts:
- a) a short, positively charged N-terminal region,
- b) a central region containing 7-13 hydrophobic amino acid residues,

# Signal Sequence

- c) and a more polar C-terminal region that includes a cleavage site.
- Only polypeptides that have signal sequence can be inserted into the endoplasmic reticulum membrane.

# Signal Sequence

- The amino end of a nascent polypeptide with a signal sequence requires the assistance of three components to pass into the lumen of the endoplasmic reticulum or to be integrated into the membrane.

# Signal Sequence

- These components are the signal recognition particle (SRP), the SRP receptor (SR), and a protein conducting channel or translocon.
- The components required for SRP cycle are as follows:-



# Signal Sequence

- 1) Signal recognition particle (SRP) consists of six proteins (SRP9, 14, 19, 54, 68, and 72 named according to their molecular masses) and a 7S RNA.



# Signal Sequence

- The SRP54 subunit is especially noteworthy because the guanine nucleotide binding site and the signal sequence binding site are located within it.

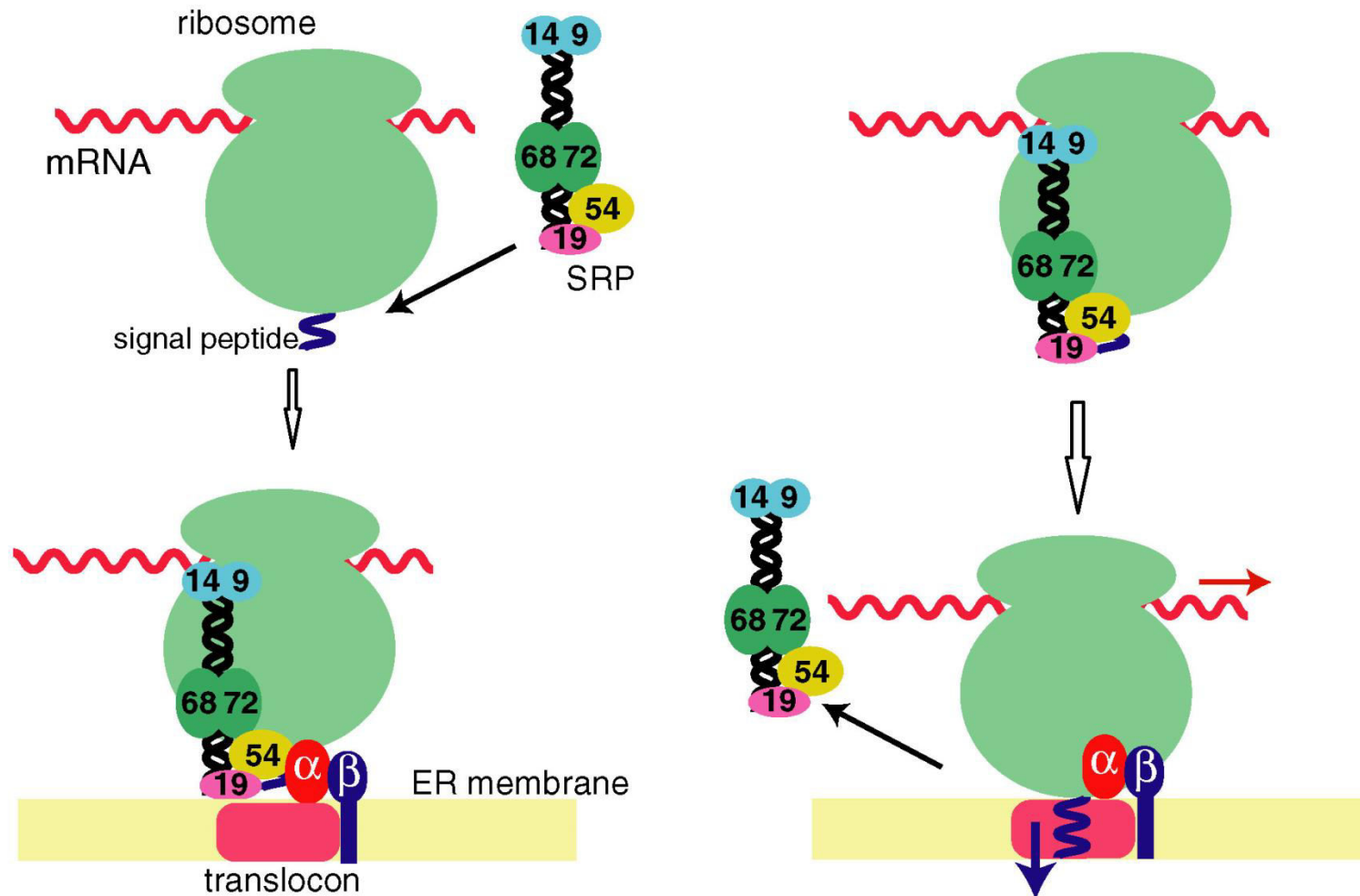
# Signal Sequence

- 2) SRP receptors contains two GTP-binding polypeptide subunits, a peripheral membrane protein  $SR\alpha$  and an integral membrane protein SR.

# Signal Sequence

- 3) Translocon. The protein conducting channel, which is made of three polypeptides (Sec61 $\alpha$ , Sec61 $\beta$  and Sec61 $\gamma$ ) acts as a passive conduit for polypeptides.

# Signal Sequence



# Signal Sequence

- Secretory and lysoosomal proteins pass completely through the endoplasmic reticulum membrane and are directed to their ultimate destination by biochemical modification such as

# Signal Sequence

- glycosylation that take place in the lumen of the endoplasmic reticulum and in the Golgi apparatus.
- Integral membrane proteins have one or more hydrophobic sequences that



# Signal Sequence

- allow them to move laterally across the translocon and into the membrane, causing the polypeptide chain to embed itself within the membrane.



# Regulation by RNAs in bacteria

- Small RNAs are involved in regulating the replication of plasmids, and others are involved in regulating gene expression.

# Regulation by RNAs in bacteria

- Some of these RNAs control transcription, for example the 6S RNA of *Escherichia coli*.
- This RNA binds to the  $\sigma^{70}$  subunit of RNA polymerase and down-regulates transcription from many  $\sigma^{70}$  promoters.

# Regulation by RNAs in bacteria

- The 6S RNA accumulates at high levels in stationary phase of bacterial growth.
- In stationary phase, an alternative  $\sigma$  factor,  $\sigma^S$ , is made.

# Regulation by RNAs in bacteria

- This  $\sigma$  factor competes with  $\sigma^{70}$  for core polymerase and directs the enzyme to promoters expressing genes for the multiple stress responses needed to survive stationary phase.

# Regulation by RNAs in bacteria

- By down-regulating transcription from  $\sigma^{70}$  promoters, 6S RNA helps this shift in expression to  $\sigma^S$  promoters.

# Regulation by RNAs in bacteria

- In recent years, attention has focused on small RNA molecules in bacteria that regulate translation and mRNA degradation.

# Regulation by RNAs in bacteria

- Interest in these small RNAs (small interfering and microRNAs) has been heightened by their similarity to RNAs that regulate gene expression in eukaryotes.



# Regulation by RNAs in bacteria

- One class of bacterial regulatory RNAs (called sRNAs) acts in *trans* to control translation of target genes, rather as microRNAs do in eukaryotes.

# Regulation by RNAs in bacteria

- They are, however, larger (80-110 nucleotides) than those eukaryotic regulatory RNAs (21 to 30 nucleotides), and they are not generally formed by processing of larger double-stranded RNA (dsRNA) precursors

# Regulation by RNAs in bacteria

- (as those eukaryotic RNA regulators are); instead, they are encoded in their final form by small genes.
- Many of these genes have been identified by bioinformatics, with more than 100 sRNAs being uncovered in *E. coli*.

# Regulation by RNAs in bacteria

- Most sRNAs work by base pairing with complementary sequences within target mRNAs and directing destruction of the mRNA, inhibiting its translation or even in some cases stimulating translation.

# Regulation by RNAs in bacteria

- Binding of an sRNA to its target mRNA is in most cases aided by the bacterial protein Hfq.

# Regulation by RNAs in bacteria

- This RNA chaperone is needed because the complementarity between the sRNAs and their target mRNAs is typically imperfect and short, and thus their interaction is weak.



# Regulation by RNAs in bacteria

- Hfq facilitates base pairing.
- Also, by binding the sRNAs even before they are paired with their targets, Hfq increases the stability of these regulators.
- A well-studied sRNA from *E. coli* is the 81-nucleotide RybB RNA.



# Regulation by RNAs in bacteria

- This sRNA binds several target mRNAs and triggers their destruction because the double-strand stretch of heteroduplex formed upon pairing is recognized as a substrate by the nuclease RNase E.

# Regulation by RNAs in bacteria

- Most of the mRNAs targeted by RybB encode iron storage proteins. Free iron is required by the cell under certain circumstances, but high levels are toxic.

# Regulation by RNAs in bacteria

- RybB regulates the levels of free iron by controlling the levels of iron storage proteins.
- RybB is expressed from a promoter recognized by a special  $\sigma$  factor called  $\sigma^E$  (like  $\sigma^S$ , a stress response  $\sigma$  factor).

# Regulation by RNAs in bacteria

- Expression of the gene encoding  $\sigma^E$  is itself regulated by RybB, and thus this sRNA is part of an autonegative regulatory loop for  $\sigma^E$ .

# Regulation by RNAs in bacteria

- The stationary-phase  $\sigma$  factor  $\sigma^S$  is encoded by the *rpoS* gene of *E. coli*.
- Translation of *rpoS* mRNA is stimulated by two sRNAs: DsrA and RprA.

# Regulation by RNAs in bacteria

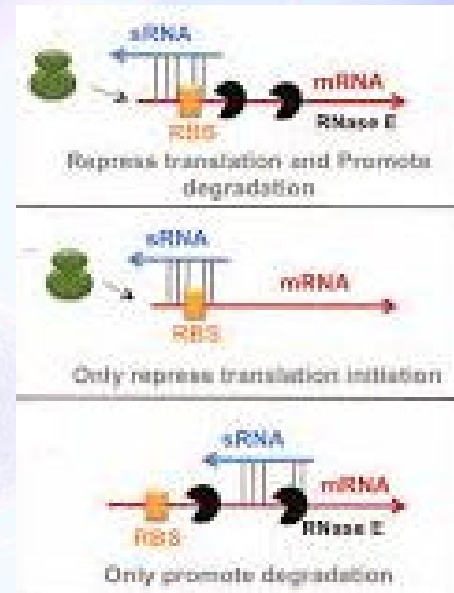
- Activation is achieved by a switch in alternative RNA base pairing: the small RNAs bind to a region of the mRNA that otherwise would pair with the ribosome-binding site, inhibiting translation.

# Regulation by RNAs in bacteria

- The rpoS gene is also acted on negatively by another small RNA, OxyS.



# Regulation by RNAs in bacteria



# Regulation by RNAs in bacteria

- These tend to be associated with genes encoding potentially toxic products, and also in regulation of some phage genes (as in  $\lambda$ ).

# Regulation by RNAs in bacteria

- These RNAs are often said to act in *cis* because they act only on the gene from which they are made (in contrast to the *trans*-acting sRNAs).

# Regulation by RNAs in bacteria

- Let us consider other examples of gene regulation mediated through alternative RNA pairing that truly operate in *cis*.
- These are RNA regulatory elements that control expression of the genes within whose mRNAs they reside.

# Regulation by RNAs in bacteria

- The most striking examples are the so-called that control metabolic operons and in biosynthetic operons.

# Regulation by RNAs in bacteria

- The *trp* genes of *E. coli* are the classic examples of the latter mechanism and are where RNA-mediated regulation was discovered.

# Riboswitches

- Riboswitches control gene expression in response to changes in the concentrations of small molecules.
- They do so through changes in RNA secondary structure.



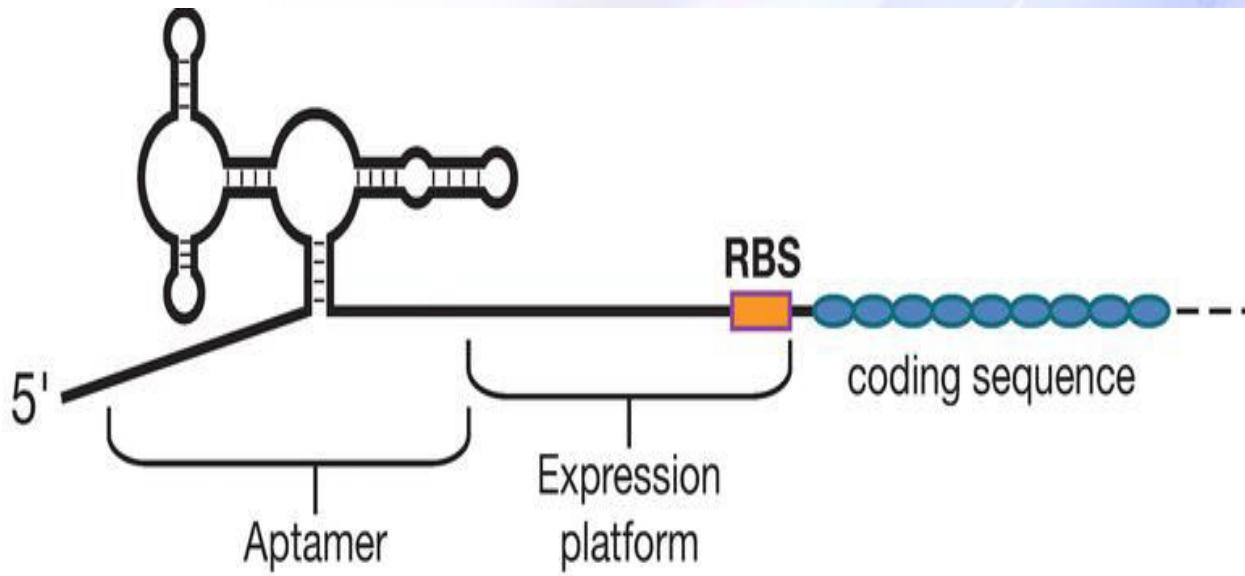
# Riboswitches

- These regulatory elements are typically found within the 5'-untranslated regions (5'-UTRs) of the genes they control.
- They can regulate expression at the level of transcription or translation.

# Riboswitches

- Each riboswitch is made up of two components:-
- the **aptamer** and,
- the **expression platform**.

# Riboswitches



# Riboswitches

- The aptamer binds the small-molecule ligand and, in response, undergoes a conformational change, which, in turn, causes a change in the secondary structure of the adjoining expression platform.

# Riboswitches

- These conformational changes alter expression of the associated gene by either terminating transcription or inhibiting the initiation of translation.

# Riboswitches

- Riboswitches are typically found upstream of genes involved in the synthesis of the metabolite ligand recognized by the riboswitch in question.

# Riboswitches

- For example, in *Bacillus subtilis*, many genes involved in the use of the amino acid methionine have a 200-nucleotide-long untranslated leader RNA that acts as a SAM (S-adenosyl methionine) – sensing riboswitch.



# Riboswitches

- RNA polymerase initiates transcription at the promoter and transcribes through this leader region before entering the coding sequence of the downstream genes.

# Riboswitches

- Once transcribed into RNA, the leader region can adopt alternative structures through alternative patterns of intramolecular base pairing.
- One arrangement includes a stem-loop transcriptional Terminator .

# Riboswitches

- SAM – the ligand for this riboswitch – binds to the aptamer and stabilizes the secondary structure that includes this transcriptional terminator.

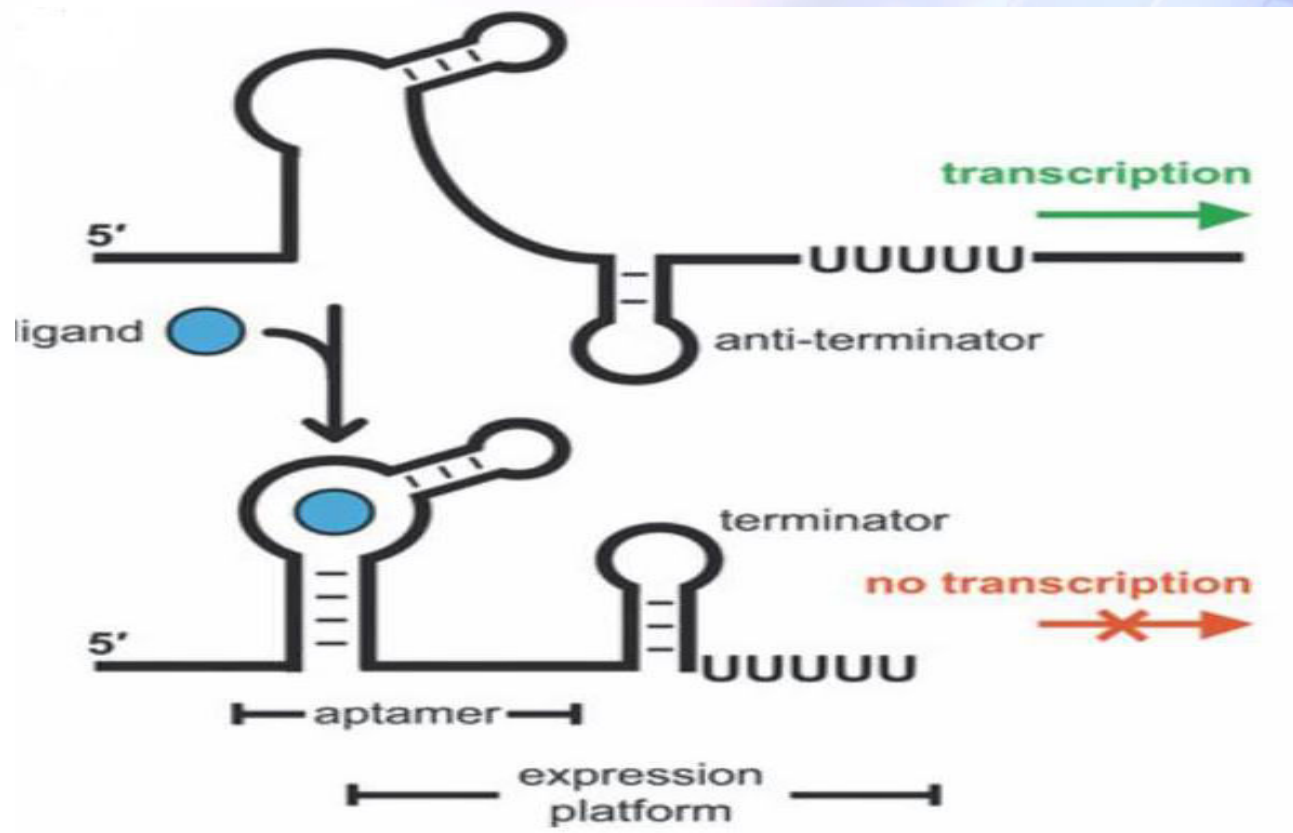
# Riboswitches

- Under these circumstances, transcription is terminated before polymerase has a chance to transcribe the downstream protein-coding segment of the gene.

# Riboswitches

- This form of transcriptional regulation is also called **attenuation**.

# Riboswitches



# Riboswitches

- In another case - at another gene - a SAM-sensing riboswitch can work by regulating translation.



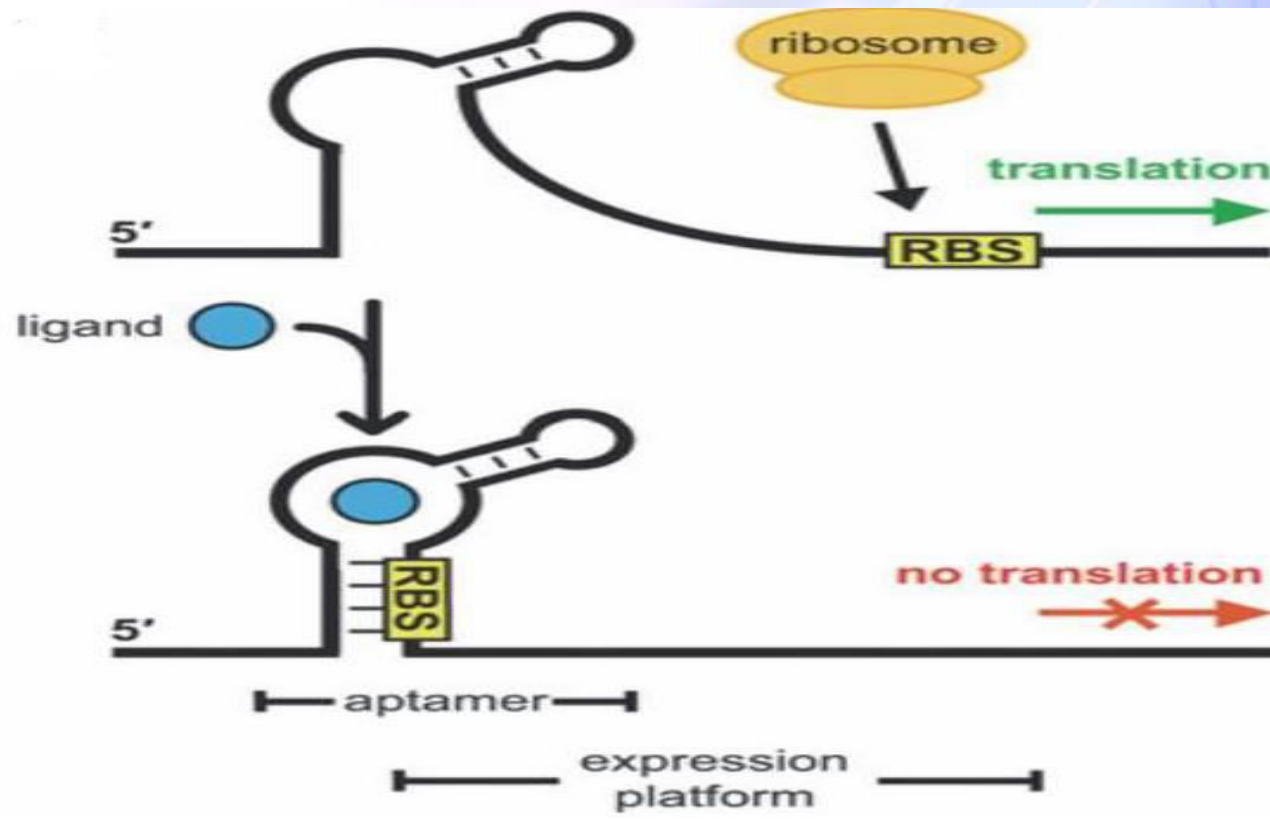
# Riboswitches

- In that case, the alternative secondary structure stabilized by SAM binding to the aptamer includes a stem-loop that, although not a transcriptional terminator, does include the ribosome-binding site.

# Riboswitches

- This conformational change sequesters the RBS and blocks ribosomes from initiating translation.
- This form of translation inhibition is thus essentially identical to that described for trans-acting sRNAs.

# Riboswitches



# Riboswitches

- Many riboswitches have been identified, and current whole-genome sequencing results suggest there are probably many hundreds or thousands found across bacterial species.

# Riboswitches

- Even well-characterized examples respond to a range of different metabolites, including lysine, vitamin B12, coenzyme thiamine pyrophosphate (TPP), flavin mononucleotide (FMN), and guanine.

# Riboswitches

- Although most prevalent in bacteria, riboswitches are found in other organisms as well, including archaea, fungi, and plants.

# Riboswitches

- In some cases in these higher organisms, riboswitches are even involved in controlling alternative splicing.



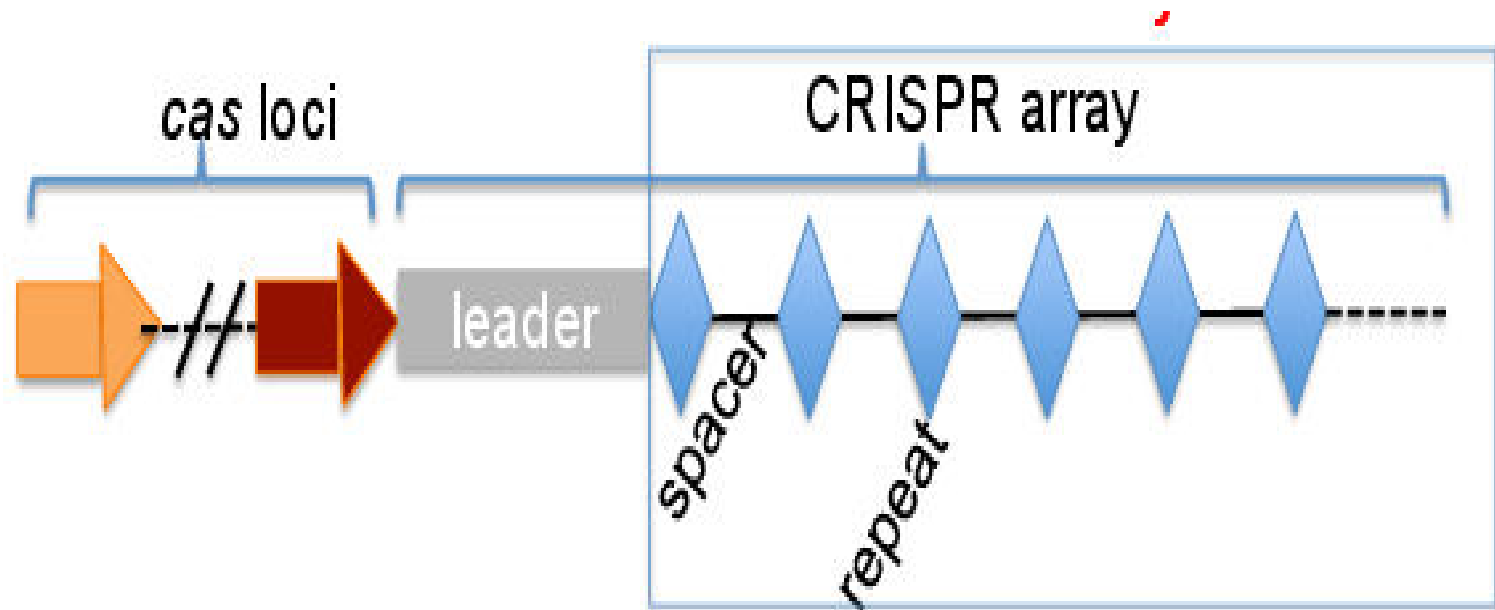
# Riboswitches

- For example, in case of the fungus *Neurospora crassa*, three TPP aptamers were identified, two of which inhibited, and the third stimulated, expression of genes through regulation of RNA splicing.

# CRISPRs

- The particular stretches of unusual but characteristically organized sequence were noticed in the genomes of several bacteria which were named as of Clustered Regularly Interspaced Short Palindromic Repeats (or CRISPRs).

# CRISPRs



# CRISPRs

- They consist of **repeated sequences** (each 30 bp long and highly conserved within a given cluster) interleaved with **spacer sequences** of similar length but highly divergent sequence

# CRISPRs

- At one end of the array is a so-called leader sequence, often A-T rich and 500 bp in length.
- CRISPRs have been found in half of all bacterial genomes sequenced, and essentially all genomes of Archaea.

# CRISPRs

- In many cases, there is only one cluster per genome, but not uncommonly there are more and the number can range up to 20 or more - and in one case almost 400 were detected in a species of *Chloroflexus*.

# CRISPRs

- How do they arise, and what do they do?
- The first clue to their origin came from the striking observation that a significant number of the spacer sequences were identical to regions of known phage or plasmids.



# CRISPRs

- This quickly led to the proposal that these arrays are involved in some sort of defense mechanism against foreign nucleic acids entering the cell.

# CRISPRs

- Experimental support for this model came when resistant bacterial cells that arose in populations challenged with a given phage were found to have incorporated spacer sequences derived from that phage.

# CRISPRs

- Likewise, reduced sensitivity to a phage could be conferred or revoked by addition or removal of relevant spacer sequences.

# CRISPRs

- Furthermore, bacteria were increasingly insensitive to infection by a given phage the more spacer sequences they acquired from that phage.
- A set of conserved protein-coding genes is tightly associated with the CRISPR sequences.

# CRISPRs

- The two most highly conserved members (*cas1* and *cas2*, for “CRISPR associated”) are found in all cases, but other *cas* genes, and more distantly conserved genes, are less so.

# CRISPRs

- These genes encode proteins involved in different aspects of CRISPR function.
- Acquisition by a cell of spacer regions from a given phage confers decreased sensitivity to further infection by that phage.

# CRISPRs

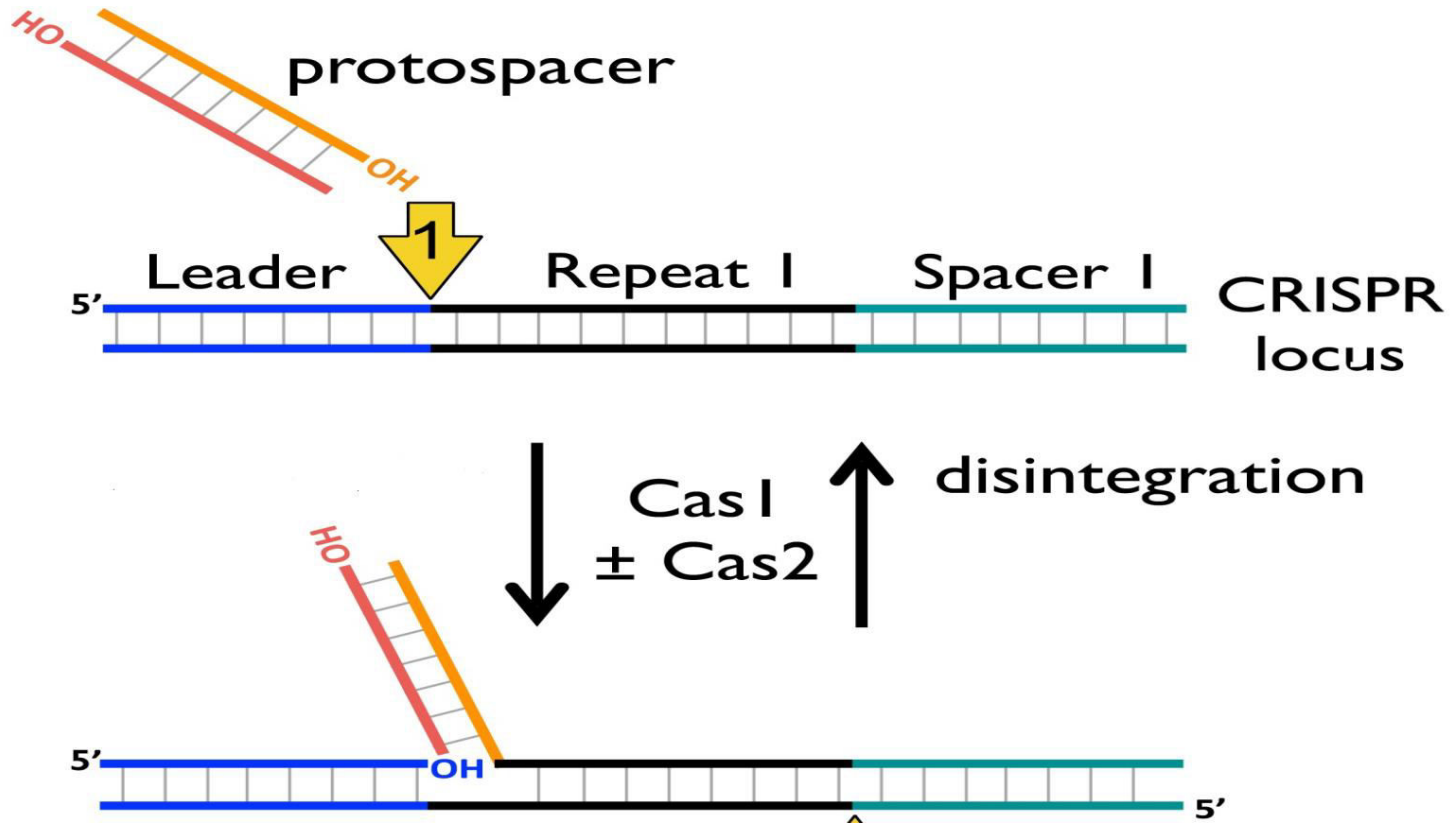
- The sequence in the virus that will become a new spacer is called the proto-spacer and is found close to a PAM (proto-spacer adjacent motif) sequence.



# CRISPRs

- When a new spacer is added to a CRISPR array, it is incorporated at the proximal end, near the leader sequence.
- Some of the *cas* genes encode proteins required for this acquisition process.

# CRISPRs



# CRISPRs

- Thus, the antiviral defense mechanism is not impaired by their absence, but the cell cannot acquire resistance to new viruses.
- The products of the *cas1*, *cas2*, and *cas4* genes fall into this category.

# CRISPRs

- *Cas1* is a putative integrase, whereas *Cas2* is a ribonuclease.
- In contrast, of other *Cas* proteins, *Cas6* is involved in expression and processing of the CRISPR cluster and *Cas3* in the interference of viral infection.

END

# Regulatory RNAs in Eukaryotes

- Eukaryotic regulatory RNAs come in many flavors characterized by their size (“long” or “short”), their origin, and the mechanisms by which they are generated and regulate gene expression.

# Regulatory RNAs in Eukaryotes

- It is now believed that between 30% and 70% of genes in higher eukaryotes are regulated to some extent by RNAs, with roles ranging from development to cellular homeostasis and protection of the cell from viruses and transposons.

# Regulatory RNAs in Eukaryotes

- Furthermore, one form of regulation viz., RNA interference (RNAi) has been adapted for use as a powerful experimental tool to manipulate gene expression in many organisms.



# Regulatory RNAs in Eukaryotes

- Several types of very short RNAs repress or silence the expression of genes with homology to those short RNAs.

# Regulatory RNAs in Eukaryotes

- Depending on the origin and context, these RNAs act by inhibiting translation of the mRNA, destruction of the mRNA, or even by transcriptional silencing of the promoter that directs expression of that mRNA.

# Regulatory RNAs in Eukaryotes

- These short RNAs are often generated by special enzymes from longer double-stranded RNAs (dsRNAs) of various origins.

# Regulatory RNAs in Eukaryotes

- The small RNAs have different names depending on their origin.
- Those made artificially or produced in vivo from dsRNA precursors are typically called **small interfering RNAs (SiRNAs)**.

# Regulatory RNAs in Eukaryotes

- Another group of regulatory RNAs is the **microRNAs (miRNAs)**.
- These miRNAs are derived from precursor RNAs that are encoded by genes expressed in cells where those miRNAs have specific regulatory functions.

# Regulatory RNAs in Eukaryotes

- A third class of short regulatory RNAs is the piwi-interaction RNAs (piRNAs), which are expressed predominantly in the germline and have features distinct from miRNAs.



# Regulatory RNAs in Eukaryotes

- Both siRNAs and miRNAs are generated from longer RNA molecules by the enzyme **Dicer**, an RNase III–like enzyme that recognizes and digests longer dsRNAs or the stem-loop structures formed by miRNA precursors.



# Regulatory RNAs in Eukaryotes

- The siRNA and miRNA products are typically 21–23 nucleotides long.
- The piRNAs (which are 24–34 nucleotides long) are derived in a manner that does not involve a dsRNA precursor.

# Regulatory RNAs in Eukaryotes

- Instead, the piRNAs are generated by processing long single-stranded transcripts covering so-called piRNA clusters found in the genome.
- This processing does not require Dicer.

# Regulatory RNAs in Eukaryotes

- These small RNAs inhibit expression of homologous target genes in three ways:-
- a) they trigger destruction of the mRNA encoded by the target gene,

# Regulatory RNAs in Eukaryotes

- b) they inhibit translation of the mRNA, or
- c) they induce chromatin modifications within the target gene and thereby silence its transcription.

# Regulatory RNAs in Eukaryotes

- Remarkably, whichever route is used in any given case, much of the same machinery is required.

# Regulatory RNAs in Eukaryotes

- This machinery includes a complex called the **RNA-induced silencing complex (RISC)**.
- A RISC contains, in addition to the small RNA, various proteins including a member of the **Argonaut** family.

# Regulatory RNAs in Eukaryotes

- The small RNA must be denatured to give a **guide RNA** - the strand that gives the RISC specificity and a **passenger RNA**, which usually gets discarded.



# Regulatory RNAs in Eukaryotes

- The resulting complex, the mature RISC, is then directed to target RNAs containing sequences complementary to the guide RNA.
- These target RNAs are degraded or their translation is inhibited.

# Regulatory RNAs in Eukaryotes

- Typically, the choice depends in part on how closely the guide RNA matches the target mRNA:-
- if the sequences are highly complementary (as is usually the case with siRNAs), the target is degraded;

# Regulatory RNAs in Eukaryotes

- if the match is not as good (i.e., if there are several base-pairing mismatches, as is often the case with miRNAs), the response is more often inhibition of translation.

# Regulatory RNAs in Eukaryotes

- In those cases in which the target RNA is degraded, Argonaute is the catalytic subunit that performs the initial mRNA cleavage; for this reason, Argonaute is often called **Slicer** and mRNA cleavage is called **slicing**.

# Regulatory RNAs in Eukaryotes

- A RISC can also be directed into the nucleus, where it recruits other proteins that modify the chromatin around the promoter of the gene complementary to the guide RNA. This modification leads to silencing of transcription.

# Regulatory RNAs in Eukaryotes

- A distinction worth making between miRNAs and siRNAs is that the former act like traditional trans-acting regulators: they are encoded by a gene but act on other genes.



# Regulatory RNAs in Eukaryotes

- In contrast, siRNAs are typically generated by transcripts of the regions on which they act and are thus described as working in *cis*.

END



# Promoters

- A gene consists of a transcriptional region and a regulatory region.
- The regulatory region can be divided into *cis*-regulatory (or *cis*-acting) elements and *trans*-regulatory (or *trans*-acting) elements.

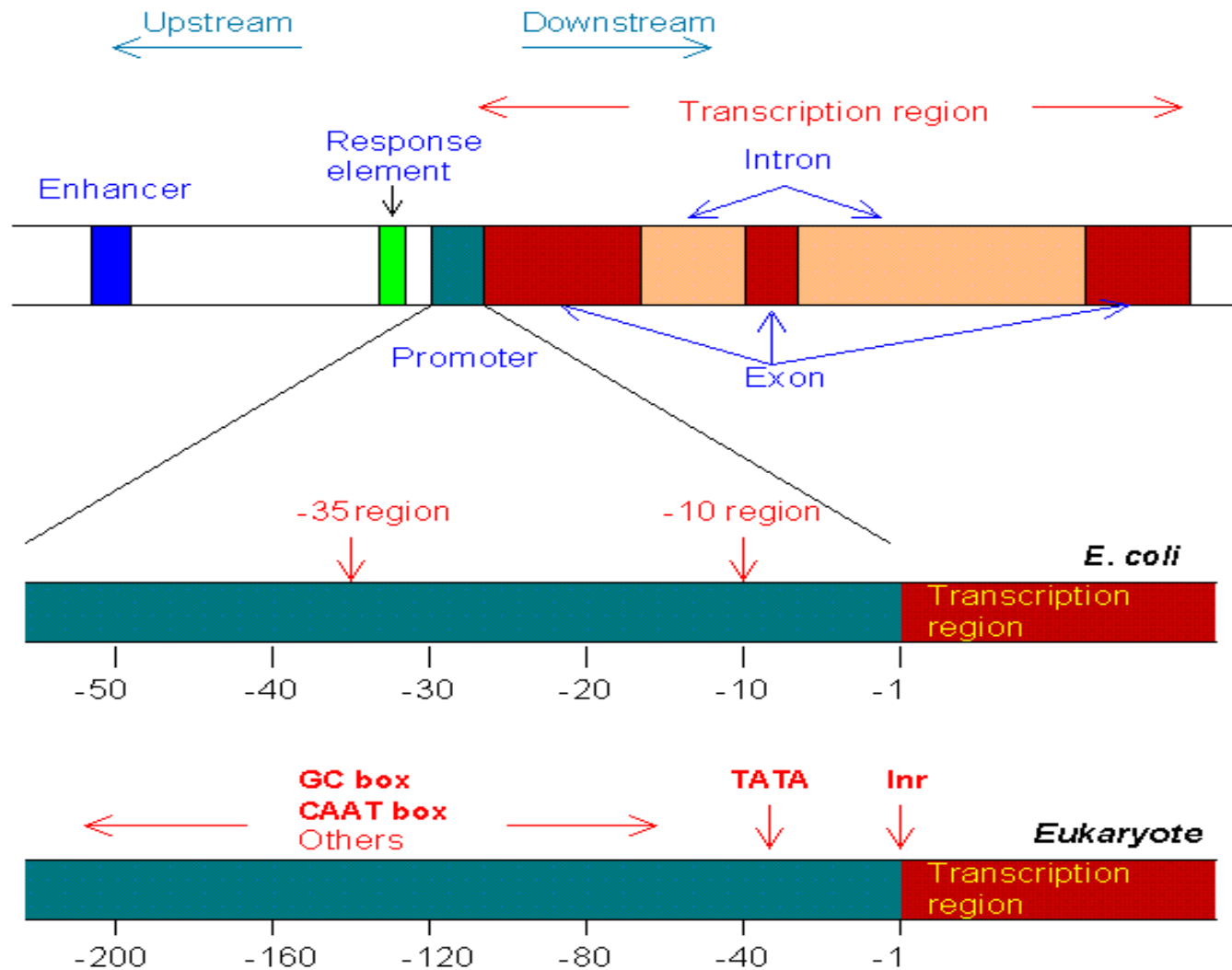
# Promoters

- The *cis*-regulatory elements are the binding sites of transcription factors which are the proteins that, upon binding with *cis*-regulatory elements, can affect (either enhance or repress) transcription.

# Promoters

- The cis-acting elements may be divided into the following four types:
- Promoter
- Enhancer
- Silencer
- Response element

# Promoters



# Promoters

- Three classes of promoters can be distinguished on the basis of their relative locations.
- Class II Promoters
- Class I Promoters
- Class III Promoters

# Promoters

- **Class II Promoter**
- The promoters recognized by RNA polymerase II are called class II promoters because these are the most complex and best studied.

# Promoters

- Class II promoters can be considered as having two parts: the **core promoter** and the **proximal promoter**.



# Promoters

- The core promoter attracts general transcription factors and RNA polymerase II at a basal level and sets the transcription start site and direction of transcription.

# Promoters

- The proximal promoter helps attract general transcription factors and RNA polymerase and includes promoter elements that can extend from about 37 bp up to 250 bp upstream of the transcription start site.

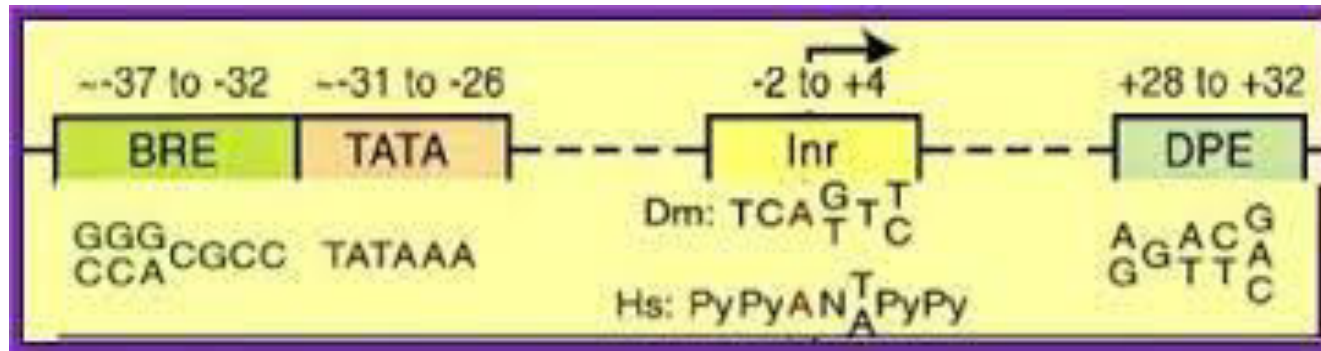
# Promoters

- The core promoter is modular and can contain almost any combination of the following elements.
- The TATA box is centered at approximately position 228 and has the consensus sequence TATA(A/T)AA(G/A);

# Promoters

- the **TFIIB recognition element** (BRE) lies just upstream of the TATA box and has the consensus sequence (G/C)(G/C)(G/A)CGCC; the **initiator (Inr)** is centered on the transcription start site, and the **downstream core element** (DCE) having three parts.

# Promoters



# Promoters

- **Class I Promoter**
- We can refer to this promoter in the singular because almost all species have only one kind of gene recognized by polymerase I: the rRNA precursor gene.

# Promoters

- The promoter has two critical regions, one of which is the **core element**, also known as the **initiator (rINR)**, is located at the start of transcription, between positions - 45 and +20.
- The other is the **upstream promoter element (UPE)**, located between positions -156 and - 107.



# Promoters

- **Class III Promoters**
- RNA polymerase III transcribes a variety of genes that encode small RNAs. These include:-
  - 1) the “classical” class III genes, including the 5S rRNA and tRNA genes; and

# Promoters

- 2) some relatively recently discovered class III genes, including the U6 snRNA gene, the 7SL RNA gene, the 7SK RNA gene, and the Epstein – Barr virus EBER2 gene.

# Promoters

- By contrast, the “classical” class III genes have promoters located entirely within the genes themselves.

END

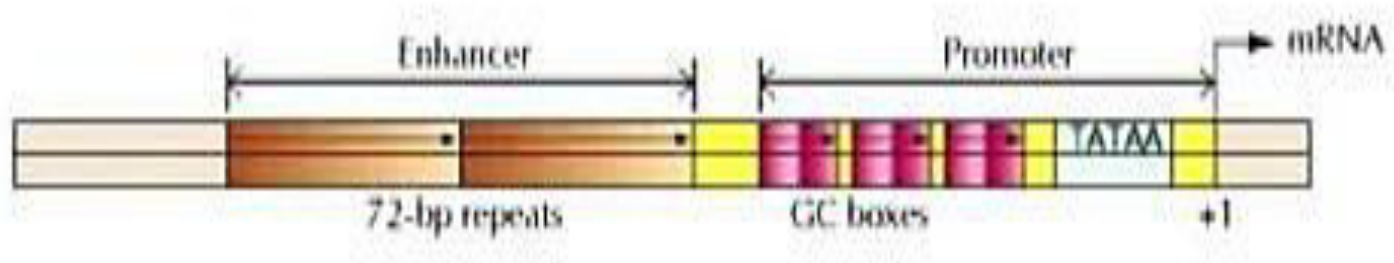
# Enhancers & Silencers

- Enhancers are elements that stimulate transcription while Silencers, by contrast, depress transcription.
- Chambon and colleagues discovered the first enhancer in the 5'-flanking region of the SV40 early gene.

# Enhancers & Silencers

- This DNA region contains a conspicuous duplication of a 72-bp sequence, called the *72-bp repeat*.
- When Benoist and Chambon made deletion mutations in this region, they observed profoundly depressed transcription in vivo.

# Enhancers & Silencers



The SV40 enhancer

# Enhancers & Silencers

- This behaviour suggested that the 72-bp repeats constituted another upstream promoter element.



# Enhancers & Silencers

- However, Paul Berg and his colleagues discovered that the 72-bp repeats still stimulated transcription even if they were inverted or moved all the way around to the opposite side of the circular SV40 genome, over 2 kb away from the promoter.

# Enhancers & Silencers

- The latter behaviour, at least, is very unpromoter-like.
- Thus, such orientation- and position independent DNA elements are called enhancers to distinguish them from promoter elements.

# Enhancers & Silencers

- Enhancers act through proteins that bind to them.
- These proteins have several names: transcription factors, enhancer-binding proteins, or activators.

# Enhancers & Silencers

- These proteins appear to stimulate transcription by interacting with general transcription factors at the promoter.
- This interaction promotes formation of a pre-initiation complex, which is necessary for transcription.

# Enhancers & Silencers

- Thus, enhancers usually allow a gene to be induced (or sometimes repressed) by activators that frequently require help from other molecules (e.g., hormones and co-activator proteins) to exert their effects.

# Enhancers & Silencers

- Enhancers are not the only DNA elements that can act at a distance to modulate transcription. Silencers also do this, but - as their name implies - they inhibit rather than stimulate transcription.



# Enhancers & Silencers

- Yeast chromosome III contains three loci of very similar sequence: MAT, HML, and HMR.
- Though MAT is expressed, the other two loci are not, and silencers located at least 1 kb away seem to be responsible for this genetic inactivity.



# Enhancers & Silencers

- They seem to be responding to an external negative influence: a silencer.
- How do silencers work?

# Enhancers & Silencers

- The available data indicate that they cause the chromatin to coil up into a condensed, inaccessible, and therefore inactive form, thereby preventing transcription of neighbouring genes.

# Enhancers & Silencers

- Sometimes the same DNA element can have both enhancer and silencer activity, depending on the protein bound to it.

# Enhancers & Silencers

- For example, the thyroid hormone response element acts as a silencer when the thyroid hormone receptor binds to it without its ligand, thyroid hormone.

# Enhancers & Silencers

- But it acts as an enhancer when the thyroid hormone receptor binds along with thyroid hormone.

END

# Monocistronic or Polycistronic mRNA

- In a bacterial cell, transcription and translation can proceed simultaneously because both these processes take place in the same cellular compartment so the protein synthetic machinery can start reading 5' end of mRNA before the 3' end is formed.

# Monocistronic or Polycistronic mRNA

- While in eukaryotic cells, both these processes occur in different compartments.
- So primary transcript is converted to mature mRNA before the protein synthesis starts in the cytoplasm.



# Monocistronic or Polycistronic mRNA

- The segment of mRNA that codes for a polypeptide chain is called an **open reading frame (ORF)** because the protein synthesis starts at a specific start codon and stops at a specific termination/stop codon.

# Monocistronic or Polycistronic mRNA

- A DNA segment corresponding to an ORF plus the translational start and stop signals for protein synthesis is called a cistron and the mRNA encoding a single polypeptide is called *monocistronic mRNA*.

# Monocistronic or Polycistronic mRNA

- Although the terms *cistron* and *gene* are sometimes used interchangeably to describe bacterial DNA segments that specify polypeptides,

# Monocistronic or Polycistronic mRNA

- the term gene has a broader meaning because it also includes the promoter region and applies to DNA segment that code for RNA molecules such as tRNA and rRNA that are not translated.

# Monocistronic or Polycistronic mRNA

- Bacterial mRNA molecules often contain two or more cistrons, thus are *polycistronic* and are actually more common than bacterial monocistronic mRNA molecules.

# Monocistronic or Polycistronic mRNA

- Each cistron within a polycistronic mRNA specifies a specific polypeptide chain.
- Furthermore, cistrons contained in polycistronic mRNA often specify proteins for a single metabolic pathway.



# Monocistronic or Polycistronic mRNA

- For example, one *Escherichia coli* mRNA has eight cistrons, each coding for a different enzyme required for histidine synthesis.
- Using polycistronic mRNA is a way for the cell to regulate synthesis of related protein coordinately.



# Monocistronic or Polycistronic mRNA

- With a polycistronic mRNA molecule, the synthesis of several related proteins – in similar quantities and at same time – can be regulated by a single signal.

# Monocistronic or Polycistronic mRNA

- In addition to reading frames and start and stop codons, other regions in mRNA are also significant.
- For example mRNA translation rarely starts exactly at the 5' end and stops exactly at 3' end .

# Monocistronic or Polycistronic mRNA

- Instead, initiation of synthesis of the first polypeptide chain of a polycistronic mRNA may begin hundreds of nucleotides from 5'-RNA terminus.

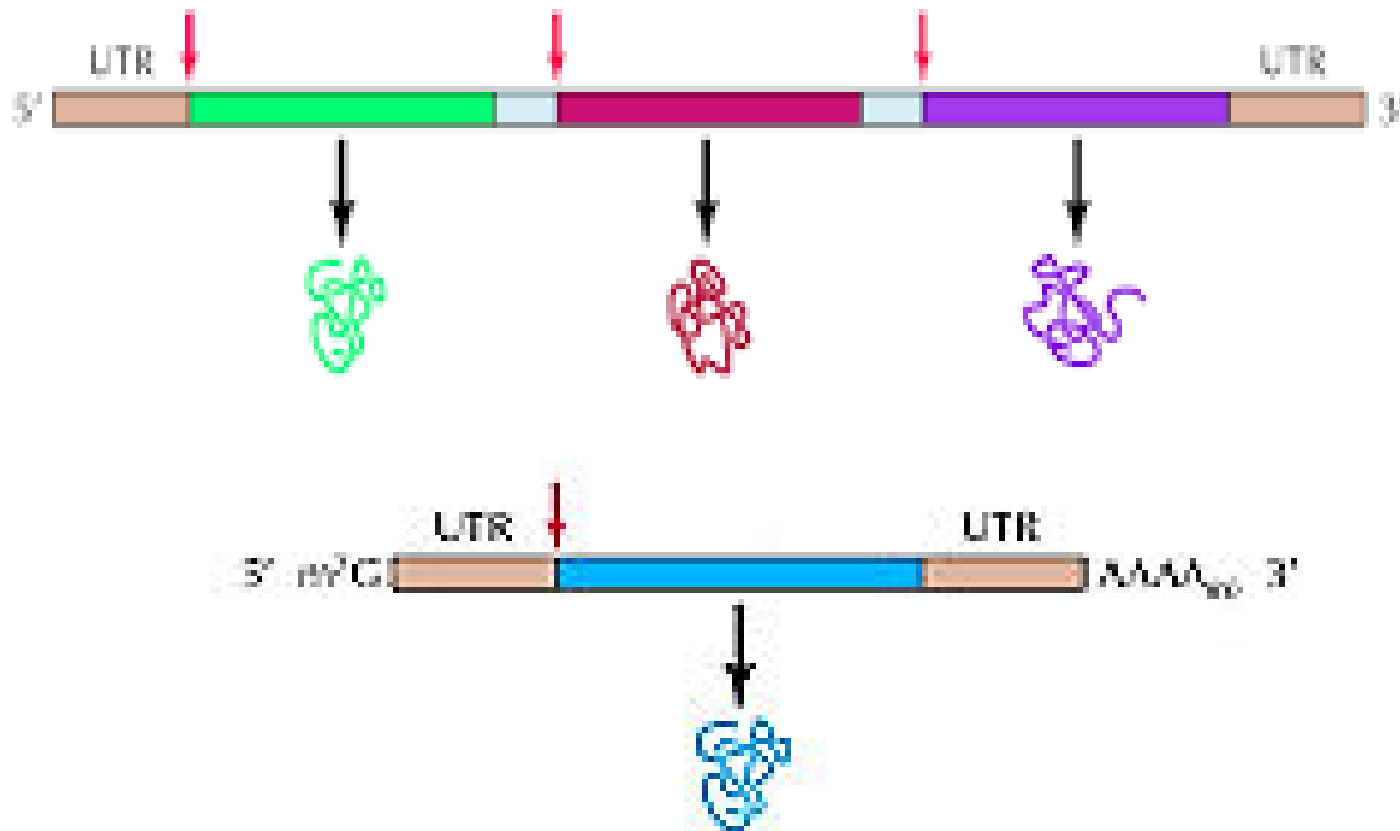
# Monocistronic or Polycistronic mRNA

- The untranslated RNA sequence before the coding region is called the 5' – untranslated region (5'- UTR) or 5' – leader while the untranslated sequence after the coding region is called 3'– untranslated region (3'- UTR).

# Monocistronic or Polycistronic mRNA

- Polycistronic mRNA molecules also usually contain intercistronic sequences (**spacers**) that are tens of bases long.

# Monocistronic or Polycistronic mRNA



# mRNA has a short life span

- An important characteristic of bacterial mRNA is that its lifetime is short compared to other types of bacterial RNA molecules.
- The half life of a typical bacterial mRNA is a few minutes.



# mRNA has a short life span

- Although mRNA's short lifetime may seem wasteful, it has an important regulatory function.
- A cell can turn off the synthesis of a protein that is no longer needed by turning off synthesis of mRNA that encodes the protein.

# mRNA has a short life span

- Soon after, none of that particular mRNA will remain and protein synthesis will cease.
- Of course, bacterial cells save energy by not being forced to synthesize proteins that they no longer need.

# mRNA has a short life span

- The particular mRNA and their life span can be analyzed by the pulse-chase experiment.
- Bacteria are briefly cultured in a medium that contains a radioactive precursor for RNA such as [ $^3\text{H}$ ] uridine.

# mRNA has a short life span

- Then the bacteria are switched to a medium that contains a high concentration of non radioactive uridine while no [ $^3\text{H}$ ] uridine and samples are removed at specific times for analyses.

# mRNA has a short life span

- The RNA are isolated and different types are separated by gel electrophoresis and detected by their radioactivity.

# mRNA has a short life span

- A typical radioactive mRNA molecule will decrease with a half life of a few minutes, whereas radioactive rRNA and tRNA will remain through many generations.

# mRNA has a short life span

- However, bacteria also contain some longlived mRNA molecules which can not be analyzed by this technique.

END



# Controlling the rate of mRNA synthesis

- Bacterial cells can control gene expression by regulating the rate at which specific genes are transcribed.
- They vary the promoter sequence so that RNA polymerase initiates transcription at some genes more efficiently than at others.

# Controlling the rate of mRNA synthesis

- However, this mechanism is not ideally suited for all situations.
- For example, the products of many genes are needed under some physiological conditions but not others, so the products of such genes must be synthesized only when circumstances demand it.

# Controlling the rate of mRNA synthesis

- Bacterial cells increase their efficiency by:-
- 1) Selecting the catabolic pathway for energy production that yields the greatest amount of energy per unit time, and
- 2) synthesizing molecules only as the need arises.

# Controlling the rate of mRNA synthesis

- Bacteria accomplish both of these objectives by turning on the transcription of specific genes when their products are needed and turning off their transcription when their products are not needed.

# Controlling the rate of mRNA synthesis

- Actually, there are no known examples of switching a system completely off.
- When transcription is in the “off” state, there always remains a basal level of gene expression.

# Controlling the rate of mRNA synthesis

- So the term “*off*” means *very slow*.
- There are also some examples in which activity is switched from fully on to partly on rather than off.



# Controlling the rate of mRNA synthesis

- In bacterial systems in which several enzymes act in sequence in a single metabolic pathway, it is often the case that either all of these enzymes are present or all are absent.



# Controlling the rate of mRNA synthesis

- This phenomenon, which is called **coordinate regulation**, results from control of the synthesis of a single polycistronic mRNA that encodes all of the gene products.
- There are several mechanisms for such regulation.

# Controlling the rate of mRNA synthesis

- The molecular mechanism of mRNA regulation can be divided into two major categories:-
  - a) negative regulation
  - b) positive regulation

# Controlling the rate of mRNA synthesis

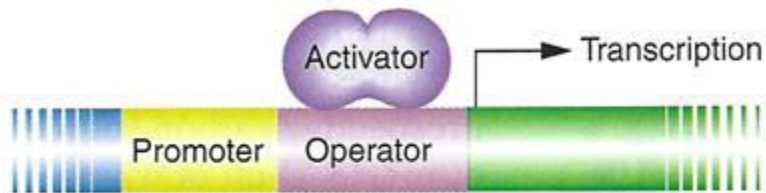
- In negative regulation, a **repressor** turns off the transcription of one or more gene.
- In positive regulation, an **activator** turns on the transcription of one or more gene.

# Controlling the rate of mRNA synthesis

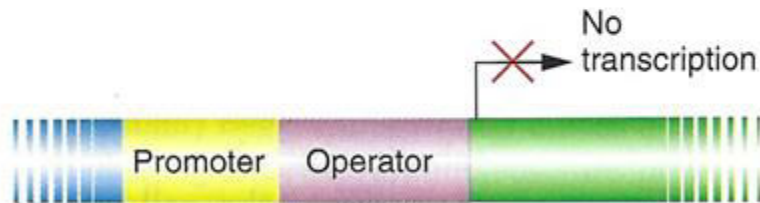
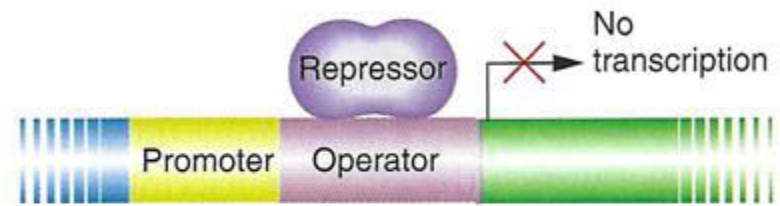
- Negative regulation and positive regulation are not mutually exclusive.
- Many genes respond to both types of regulation.

# Controlling the rate of mRNA synthesis

Positive regulation



Negative regulation



(No activator)



(No repressor)

# Lactose Operon

- In *E. coli*, two proteins are necessary for lactose metabolism.
- These include the enzyme  $\beta$ -galactosidase and a carrier molecule, lactose permease, which transports lactose (and other galactosides) into the cell.



# Lactose Operon

- The existence of the two proteins was first shown by a combination of genetic experiments and biochemical analysis.
- First, hundreds of Lac<sup>-</sup> mutants (unable to use lactose as a carbon source) were isolated.



# Lactose Operon

- By genetic manipulation, some of these mutations were moved from the *E. coli* chromosome to an *F'**lac* plasmid (a plasmid carrying the genes for lactose utilization).

# Lactose Operon

- Then partial diploids having the genotypes  $F'/lac^- / lac^+$  or  $F'/lac^+ / lac^-$  were constructed.
- It was observed that these diploids always have a  $Lac^+$  phenotype (can make  $\beta$ -galactosidase).

# Lactose Operon

- This shows that none of the *lac*<sup>-</sup> mutants make an inhibitor that blocks *lac* gene function.
- Partial diploids were also constructed in which both the chromosome and the F'/*lac* plasmid were *lac*<sup>-</sup>.

# Lactose Operon

- Using different pairs of  $lac^-$  mutants, some pairs were observed to have a  $Lac^+$  phenotype while others were observed to have a  $Lac^-$  phenotype.

# Lactose Operon

- This complementation test showed that all of the mutants initially isolated fell into two groups which were called *lacZ* and *lacY*.

# Lactose Operon

- The existence of two complementation groups was good evidence that there are at least two genes in the *lac* system.
- The *lacZ* gene is the structural gene for  $\beta$ -galactosidase.

# Lactose Operon

- This enzyme catalyzes the hydrolysis of o-nitrophenyl- $\beta$ -galactoside (a lactose analog) and results in the production of o-nitrophenoxide which is a yellow product.



# Lactose Operon

- The function of the *lacY* gene product as lactose permease was strongly suggested by experiments that showed that *lacZ*<sup>+ *lacY*<sup>-</sup> cells can not transport lactose into the cell while *lacY*<sup>-</sup> *lacZ*<sup>+</sup> cells can do so.</sup>

# Lactose Operon

- Investigators discovered a third gene, lacA, the product of which is a  $\beta$ -galactoside transacetylase, which transfers an acetyl group from acetyl-CoA to lactose analog.

# Lactose Operon

- The reason that *lacA* was not detected at the same time as *lacZ* and *lacY* is that its gene product, transacetylase, is not required for lactose catabolism.

# Lactose Operon

- The precise role of *lacA* is still a matter of supposition.
- One hypothesis is that transacetylase detoxifies lactose analogs that would harm cells.

END

# *lac* structural genes are regulated

- When *E. coli* with a *lac*<sup>+</sup> genotype is cultured in a lactose-free medium, the intracellular concentration of  $\beta$ -galactosidase, permease and tranacetylase are exceedingly low.

# ***lac* structural genes are regulated**

- When lactose is added to the growth medium, however, the concentration of these protein increases simultaneously to about  $10^5$  molecules per cell (about 1% of total cellular protein).



# ***lac* structural genes are regulated**

- Furthermore, lactose addition triggers the synthesis of *lac* mRNA as evidenced in studies in which mRNA, labeled with [<sup>32</sup>P] phosphate at various times after lactose addition, is hybridized to DNA that carries *lac* genes.



# *lac* structural genes are regulated

- Enzymes such as  $\beta$ -galactosidase, permease and tranacetylase are said to be **inducible enzymes** because their rate of synthesis increases in response to the addition of a small molecule (lactose) to the medium.

# *lac* structural genes are regulated

- Other enzymes, called **repressible enzymes**, exhibit a decreased rate of synthesis in response to the addition of a small molecule in the medium.

# ***lac* structural genes are regulated**

- For instance, the addition of tryptophan to the growth medium causes *E. coli* to greatly decrease the rate at which it produces enzymes needed for tryptophan synthesis.

# *lac* structural genes are regulated

- Still other enzymes, called **constitutive enzyme**, are synthesized at fixed rates under all growth conditions.

# ***lac* structural genes are regulated**

- **Constitutive enzyme** usually perform basic cellular “housekeeping” functions needed for normal cell maintenance.

# ***lac* structural genes are regulated**

- Lactose is rarely used in experiments to study induction of the lactose enzymes because the  $\beta$ -galactosidase that is synthesized catalyzes lactose cleavage.

# ***lac* structural genes are regulated**

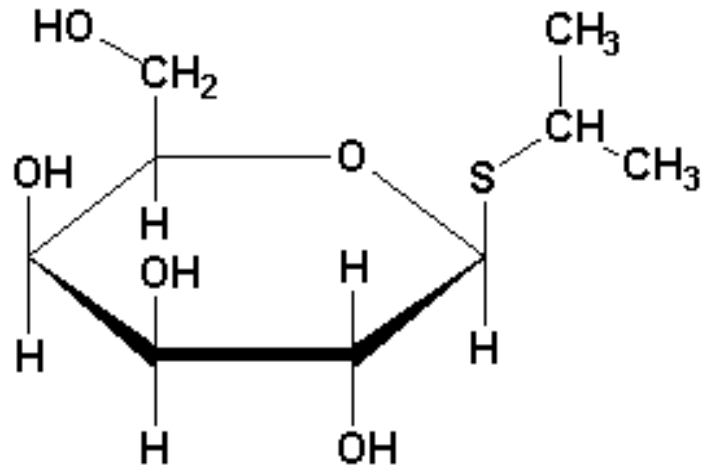
- As a result, the lactose concentration continually decreases, which complicates the analysis of kinetic experiments.



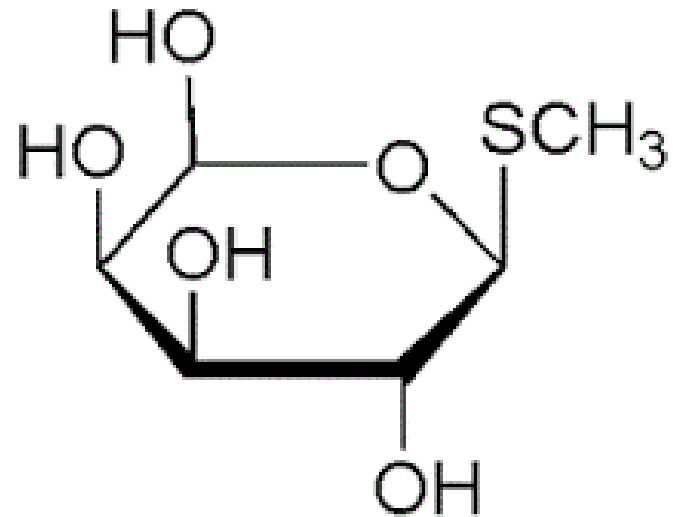
# ***lac* structural genes are regulated**

- Instead, two sulfur-containing lactose analogs, isopropylthiogalactoside (IPTG) and thiomethylgalactoside (TMG) are used.

# *lac* structural genes are regulated



**Isopropyl Thiogalactoside (IPTG)**



# *lac* structural genes are regulated

- These are effective inducers without being substrates of  $\beta$ -galactosidase.
- Inducers having this property are called gratuitous inducers.

END

# Regulation of *lac* mRNA

- Two French investigators Jacques Monod and Francois Jacob performed a series of genetic and biochemical experiments in the late 1950's that elucidated the mechanism of regulation of *lac* system.

# Regulation of *lac* mRNA

- The began by isolating constitutive *E. coli* mutants that make *lac* mRNA (make all the three enzymes) in the presence as well as absence of the inducer.

# Regulation of *lac* mRNA

- Then they constructed a variety of partial diploid cells containing constitutive mutants and observed the cell's ability to synthesize  $\beta$ -galactosidase.

# Regulation of *lac* mRNA

- These mutations appeared to be of two types, termed *lacI* and *lacO<sup>c</sup>*.
- The *lacI*<sup>-</sup> mutations behaved like typical minus mutations in most genes and are recessive.



# Regulation of *lac* mRNA

- Because *lac* mRNA synthesis is off in a *lacI*<sup>+</sup> cell and is on in a *lacI*<sup>-</sup> mutant, the *lacI* gene is apparently a regulatory gene.

# Regulation of *lac* mRNA

- The *lacI* gene codes for a product that acts as an inhibitor to keep the *lac* structural genes turned off.
- The *lacI*<sup>-</sup> mutant lacks the inhibitor and thus is constitutive.

# Regulation of *lac* mRNA

- A  $lacI^+ / lacI^-$  partial diploid has one good copy of the *lacI* gene product, so the system is inhibited.
- Monod & Jacob called the *lacI*-gene product, the **Lac repressor**.

# Regulation of *lac* mRNA

- The most likely explanation for the constitutive lactose system is that the mutant strain synthesizes a truncated repressor protein that can't block transcription of the *lac* genes.

# Regulation of *lac* mRNA

- This conclusion was confirmed when the Lac repressor was purified and characterized.
- Genetic mapping experiments placed the *lacI* gene adjacent to the *lacZ* gene and established the order of the genes as *lacI lacZ lacY lacA*.

# Regulation of *lac* mRNA

- A striking property of the *lacO<sup>c</sup>* mutations is that in certain cases, they are dominant.
- The significance of the dominance of the *lacO<sup>c</sup>* mutations becomes clear from the properties of the partial diploids.



# Regulation of *lac* mRNA

- Both the diploids  $lacO^c lacZ^- / lacO^+ lacZ^+$  and  $lacO^c lacZ^+ / lacO^+ lacZ^-$  are  $Lac^+$ , because there is a functional *lacZ* gene.
- In case of former,  $\beta$ -galactosidase synthesis is inducible even though a  $lacO^c$  mutation is present.



# Regulation of *lac* mRNA

- In case of former,  $\beta$ -galactosidase synthesis is inducible even though a  $\text{lacO}^c$  mutation is present.

# Regulation of *lac* mRNA

- An immunological test capable of detecting a mutant  $\beta$ -galactosidase showed that the mutant enzyme is synthesized constitutively in a  $lacO^c lacZ^- / lacO^+ lacZ^+$  partial diploid,

# Regulation of *lac* mRNA

- whereas the wild type enzyme is synthesized only if an inducer is added.

# Regulation of *lac* mRNA

- This experiment takes advantage of the fact that a purified antibody to  $\beta$ -galactosidase will also react with mutant protein as long as the structural differences between wild and mutant proteins are not too great.

# Regulation of *lac* mRNA

- A reaction of this type in which an antibody that is raised in response to one protein is used to detect a closely related protein is called a **cross-reaction** and closely related protein is called **cross-reacting material (CRM)**.

# Regulation of *lac* mRNA

- Thus the presence of CRM, which can be detected by a variety of standard immunological procedures, is indicative of the presence of mutant protein.

# Regulation of *lac* mRNA

- Genetic mapping experiments showed that all the *lacO<sup>c</sup>* mutations are located between genes *lacI* and *lacZ*, so the gene order of the five elements of the *lac* system is *lacI lacO lacZ lacY lacA*.



# Regulation of *lac* mRNA

- Together these experiments lead to the conclusion that *lacO<sup>c</sup>* mutations define a site or a noncoding region of the DNA rather than a gene and that the *lacO* region determines whether synthesis of the

# Regulation of *lac* mRNA

- product of the adjacent *lacZ* gene is inducible or constitutive.
- The *lacO* site is called the **operator**.

END

# The operon model

- Monod and Jacob proposed the **operon model** in 1961 to explain how the lac system is regulated.

# The operon model

- The term operon refers to two or more contiguous genes and the genetic elements that regulate their transcription in a coordinate fashion.

# The operon model

- Promoters had not yet been discovered when Monod & Jacob proposed the operon model but were readily incorporated into the operon model after their discovery.
- The five major features of the model are:-

# The operon model

- 1. The products of the *lacZ*, *lacY* and *lacA* genes are encoded in a single polycistronic *lac* mRNA molecule.
- 2. The promoter for this mRNA molecule is immediately adjacent to the *lac* region.

# The operon model

- Promoter mutations (p-) that are completely incapable of making  $\beta$ -galactosidase, permease, and transacetylase have been isolated. The promoter is located between *lacI* and *lacO*.



# The operon model

- 3. The operator is a sequence of bases (in the DNA) to which the repressor protein binds.
- 4. When the repressor protein is bound to the operator, *lac* mRNA transcription can't take place.

# The operon model

- 5. Inducers stimulate lac mRNA synthesis by binding to the repressor. This binding alters the repressor's conformation so it can't bind to the operator. In the presence of an inducer, therefore,

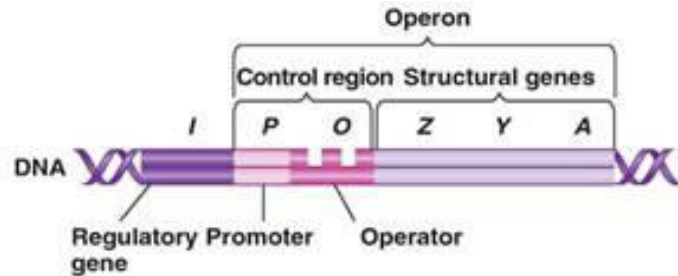
# The operon model

- The operator is unoccupied and the promoter is available for initiation of mRNA synthesis. This state is called **derepression**.

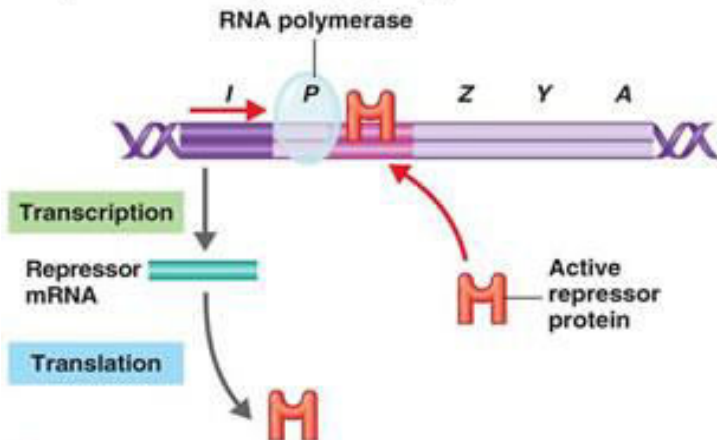
# The operon model

- This simple model explains many of the features of the *lac* system and of other negatively regulated genetic systems.
- However, *lac* operon is also subjected to positive regulation.

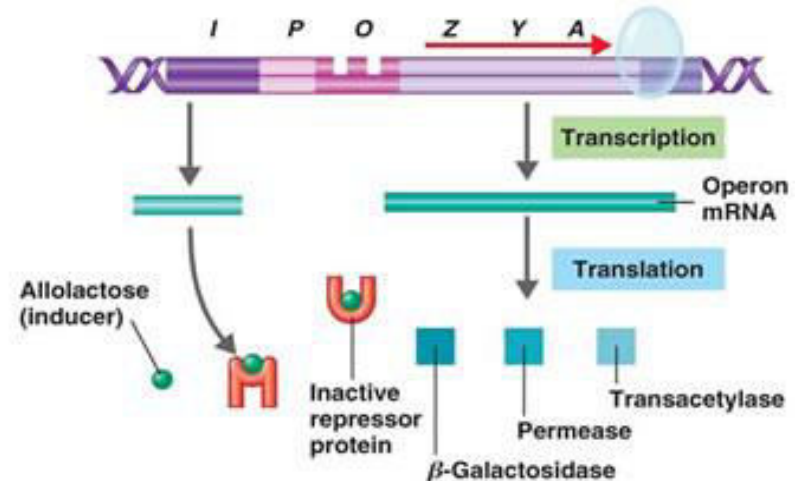
# The operon model



- 1 Structure of the operon.** The operon consists of the promoter (*P*) and operator (*O*) sites and structural genes that code for the protein. The operon is regulated by the product of the regulatory gene (*I*).



- 2 Repressor active, operon off.** The repressor protein binds with the operator, preventing transcription from the operon.



- 3 Repressor inactive, operon on.** When the inducer allolactose binds to the repressor protein, the inactivated repressor can no longer block transcription. The structural genes are transcribed, ultimately resulting in the production of the enzymes needed for lactose catabolism.

# Allolactose; the inducer of lactose operon

- Two related problems became evident as the operon model was tested.



# Allolactose; the inducer of lactose operon

- First, inducers must enter a cell if they are to bind to repressor molecules , yet lactose transport requires permease, and permease synthesis requires induction.



# Allolactose; the inducer of lactose operon

- Thus, it must be explained how the inducer gets into the cell in the first place.

# Allolactose; the inducer of lactose operon

- Second, the isolated Lac repressor does not bind lactose but does bind a lactose isomer called **allolactose**.
- Remarkably,  $\beta$ -galactosidase also converts a small proportion of lactose to allolactose.

# Allolactose; the inducer of lactose operon

- Therefore, induction of the synthesis of  $\beta$ -galactosidase by lactose requires that  $\beta$ -galactosidase be present.

# Allolactose; the inducer of lactose operon

- Both problems are solved in the same way; in the uninduced state, a small amount of *Lac* mRNA is synthesized (roughly one mRNA molecule per cell per generation).

# Allolactose; the inducer of lactose operon

- This synthesis, called the **basal synthesis**, occurs because the binding of the repressor to the operator is never infinitely strong.

# Allolactose; the inducer of lactose operon

- Thus, even though the repressor binds tightly to the operator, it occasionally comes off and an RNA polymerase molecule can initiate transcription during the instant that the operator is free.

# Allolactose; the inducer of lactose operon

- When lactose is added to a *Lac*<sup>+</sup> culture growing in lactose free medium, the few permease molecules transport a few lactose molecules into the cell and few  $\beta$ -galactosidase molecules convert some of lactoses to allolactose.



# Allolactose; the inducer of lactose operon

- An allolactose molecule then binds to a repressor molecule that is sitting on the operator, and the repressor is inactivated and falls off the operator.

# Allolactose; the inducer of lactose operon

- Synthesis of *Lac* mRNA begins and these mRNA molecules are translated to produce hundreds of  $\beta$ -galactosidase and permease molecules which allow lactose molecules to enter into the cells.

# Allolactose; the inducer of lactose operon

- Most of the lactose molecules are cleaved to yield glucose and galactose, but some are converted to allolactose molecules, which binds to and inactivate all of the intracellular repressor molecules.

# Allolactose; the inducer of lactose operon

- Thus lac mRNA is synthesized at high rate and the permease and  $\beta$ -galactosidase concentrations become quite high.
- The glucose produced is used as a carbon source for energy.

# Allolactose; the inducer of lactose operon

- The glucose produced by cleavage reaction is used as a carbon source for energy while the galactose produced is converted to glucose-1-phosphate by a set of enzymes which are also inducible.

# Allolactose; the inducer of lactose operon

- Ultimately, all of the lactose in the growth medium and within the cells is consumed.
- Then the allolactose concentration drops so that there is not sufficient allolactose to bind to a repressor.



# Allolactose; the inducer of lactose operon

- The repressor binds to the operator, re-establishing repression and thereby blocking further synthesis of *lac* mRNA.
- In bacteria, most mRNA molecules have half life of only a few minutes.



# Allolactose; the inducer of lactose operon

- Hence, in less than one generation, there is little remaining *lac* mRNA and synthesis of  $\beta$ -galactosidase and permease ceases.
- These proteins are quite stable but are gradually diluted as the cell divide.

# Allolactose; the inducer of lactose operon

- Note that if lactose were added again one generation after the original lactose had been depleted, cleavage of lactose would begin immediately as the cells would already have adequate  $\beta$ -galactosidase and permease.

END

# The *Lac* repressor

- Walter Gilbert and Benno Muller-Hill successfully isolated the *Lac* Repressor from *E. coli* extracts in 1966.
- The *Lac* repressor is a homotetramer with a molecular mass of 154 kDa.

# The *Lac* repressor

- Each subunit of the *Lac* repressor is made up of 360 amino acids and can bind one molecule of IPTG.
- Crude cell extracts bind about 20-40 molecules of IPTG per cell, so there are roughly 5-10 repressor molecules per cell.

# The *Lac* repressor

- The specific binding of repressor to the operator sequence and the inhibition of this binding by an inducer have been demonstrated by using purified repressor.

# The *Lac* repressor

- An important procedure for studying repressor-operator binding is the nitrocellulose filter assay.
- Protein stick to these filters but DNA does not.



# The *Lac* repressor

- If a mixture of repressor and radioactive *lac* DNA is passed through such a filter, radioactivity will be retained on the filter if the protein and the *lac* DNA form a complex.



# The *Lac* repressor

- The results indicate that the *Lac* repressor binds to DNA with a normal *lac* operator but fails to bind to DNA with a *lacO<sup>c</sup>* mutant operator.

END

# The *Lac* repressor

- Furthermore, IPTG prevents the *Lac* repressor from binding to DNA with a normal *lac* operator.
- These studies confirm the major predictions of the operon model.

# The *Lac* repressor

- After the *lac* operon was sequenced, it eventually became clear that it has two additional operators.
- The original operator with its centre of symmetry at position +11 is now designated as *lacO*<sub>1</sub>.

# The *Lac* repressor

- Auxiliary operators *lacO*<sub>2</sub> and *lacO*<sub>3</sub> have their centers of symmetry at positions +412 and -82, respectively.
- Thus, *lacO*<sub>3</sub> is upstream of the *lac* promoter while *lacO*<sub>2</sub> is located in *lacZ*.

# The *Lac* repressor

- The discoveries of *lacO*<sub>2</sub> and *lacO*<sub>3</sub> led investigators to ask whether these auxiliary operators participate in the *lac* operon regulation.

# The *Lac* repressor

- Muller-Hill and coworkers performed series of genetic experiments in which they altered one or more *lac* operators and then determined the alteration's effect on repression.

# The *Lac* repressor

- Point mutations in *lacO*<sub>1</sub> caused a 5 – 50 folds decrease in repression, but some repression was still observed.



# The *Lac* repressor

- Destruction of either *lacO*<sub>2</sub> or *lacO*<sub>3</sub> caused a twofold decrease in repression and destruction of both auxiliary operators caused a 70-fold decrease in repression.

# The *Lac* repressor

- The ability to repress the *lac* system was completely lost in cells in which *lacO*<sub>1</sub> and either of the auxiliary operators were destroyed.

# The *Lac* repressor

- These results indicate that *lacO*<sub>1</sub> play a major role in repression but the two auxiliary operators also make important contributions.

END

# The *Lac* repressor

- Furthermore, IPTG prevents the *Lac* repressor from binding to DNA with a normal *lac* operator.
- These studies confirm the major predictions of the operon model.

# The *Lac* repressor

- After the *lac* operon was sequenced, it eventually became clear that it has two additional operators.
- The original operator with its centre of symmetry at position +11 is now designated as *lacO*<sub>1</sub>.

# The *Lac* repressor

- Auxiliary operators *lacO*<sub>2</sub> and *lacO*<sub>3</sub> have their centers of symmetry at positions +412 and -82, respectively.
- Thus, *lacO*<sub>3</sub> is upstream of the *lac* promoter while *lacO*<sub>2</sub> is located in *lacZ*.

# The *Lac* repressor

- The discoveries of *lacO*<sub>2</sub> and *lacO*<sub>3</sub> led investigators to ask whether these auxiliary operators participate in the *lac* operon regulation.



# The *Lac* repressor

- Muller-Hill and coworkers performed series of genetic experiments in which they altered one or more *lac* operators and then determined the alteration's effect on repression.

# The *Lac* repressor

- Point mutations in *lacO*<sub>1</sub> caused a 5 – 50 folds decrease in repression, but some repression was still observed.

# The *Lac* repressor

- Destruction of either *lacO*<sub>2</sub> or *lacO*<sub>3</sub> caused a twofold decrease in repression and destruction of both auxiliary operators caused a 70-fold decrease in repression.

# The *Lac* repressor

- The ability to repress the *lac* system was completely lost in cells in which *lacO*<sub>1</sub> and either of the auxiliary operators were destroyed.

# The *Lac* repressor

- These results indicate that *lacO*<sub>1</sub> play a major role in repression but the two auxiliary operators also make important contributions.

END

# Catabolite Repression

- The function of the  $\beta$ -galactosidase in lactose metabolism is to hydrolyze lactose to form glucose and galactose.

# Catabolite Repression

- If the growth medium contains both glucose and lactose, then in the interest of efficiency, there is no need for a cell to turn on the *lac* operon.
- Monod in 1940s showed that cells behave according to this logic.



# Catabolite Repression

- *E. coli* cells incubated in the presence of glucose and lactose do not start to make  $\beta$ -galactosidase until all the exogenous glucose is consumed.

# Catabolite Repression

- These findings were later extended to lactose permease and transacetylase, the two other proteins specified by *lac* operon.
- The reason that lactose enzymes are not made when glucose is present is that no *lac* mRNA is made.

# Catabolite Repression

- Transcription-level inhibition of the lactose enzymes and a variety of other inducible enzymes by glucose (or other readily used carbon source) is called **catabolite repression**.

# Catabolite Repression

- The mechanism by which glucose inhibits  $\beta$ -galactosidase synthesis remained a complete mystery for about 20 year after Monod first observed the phenomenon.

# Catabolite Repression

- Richard S. Makman and Earl W. Sutherland found an important clue to the mystery in 1965 when they observed that the intracellular concentration of 3',5'-cyclic adenylate (cAMP) drops from

# Catabolite Repression

- about  $10^{-4}$  to  $10^{-7}$  when glucose is added to a growing culture of *E. coli*.
- Genetic studies confirmed the involvement of cAMP in the catabolite repression.

# Catabolite Repression

- Two mutant classes were isolated that could not synthesize *lac* enzymes when cultured in a medium containing lactose but not glucose.



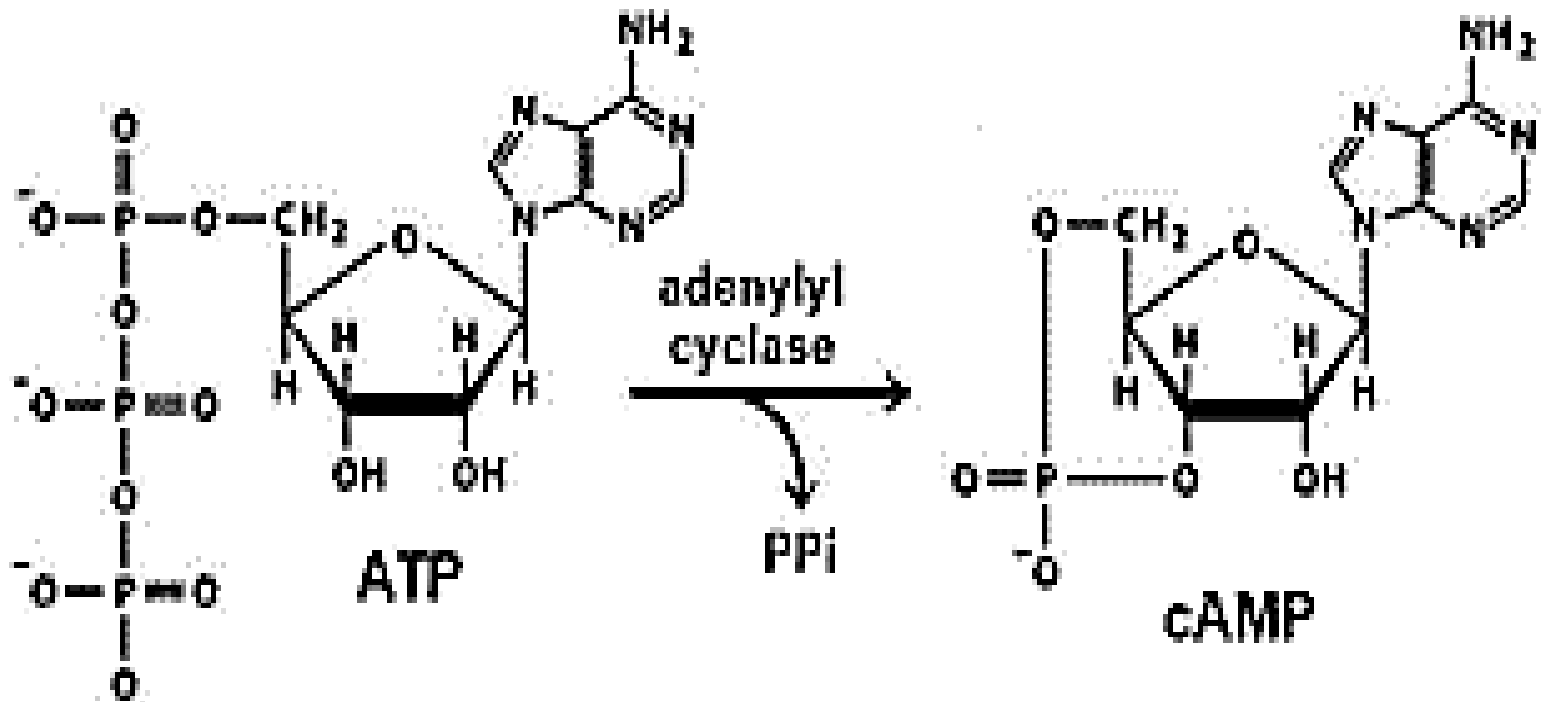
# Catabolite Repression

- Class I mutants regained the ability to synthesize *lac* enzymes when cAMP was added to the growth medium but class II mutants did not.

# Catabolite Repression

- Subsequent studies showed that class I mutants have defects in **adenylate cyclase**, the enzyme that converts ATP to cAMP.
- Adenylate cyclase exists in an active form that is phosphorylated and an inactive form that is dephosphorylated.

# Catabolite Repression



# Catabolite Repression

- Class II mutants have defects in a protein that binds cAMP.
- This protein, called the **cAMP receptor protein (CRP)** or the **catabolite activator protein (CAP)**, is encoded by the *crp* gene.

# Catabolite Repression

- *In vitro* studies have shown that CRP and cAMP form a **cAMP•CRP complex**, which is needed to activate the *lac* system.

END

# Catabolite Repression

- The **cAMP•CRP** is a **positive regulator** or **activator**, in contrast to the repressor, and the *lac* operon is independently regulated both positively and negatively.

# Catabolite Repression

- Based on the information available, it may be proposed that glucose somehow inhibits phosphorylation of adenylate cyclase, thereby preventing cAMP formation.



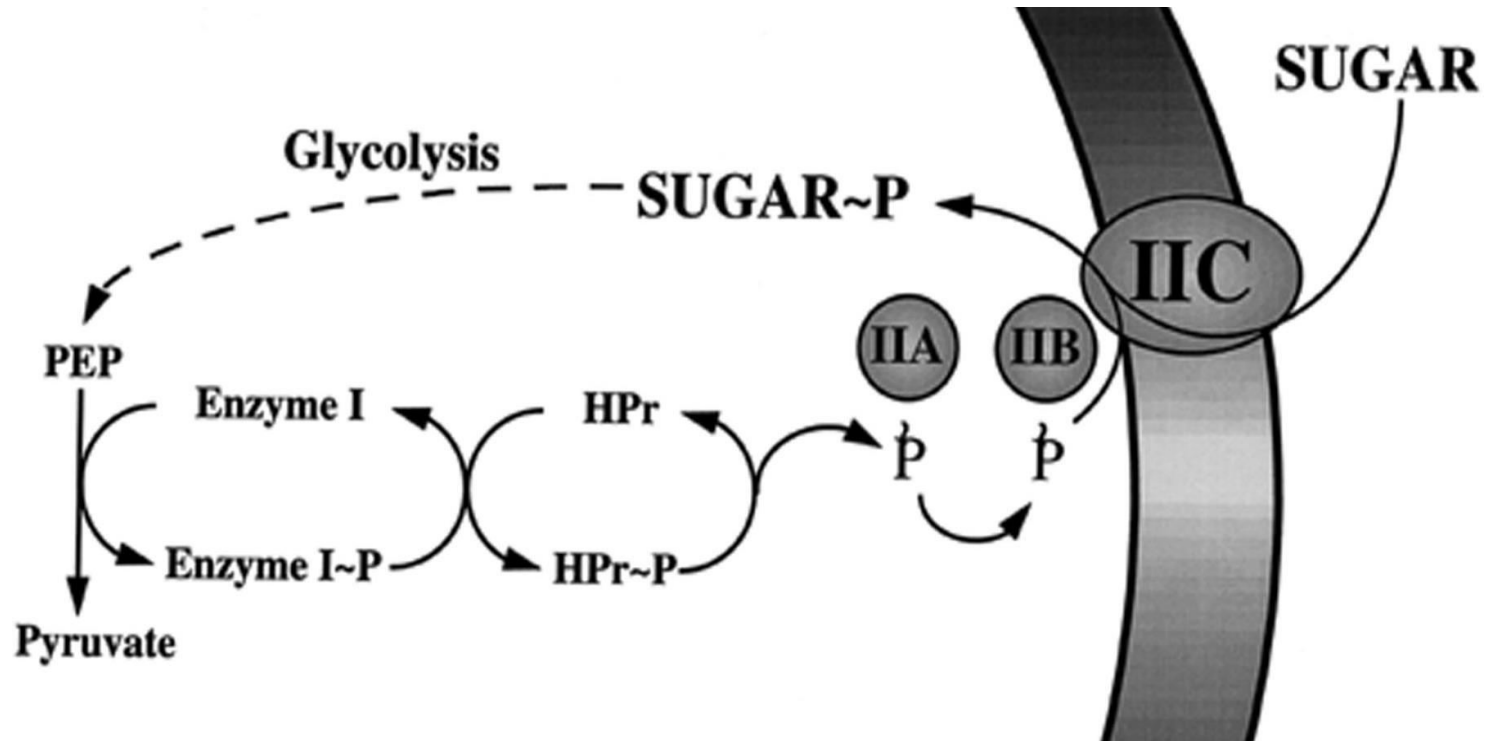
# Catabolite Repression

- The next challenge was to find the link between glucose metabolism and adenylate kinase phosphorylation.
- Studies indicated that the link is a glucose-specific, phosphoenolpyruvate-dependant phosphotransferase system (PTS).

# Catabolite Repression

- This system uses energy supplied by phosphoenolpyruvate (PEP) to phosphorylate glucose as it transports the sugar across the inner cell membrane.
- The system requires four proteins.

# Catabolite Repression



# Catabolite Repression

- Two of these, enzyme I (E-I) and the histidine containing protein (Hpr), are also components of other sugar transporter systems and therefore are unlikely to be direct participants in a glucose-specific phenomenon.

# Catabolite Repression

- The two other proteins, enzyme II (EIIA) and enzyme IIBC (EBIIC), are specific for the glucose transport system and are more likely participants in a glucose-specific phenomenon.

# Catabolite Repression

- EIIA participates in the catabolite repression by two different mechanisms.
- The first mechanism is based on the fact that EIIA can transfer a phosphoryl group from HPr-P to either EIIBC or adenylate cyclase.

# Catabolite Repression

- The preferred substrate is EIIBC, which then transfers the phosphoryl group to glucose to form glucose-6-phosphate.



# Catabolite Repression

- However, when glucose is unavailable, EIIBC will be fully phosphorylated and EIIA-P has no other alternative but to transfer its phosphoryl group to adenylate cyclase.

# Catabolite Repression

- Phosphorylation changes the inactive dephosphorylated form of adenylate cyclase to the active phosphorylated form of adenylate cyclase, which then converts ATP to cAMP.

# Catabolite Repression

- Thus, glucose interferes with the conversion of the inactive form of adenylate cyclase to the active form.

# Catabolite Repression

- The second mechanism by which glucose influences lac operon transcription is called **inducer exclusion**, which also involves the glucose transport system.

# Catabolite Repression

- When glucose is present, EI<sub>1</sub>A-P transfers its phosphate group through EI<sub>2</sub>B to the sugar and the dephosphorylated form of EI<sub>1</sub>A binds to lactose permease and inactivates it.

# Catabolite Repression

- Lactose permease inactivation prevents from entering the cell and being converted to allolactose.
- In the absence of allolactose, the repressor remains bound to the operator and the *lac* operon is turned off.

# Catabolite Repression

- The cAMP modulation and inducer exclusion mechanisms may not be the only ones that contribute to glucose's ability to inhibit *lac* mRNA formation.
- For example, regulation of CRP synthesis may also be important.

END



# The cAMP•CRP complex

- In the absence of cAMP•CRP complex, the *lac* promoter is quite weak because its – 10 box differs significantly from the consensus sequence.

# The cAMP•CRP complex

- A mutant *lac* promoter with a – 10 box that has the consensus sequence does not require cAMP•CRP complex for transcription activation.

# The cAMP•CRP complex

- It therefore seems reasonable to propose that interactions between cAMP•CRP complex and RNA polymerase holoenzyme increase the holoenzyme's affinity for the *lac* promoter.

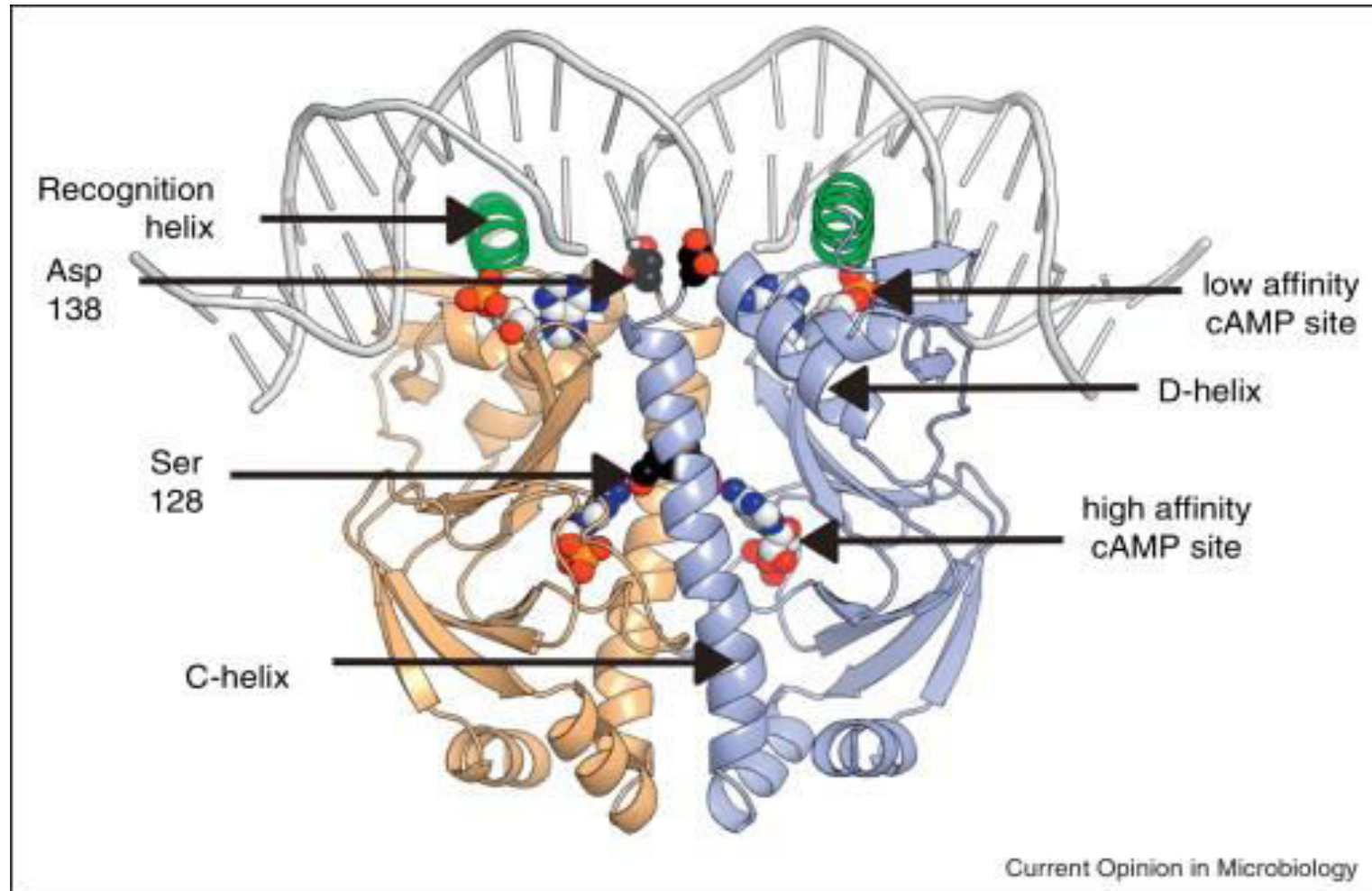
# The cAMP•CRP complex

- Thomas Steitz and coworkers determined the crystal structure of the cAMP•CRP complex bound to DNA.
- CRP consists of two identical polypeptide chains of 209 amino acid residues.

# The cAMP•CRP complex

- Each chain consists of an N-terminal domain and a C-terminal domain.
- The N-terminal domain contains a pocket for binding cAMP while the C-terminal domain contains a motif that binds to DNA.

# The cAMP•CRP complex





# The cAMP•CRP complex

- In the absence of cAMP, CRP•DNA interactions are non specific and weak.
- The cAMP•CRP complex however, binds very tightly to a specific DNA sequence designated as the **activator site (AS)**.



# The cAMP•CRP complex

- In the *lac* operon, the center of AS is 61.5 bp upstream from the transcription start site.
- Promoters like the *lac* promoter that have AS at position – 61.5 are designated as class I cAMP-dependant promoter.

# The cAMP•CRP complex

- Binding of cAMP•CRP to the consensus sequence is so tight that an operon containing this sequence would be permanently switched on.
- It is therefore not surprising that actual cAMP•CRP activator sites differ from the consensus sequence.

# The cAMP•CRP complex

- Activator sites in different operons compete for the cAMP•CRP with the activator preferentially binding to sequences that most closely resemble the consensus sequence.

# The cAMP•CRP complex

- cAMP•CRP bound to DNA migrates in an anomalous fashion when subjected to non denaturing gel electrophoresis.
- The most likely explanation of this is that the protein causes the DNA to bend when it binds to it.

# The cAMP•CRP complex

- X-ray crystallography studies confirmed this interpretation, showing that the cAMP•CRP complex sharply bends DNA by an angle of between  $80^{\circ}$  to  $90^{\circ}$ .

# The cAMP•CRP complex

- In addition to specific contacts with AS, the cAMP•CRP complex also makes specific contacts with RNA polymerase holoenzyme.

END



# Galactose Operon

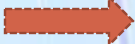
- The *gal* operon, one of the best studied catabolite sensitive operon, contains four structural genes *galK*, *galT*, *galE* and *galM* which specifies the enzymes galactokinase, galactose transferase, galactose epimerase and mutarotase.



# Galactose Operon

- Mutarotase converts the  $\beta$ -D-galactose formed when  $\beta$ -galactosidase hydrolyzes lactose into  $\alpha$ -D-galactose.

# Galactose Operon

- Then galactokinase, galactose transferase and epimerase act in a sequence of steps to yield the overall reaction:-
- Galactose + ATP  Glucose-1-phosphate + ADP

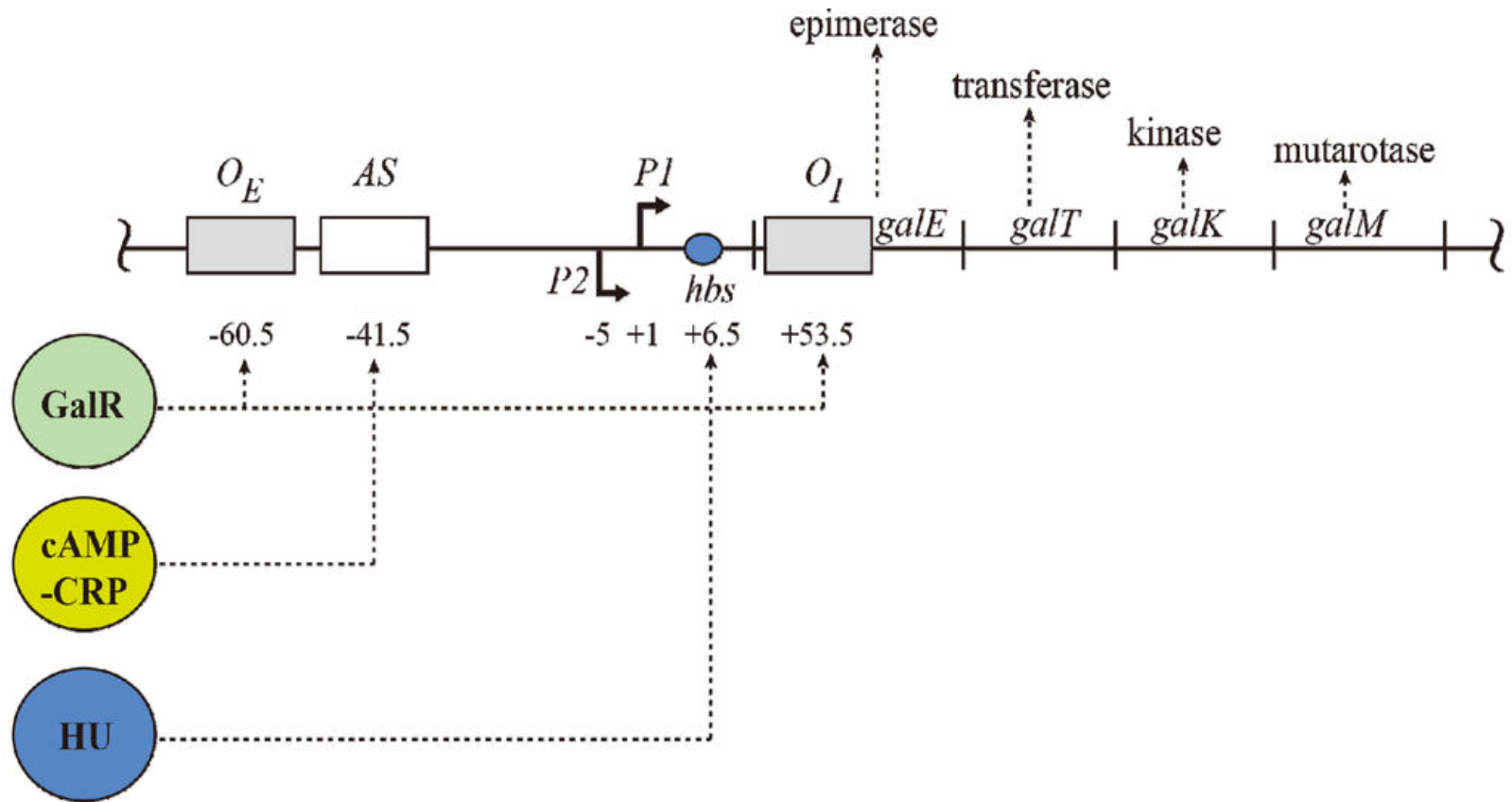
# Galactose Operon

- The galactose (*gal*) operon, like the *lac* operon, is regulated by a repressor and cAMP•CRP.
- The structural gene for the repressor, *galR*, is located far from structural genes for the galactose enzymes.

# Galactose Operon

- The *gal* operon also has two operators,  $galO_E$  ( $O_E$ ) and  $galO_I$  ( $O_I$ ), a promoter region, and a cAMP•CRP AS.
- The promoter region contain two promoters  $P_1$  and  $P_2$ .

# Galactose Operon



# Galactose Operon

- One operators,  $galO_E$ , is upstream from the promoter region and the other  $galO_I$ , is in  $galE$ .
- A mutation in either operator causes at least partial constitutivity, indicating that both operators participate in repression.

# Galactose Operon

- The Gal repressor, which binds to both operators, is made of two identical polypeptide chains.
- A null mutation in *galR* makes both promoters constitutive.



# Galactose Operon

- The amino acid sequence of Gal repressor is similar to that of Lac repressor.
- There is one important difference, however,

# Galactose Operon

- the Gal repressor does not have the C-terminal residues that allow the Lac repressor to form a four helical bundle and become a stable tetramer.

# Galactose Operon

- The Gal repressor has a C-terminal domains that binds the inducer galactose and an NTD with a helix-turn-helix motif that binds to a galO<sub>E</sub> or galO<sub>I</sub> half-site.

# Galactose Operon

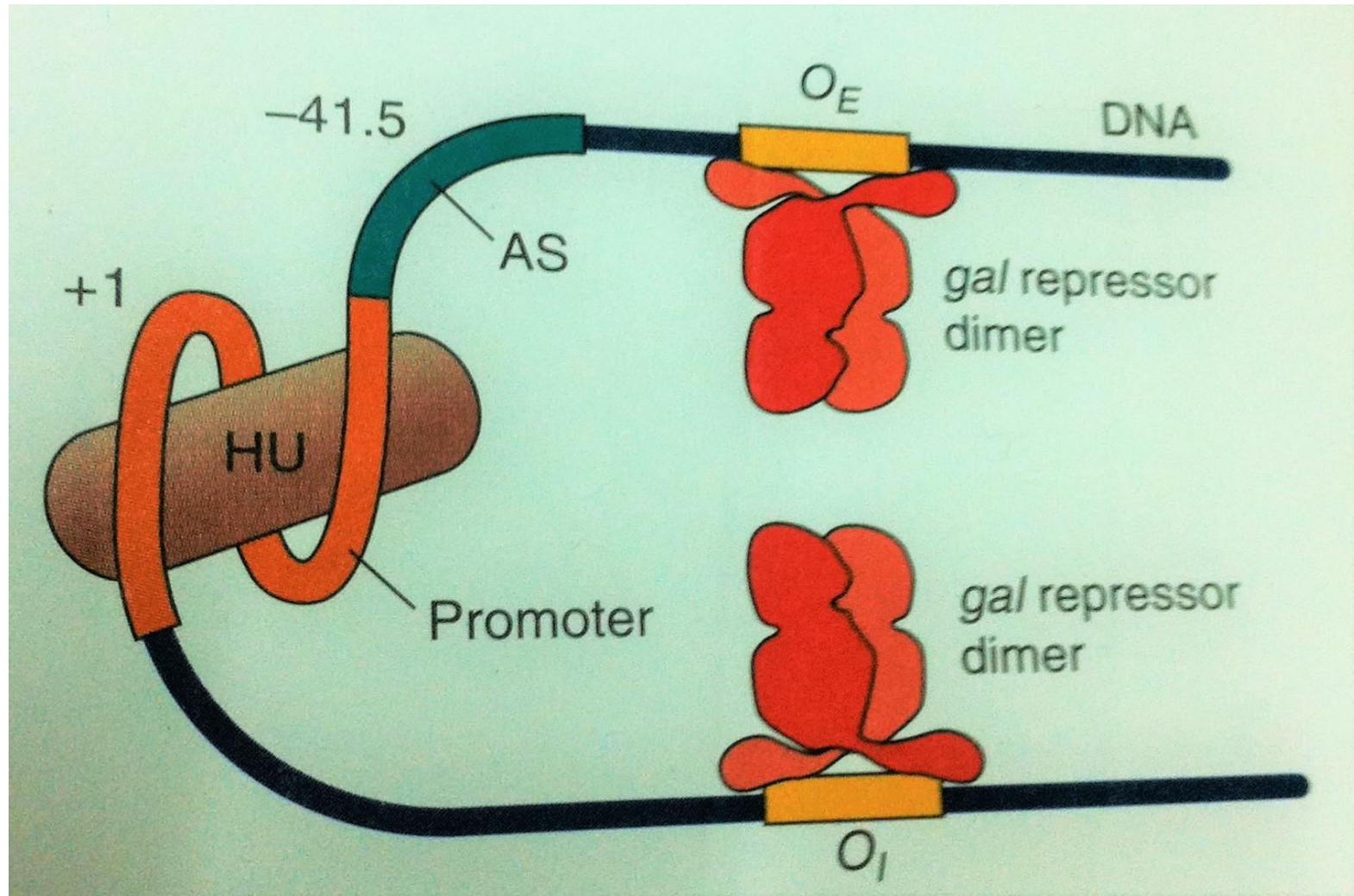
- The Gal repressor is a dimer and not a tetramer and the studies which indicated that both  $galO_E$  and  $galO_I$  are required for repression, suggested some type of cooperative interaction.

# Galactose Operon

- Sanker Adhya and coworkers solved the problem by showing that a histone-like protein called HU binds to the *gal* promoter region only when the Gal repressors are bound at both  $galO_E$  or  $galO_I$  causing the DNA to loop.



# Galactose Operon



# Galactose Operon

- The two *gal* promoter ( $P_1$  &  $P_2$ ) are separated by 5 bp and as a result, the transcription initiation site for  $P_1$  is 5 bp upstream from that for  $P_2$ .

END



# Galactose Operon

- Polycistronic mRNA molecules formed in response to either promoter contain all the information required to synthesize the four galactose enzymes.

# Galactose Operon

- Binding cAMP•CRP to AS repress transcription from  $P_2$  but activates transcription from  $P_1$ .

# Galactose Operon

- Moreover,  $P_2$  can support *gal* mRNA transcription in the presence of glucose when intracellular cAMP•CRP levels are low but  $P_1$  cannot do so.

# Galactose Operon

- The fact that AS is 20 bp closer to  $P_1$  in the *gal* operon than it is to  $P_{lac}$  in the *lac* operon leads one to predict that cAMP•CRP and RNA polymerase interactions should be different in the *gal* and *lac* operons.

# Galactose Operon

- This prediction is correct.
- Promoters like  $P_1$  in *gal* operon that have AS at position – 41.5 are called **class II cAMP-dependant promoters**.

# Galactose Operon

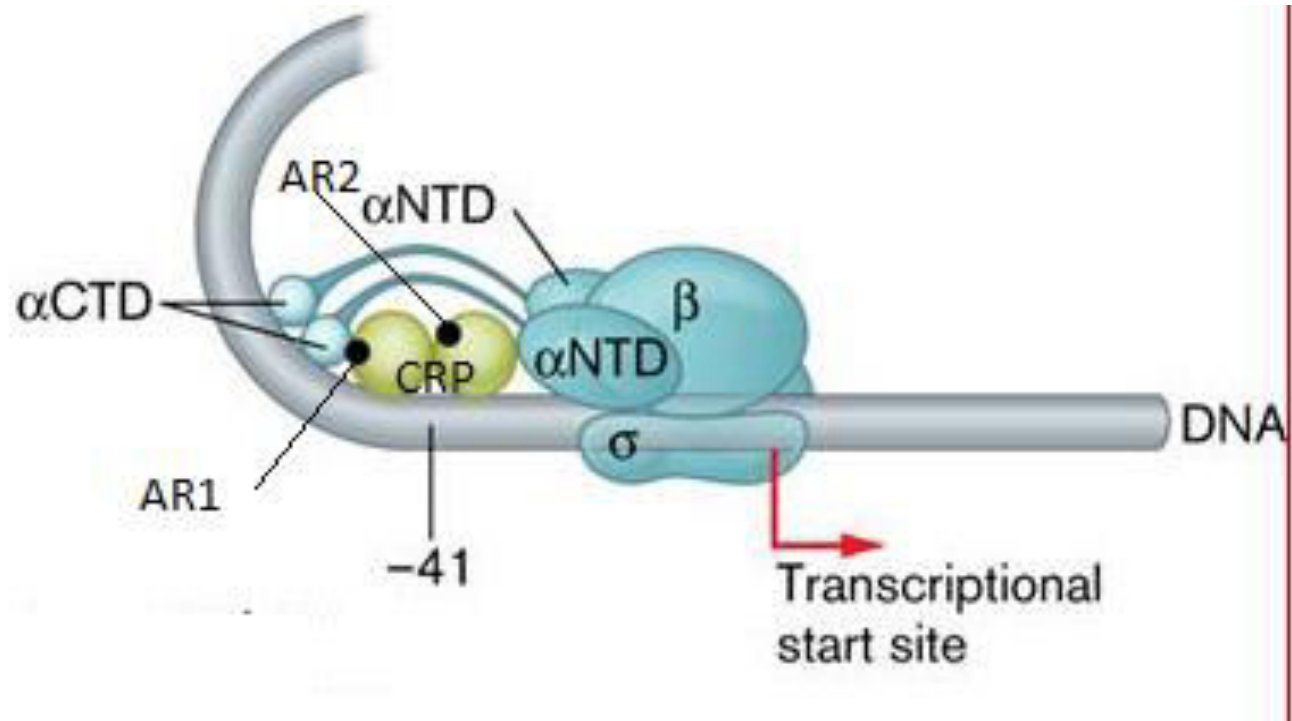
- Three activator regions on CRP – **AR1, AR2** and **AR3** – interact with RNA polymerase holoenzyme.

# Galactose Operon

- AR1 on the upstream unit of the CRP dimer and AR2 and AR3 on the downstream subunit of dimer interact with  $\alpha$ CTD,  $\alpha$ NTD, and the  $\sigma^{70}$  subunit, respectively.



# Galactose Operon



# Galactose Operon

- Why does the *gal* operon has two promoters?
- The answer appears to be that galactose has two roles in cellular metabolism.
- It is both a carbon source and a precursor for lipopolysachharide synthesis.

# Galactose Operon

- When galactose is not available in the growth medium, cells require the epimerase specified by the *gal* operon to convert glucose to galactose, which is then used to make lipopolysachharides.

# Galactose Operon

- Synthesis from the  $P_2$  promoter permits the low level of epimerase formation required to convert glucose to galactose-1-phosphate so that the lipopolysachharides can be formed.

# Galactose Operon

- If  $P_1$  were the only promoter, then epimerase could not be made when glucose is present because  $P_1$  require cAMP•CRP activation.

# Galactose Operon

- On the other hand, if  $P_2$  were the only promoter, then galactose could not fully induce the operon when galactose was the sole carbon source because cAMP•CRP inhibits  $P_2$ .

# Galactose Operon

- Thus, for the sake of both necessity and economy, a cAMP•CRP – independent promoter ( $P_2$ ) is needed for background constitutive synthesis and a cAMP•CRP – dependent promoter



# Galactose Operon

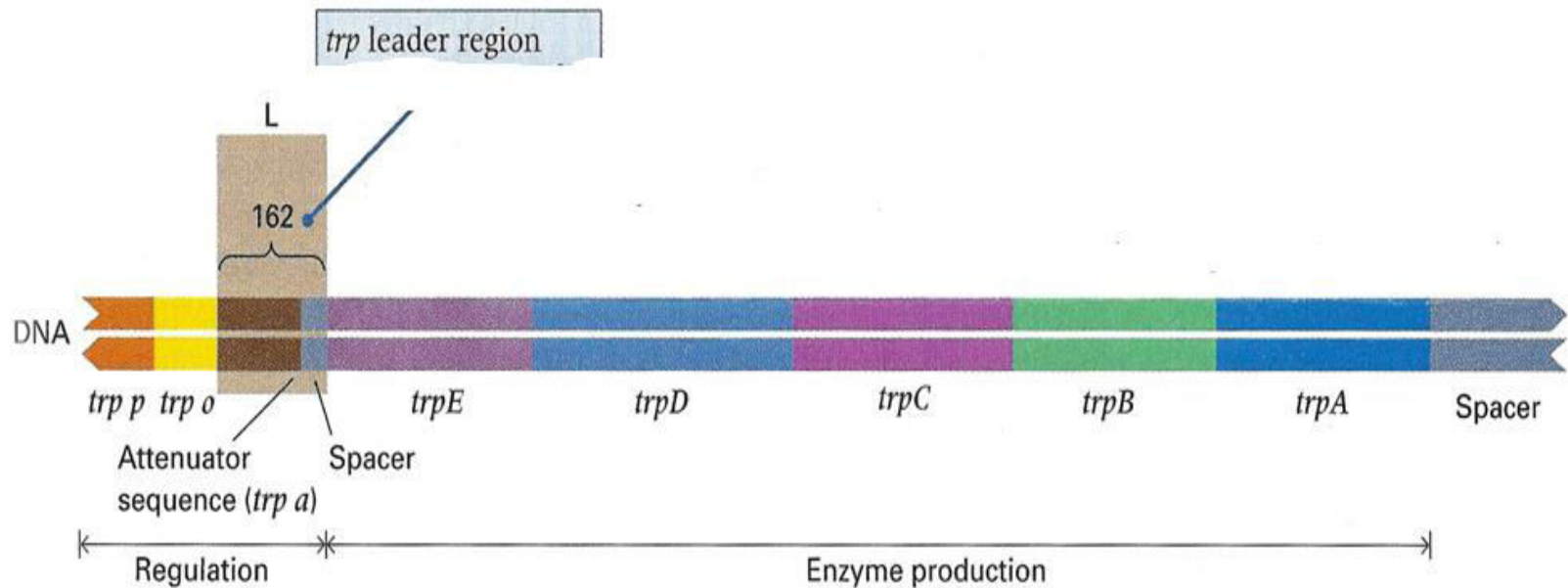
- ( $P_1$ ) is needed to regulate high-level synthesis.
- Furthermore, the regulation is efficient only if  $P_2$  is inhibited by cAMP•CRP.

END

# Tryptophan Operon

- The *trp* operon consists of a promoter, an operator, a leader (*trpL*), an attenuator, and five structural genes designated as *trpE*, *trpD*, *trpC*, *trpB* and *trpA*.

# Tryptophan Operon



# Tryptophan Operon

- Consistent with the fact that the *trp* operon specifies biosynthetic rather than degradative enzymes, it does not have a cAMP•CRP activation site.

# Tryptophan Operon

- The operator is part of a coarse on-off control, and the leader and attenuator allow for finer control.
- The leader sequence, which codes for a short peptide is the first region of the *trp* operon to be transcribed.

# Tryptophan Operon

- RNA polymerase then moves forward to transcribe the five structural genes in the order *trpE* → *trpA* to form the polycistronic *trp* mRNA.

# Tryptophan Operon

- The five polypeptides specified by the structural genes form the three enzymes that are essential for tryptophane biosynthesis.



# Tryptophan Operon

- Early studies revealed that the *trp* system is turned off when tryptophane is added to an *E. coli* culture, suggesting that the operon is repressed rather than induced.

# Tryptophan Operon

- Repression, like induction, involves negative regulation of transcription initiation by a regulatory protein.
- However, the regulatory protein does not bind to the operator until after forming a complex with tryptophan.

# Tryptophan Operon

- The structural gene for the Trp regulatory protein, *trpR*, is located a considerable distance away from the *trp* operon.

# Tryptophan Operon

- This distance does not present a problem because the regulatory protein diffuses throughout the cell.
- The biologically active form of the regulatory protein is a homodimer.

# Tryptophan Operon

- Each subunit contains a helix-turn-helix motif that can bind to a *trp* operator half-site.
- The *trp* operator and promoter regions have significant overlap.
- Therefore, binding of the TrpR•tryptophane complex and RNA polymerase are mutually exclusive.

# Tryptophan Operon

- The TrpR•tryptophane complex functions as a coarse on-off switch that turns the tryptophane operon off when tryptophane levels are high.

# Tryptophan Operon

- A fine control mechanism also exists that allows cells to regulate their tryptophane enzyme concentration according to the tryptophane concentration.



# Tryptophan Operon

- An important clue to the existence of this fine control mechanism came from an experiment performed by Charles Yanofsky in 1972, which showed that *E. coli* mutants that lack a functional TrpR protein

# Tryptophan Operon

- increase *trp* operon transcription after being starved for tryptophan.
- If TrpR•tryptophan complex were the only regulatory factor, then transcription of the *trp* operon should not have increased.

# Tryptophan Operon

- Further studies revealed that *trp* mRNA has a 162 nucleotide sequence before the first codon in *trpE*, designated as the **leader** or *trpL*, which plays an essential role in the fine control mechanism.

# Tryptophan Operon

- Constitutive mutants exhibit a sixfold increase in tryptophane enzyme synthesis when bases 123 – 150 within the leader are deleted.
- This 28 base sequence is called the **attenuator**.

# Tryptophan Operon

- The attenuator can fold into a stem-and-loop structure with the potential to function as a rho-independent transcription terminator.

# Tryptophan Operon

- Deleting the attenuator removes the transcription termination site, allowing RNA polymerase to complete *trp* mRNA synthesis.

END

# Tryptophan Operon

- The *trp* leader has four complementary segments that can interact to form two sets of mutually exclusive hairpin structures.
- Segment 1 & 2 can base pair to form hairpin 1•2 while segment 3 & 4 can form hairpin 3•4.



# Tryptophan Operon

- Alternatively, segment 2 & 3 can base pair to form hairpin 2•3.
- The 3•4 hairpin, which contains the attenuator sequence, is a transcription terminator.

# Tryptophan Operon

- It follows that RNA polymerase would be able to synthesize full length *trp* mRNA if conditions were somehow favorable for 2•3 hairpin formation rather than 3•4 hairpin formation.

# Tryptophan Operon

- RNA polymerase molecules that escape repression begin synthesizing *trp* mRNA.
- RNA polymerase continues transcribing *trp* operon until it encounters a pause site located just after segment 2 of the leader sequence.

# Tryptophan Operon

- Because the leader sequence is at the 5' end of *trp* mRNA, it is the first sequence available for translation.
- Ribosomes begin translating the leader sequence at its AUG start codon.

# Tryptophan Operon

- When the moving ribosome reaches the paused RNA polymerase, the paused RNA polymerase is released.
- Pausing serves the important role of synchronizing the transcription and translation process.

# Tryptophan Operon

- The subsequent fate of the RNA polymerase depends on the tryptophan concentration i.e.,
  - 1. high tryptophan concentration or
  - 2. low tryptophan concentration.



# Tryptophan Operon

- In the first case, ribosome moves past the tryptophan codon in the segment 1 and continues to translate the leader region until it encounters the stop codon (UGA) between segment 1 & 2 and falls off the nascent mRNA molecule.



# Tryptophan Operon

- Once free of the ribosome, segment 1 pairs with segment 2 to form the 1•2 hairpin.
- RNA polymerase continues to transcribe the leader region, synthesizing segment 3 and then segment 4.

# Tryptophan Operon

- These two segments pair to form the rho independent transcription terminator, which causes RNA polymerase to fall off the DNA template, preventing *trpE* transcription.

# Tryptophan Operon

- In the second case when the concentration of tryptophan is very low, the bulky ribosome will pause at the tryptophane codon (UGG) on segment 1, preventing segment 1 from pairing with segment 2 to form the 1•2 hairpin.

# Tryptophan Operon

- Segment 2 is therefore free to interact with segment 3 to form 2•3 hairpin (the so called **antiterminator**) as soon as the RNA polymerase completes the synthesis of segment 3.

# Tryptophan Operon

- Then RNA polymerase can continue to synthesize the complete *trp* mRNA because the 3•4 hairpin (the transcription terminator) is not formed.

# Tryptophan Operon

- Thus, if tryptophan is present in excess, transcription termination occurs at the attenuator and little enzyme is synthesized.

# Tryptophan Operon

- If tryptophan is absent, transcription termination does not occur and tryptophan enzymes are formed.

END



# Gene Expression Is Controlled by Regulatory Proteins

- Genes are very often controlled by extracellular signals; in the case of bacteria, these signals typically mean molecules present in the growth medium.

# Gene Expression Is Controlled by Regulatory Proteins

- These signals are communicated to genes by regulatory proteins, which come in two types: positive regulators, or **activators**, and negative regulators, or **repressors**.

# Gene Expression Is Controlled by Regulatory Proteins

- Typically, these regulators are DNA-binding proteins that recognize specific sites at or near the genes they control.
- An activator increases transcription of the regulated gene, and repressors decrease or eliminate that transcription.

# Gene Expression Is Controlled by Regulatory Proteins

- How do these regulators work?
- First, RNA polymerase binds to the promoter in a closed complex (in which the DNA strands remain together).

# Gene Expression Is Controlled by Regulatory Proteins

- The polymerase–promoter complex then undergoes a transition to an open complex in which the DNA at the start site of transcription is unwound and the polymerase is positioned to initiate transcription.

# Gene Expression Is Controlled by Regulatory Proteins

- This is followed by promoter escape, the step in which polymerase leaves the promoter and starts transcribing.
- Polymerase then proceeds through the elongation phase before finally terminating.



# Gene Expression Is Controlled by Regulatory Proteins

- Which steps are stimulated by activators and inhibited by repressors depends on the promoter and regulators in question.



# Gene Expression Is Controlled by Regulatory Proteins

- The most common step at which gene expression is regulated is the initiation of transcription.
- There are two reasons why this might make sense.

# Gene Expression Is Controlled by Regulatory Proteins

- First, transcription initiation is the most energetically efficient step to regulate.
- By this we mean that deciding whether or not to express a gene at the first step ensures that no energy or resources are wasted.

# Gene Expression Is Controlled by Regulatory Proteins

- Second, regulation at this first step is easier to do.
- There is only a single copy of each gene and so typically only a single promoter on a single DNA molecule must be regulated to control expression of a given gene.

# Gene Expression Is Controlled by Regulatory Proteins

- Why then is not all regulation focused on the step of transcription initiation?
- Regulating later steps can have two advantages.

# Gene Expression Is Controlled by Regulatory Proteins

- First, it allows for more inputs: if a gene is regulated at more than one step, more signals can modulate its expression, or the same signals can do so even more effectively.

# Gene Expression Is Controlled by Regulatory Proteins

- Second, regulation at steps later than transcription initiation can reduce the response time.
- Thus, consider again the example of translational regulation.



# Gene Expression Is Controlled by Regulatory Proteins

- If a signal relieves repression of this step, the protein product encoded by the gene will be produced immediately upon receipt of that signal.
- This reduced response time might obviously be advantageous in some situations.



# Gene Expression Is Controlled by Regulatory Proteins

- Many promoters are regulated by activators that help RNA polymerase bind DNA and by repressors that block that binding.

# Gene Expression Is Controlled by Regulatory Proteins

- At many promoters, in the absence of regulatory proteins, RNA polymerase binds only weakly.
- This is because one or more of the promoter elements is absent or imperfect.

# Gene Expression Is Controlled by Regulatory Proteins

- When polymerase does occasionally bind, however, it spontaneously undergoes a transition to the open complex and initiates transcription.

# Gene Expression Is Controlled by Regulatory Proteins

- This gives a low level of **constitutive expression** called the **basal level**.
- Binding of RNA polymerase is the rate-limiting step in this case.

# Gene Expression Is Controlled by Regulatory Proteins

- To control expression from such a promoter, a repressor need only bind to a site overlapping the region bound by polymerase.

# Gene Expression Is Controlled by Regulatory Proteins

- In that way, the repressor blocks polymerase binding to the promoter, thereby preventing transcription, although it is important to note that repression can work in other ways as well.

# Gene Expression Is Controlled by Regulatory Proteins

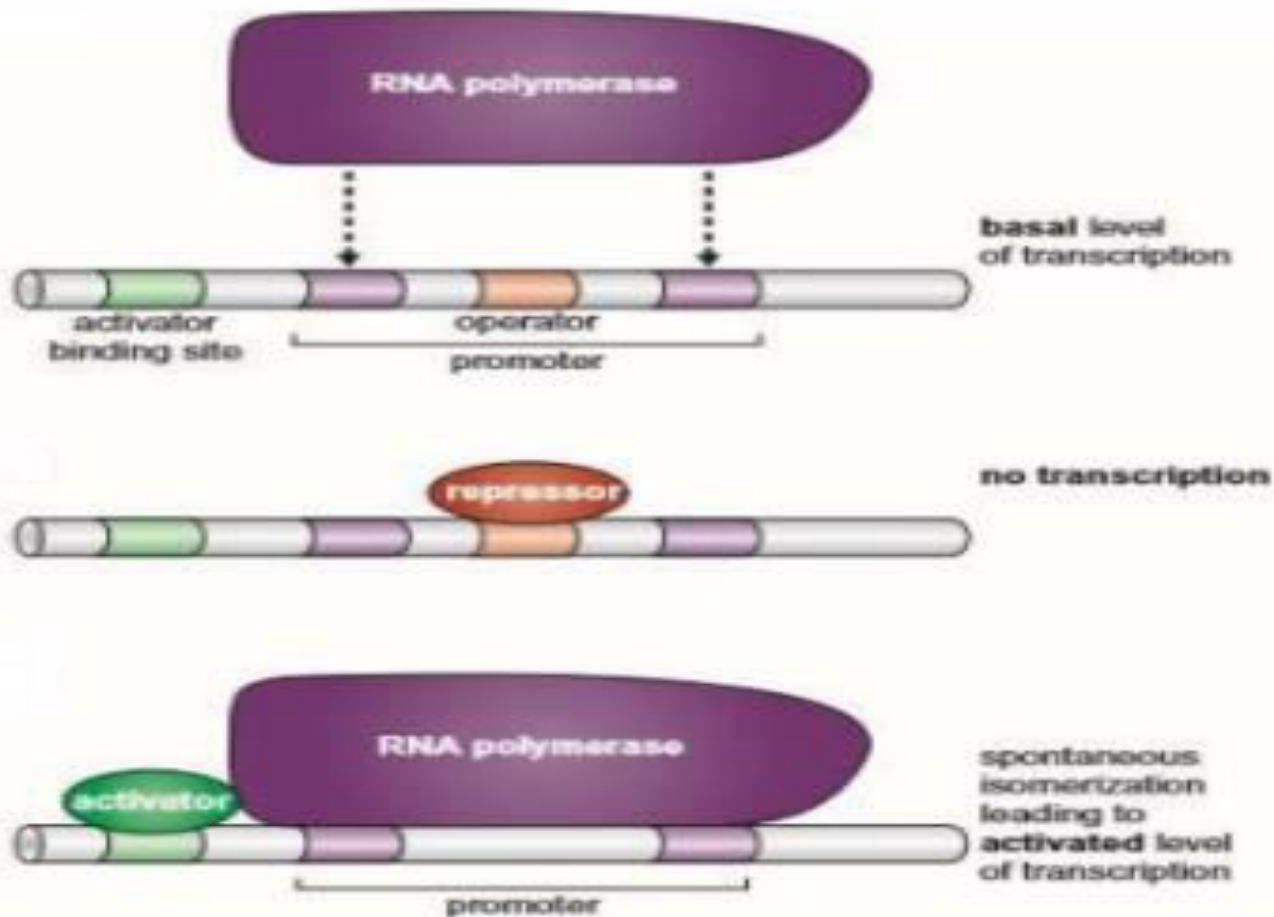
- The site on DNA where a repressor binds is called an **operator**.
- To activate transcription from this promoter, an activator can just help the polymerase bind the promoter.
- Typically, this is achieved as follows:



# Gene Expression Is Controlled by Regulatory Proteins

- the activator uses one surface to bind to a site on the DNA near the promoter; with another surface, the activator simultaneously interacts with RNA polymerase, bringing the enzyme to the promoter.

# Gene Expression Is Controlled by Regulatory Proteins



# Gene Expression Is Controlled by Regulatory Proteins

- This mechanism, often called **recruitment**, is an example of **cooperative binding** of proteins to DNA.
- The interactions between the activator and polymerase, and between activator and DNA, serve merely “adhesive” roles.

# Gene Expression Is Controlled by Regulatory Proteins

- The enzyme is active and the activator simply brings it to the nearby promoter.
- Once there, it spontaneously isomerizes to the open complex and initiates transcription.

END

## Some Activators and Repressors Work by Allostery

- Let us consider another class of promoter in which RNA polymerase binds efficiently unaided and forms a stable closed complex.

## Some Activators and Repressors Work by Allostery

- But that closed complex does not spontaneously undergo transition to the open complex.
- At this promoter, an activator must stimulate the transition from a closed to open complex, since that transition is the rate-limiting step.

## Some Activators and Repressors Work by Allostery

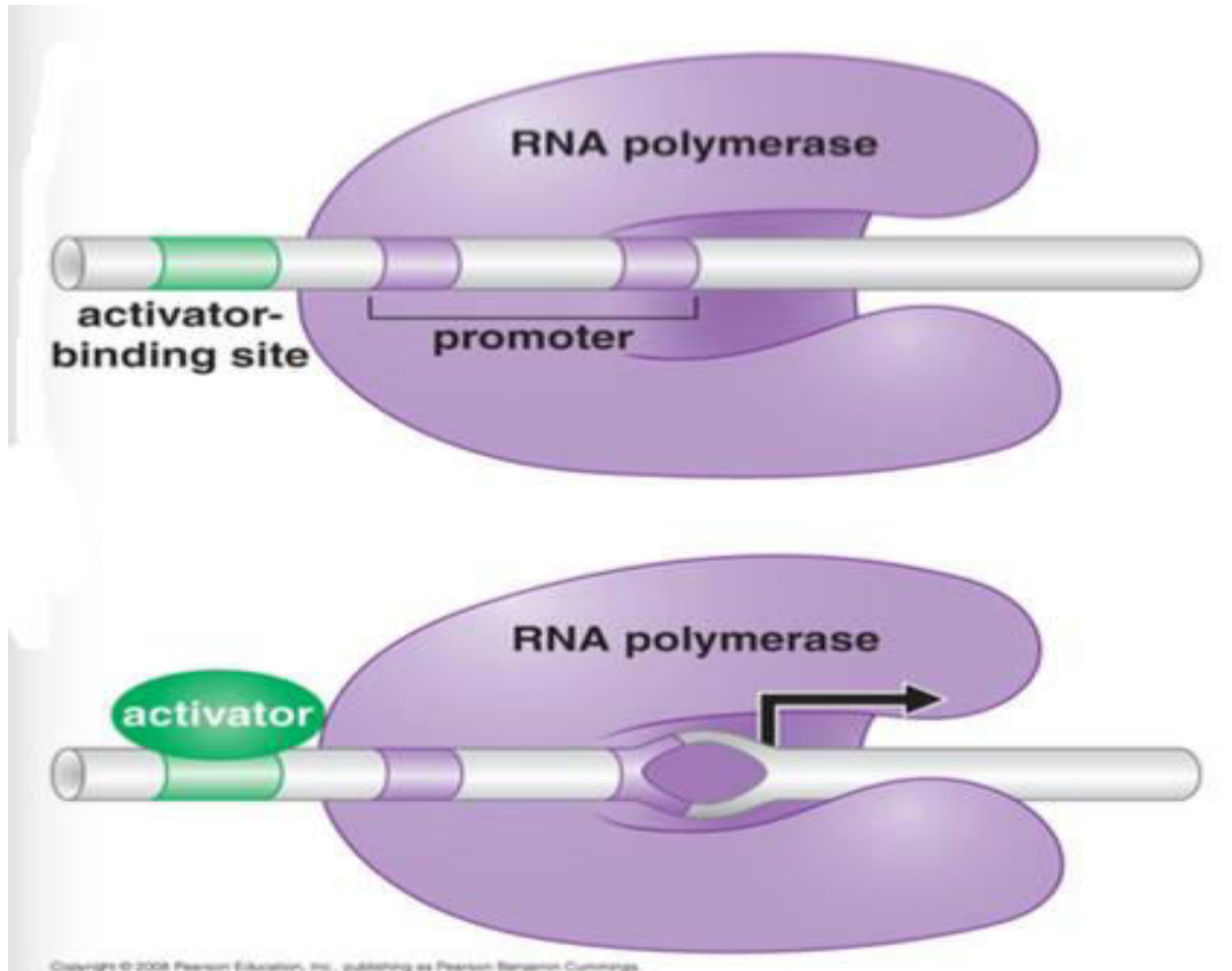
- Activators that stimulate this kind of promoter work by triggering a conformational change in either RNA polymerase or DNA.



## Some Activators and Repressors Work by Allostery

- It means that, they interact with the stable closed complex and induce a conformational change that causes transition to the open complex.
- This mechanism is an example of **allostery**.

# Some Activators and Repressors Work by Allostery



## Some Activators and Repressors Work by Allostery

- For example, the *glnA* promoter and *NtrC* the activator.
- In still another class of promoter, transcription initiation is limited at the step of promoter escape.

## Some Activators and Repressors Work by Allostery

- One example of such a promoter directs expression of the *malT* gene.
- In the absence of an activator, it undergoes abortive initiation, and only in the presence of an activator will it efficiently escape into elongation.

## Some Activators and Repressors Work by Allostery

- Repressors can work in ways other than just blocking the binding of RNA polymerase.
- For example, some repressors interact with polymerase at the promoter and inhibit transition to the open complex, or promoter escape.

## Some Activators and Repressors Work by Allostery

- Generally, DNA-binding proteins that interact with each other bind to adjacent sites.
- But some proteins interact with each other even when bound to sites well separated on the DNA.



## Some Activators and Repressors Work by Allostery

- To accommodate this interaction, the DNA between the sites loops out, bringing the sites into proximity with one another.
- Distant DNA sites can be brought closer together to help loop formation.



## Some Activators and Repressors Work by Allostery

- In bacteria, for example, there are cases in which a protein binds between an activator-binding site and the promoter and helps the activator interact with polymerase by bending the DNA in a favorable direction.

## Some Activators and Repressors Work by Allostery

- There are also cases where such a protein hinders loop formation and activation by bending the DNA in an unfavorable direction.

END

# Cooperative Binding and Allostery

- It is already pointed out that gene activation can be mediated by simple cooperative binding.
- The activator interacts simultaneously with DNA and with polymerase and so recruits the enzyme to the promoter.

# Cooperative Binding and Allostery

- How activation can be mediated by allosteric events: an activator interacts with polymerase already bound to the promoter and, by inducing a conformational change in the enzyme or the promoter, stimulates transcription initiation.

# Cooperative Binding and Allostery

- Both cooperative binding and allostery have additional roles in gene regulation.
- For example, groups of regulators often bind DNA cooperatively:-

# Cooperative Binding and Allostery

- two or more activators and/or repressors interact with each other and with DNA and thereby help each other bind near a gene they all regulate.



# Cooperative Binding and Allostery

- This kind of interaction can produce sensitive switches that allow a gene to go from completely off to fully on in response to only small changes in conditions.



# Cooperative Binding and Allostery

- Cooperative binding of activators can also serve to integrate signals: some genes are activated only when multiple signals (and thus multiple regulators) are simultaneously present.

# Cooperative Binding and Allostery

- A particularly striking and well-understood example of cooperativity in gene regulation is provided by bacteriophage  $\lambda$ .

# Cooperative Binding and Allostery

- Allostery is not only a mechanism of gene activation, but also often the way regulators are controlled by their specific signals.

# Cooperative Binding and Allostery

- Thus a typical bacterial regulator can adopt two conformations: in one, it can bind DNA; in the other, it cannot.

# Cooperative Binding and Allostery

- Binding of a signal molecule locks the regulatory protein in one or another conformation, thereby determining whether or not it can act.

# Cooperative Binding and Allostery

- The bulk of gene regulation takes place at the initiation of transcription.
- This is true in eukaryotes just as it is in bacteria. But regulation is certainly not restricted to that step in either class of organisms.

# Cooperative Binding and Allostery

- In bacteria, gene regulation also occurs at the level of transcriptional elongation and termination and even at translational levels.



# Cooperative Binding and Allostery

- There are also cases involving regulation by RNAs e.g., attenuation, riboswitches, and small RNAs.

END

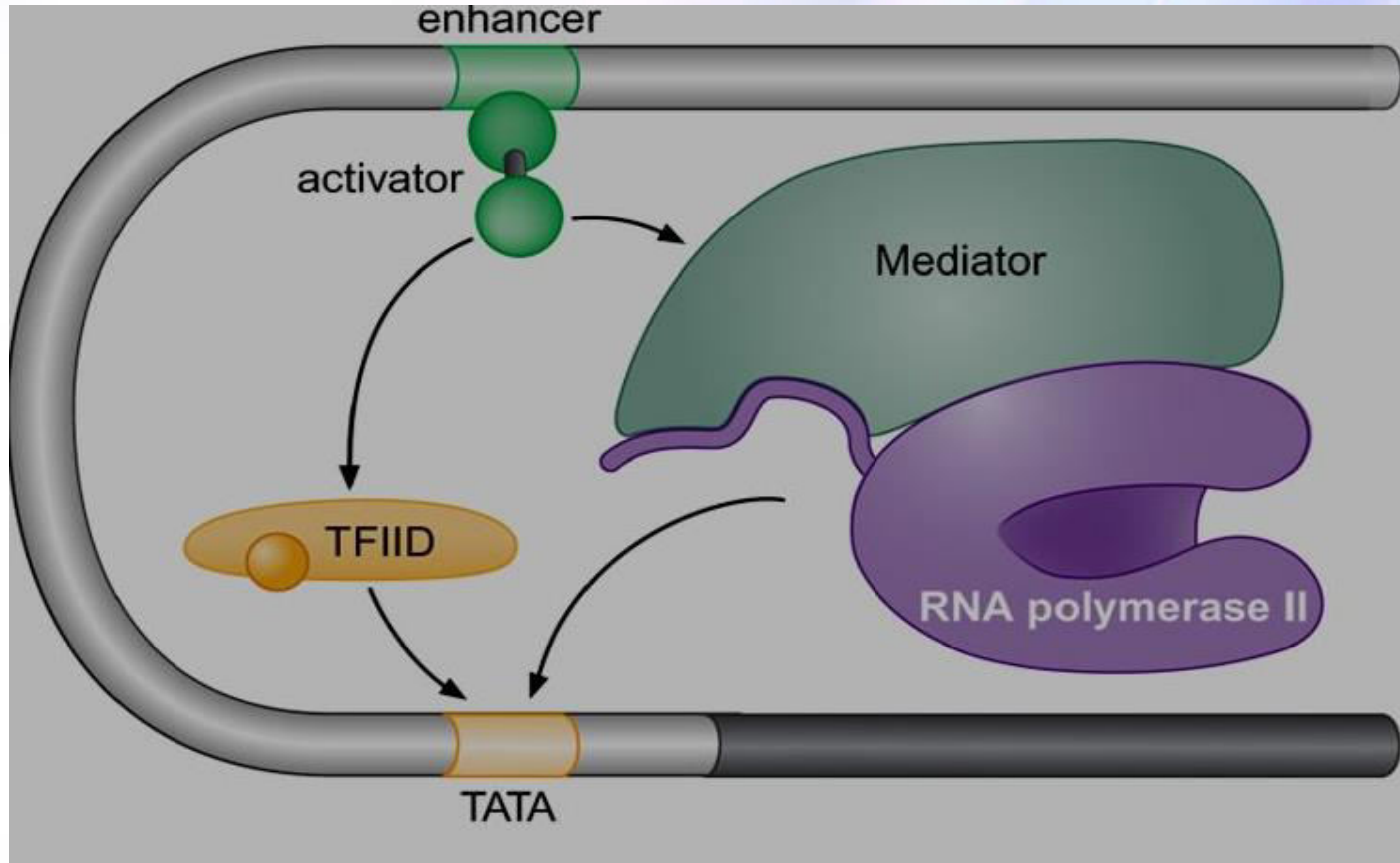
# Activators Recruit the Transcriptional Machinery

- The eukaryotic transcriptional machinery contains numerous proteins in addition to RNA polymerase.
- Many of these proteins come in preformed complexes such as mediator and the TFIID Complex.

# Activators Recruit the Transcriptional Machinery

- Activators interact with one or more of these complexes and recruit them to the gene.
- Other components that are not directly recruited by the activator bind cooperatively with those that are recruited.

# Activators Recruit the Transcriptional Machinery



# Activators Recruit the Transcriptional Machinery

- At most genes , the transcriptional machinery appears at the promoter only upon activation of the gene.
- That is, the machinery is not pre-bound, confirming that the role of the activator is to recruit it.

# Activators Recruit the Transcriptional Machinery

- In bacteria, we saw that genes activated by recruitment (such as the lac genes) can be activated in so-called activator bypass experiments.



# Activators Recruit the Transcriptional Machinery

- In such an experiment, activation is observed when RNA polymerase is recruited to the promoter without using the natural activator –polymerase interaction.
- Similar experiments work in yeast.



# Activators Recruit the Transcriptional Machinery

- Thus, the GAL1 gene (normally activated by Gal4) can be activated equally well by a fusion protein containing the DNA-binding domain of the bacterial protein LexA fused directly to a component of the Mediator complex.

# Activators Recruit the Transcriptional Machinery

- It is important to note that these experiments do not exclude the possibility that at least some activators not only recruit parts of the transcriptional machinery but also induce allosteric changes in them.

# Activators Recruit the Transcriptional Machinery

- Such changes might stimulate the efficiency of transcription initiation.

# Activators Recruit the Transcriptional Machinery

- Nevertheless, the recruitment of the machinery to one or another gene is the basis of specificity; that is, which gene is activated depends on which gene has the machinery recruited to it.

# Activators Recruit the Transcriptional Machinery

- In addition, the success of the activator bypass experiments suggests that any allosteric events that occur during initiation do not require the activator to do anything beyond recruiting proteins to the gene.

# Activators Recruit Nucleosome Modifiers

- In addition to direct recruitment of the transcriptional machinery, recruitment of nucleosome modifiers can help activate a gene packaged within chromatin.



# Activators Recruit Nucleosome Modifiers

- Nucleosome modifiers come in two types:
- 1) those that add chemical groups to the tails of histones, such as **histone acetyltransferases** (HATs ), which add acetyl groups, and



# Activators Recruit Nucleosome Modifiers

- 2) those that displace (or “remodel”) the nucleosomes, such as the ATP-dependent activity of **SWI/SNF**.
- How do these modifications help activate a gene?

# Activators Recruit Nucleosome Modifiers

- Two basic models explain how changes in nucleosomes can help the transcriptional machinery bind at the promoter.

# Activators Recruit Nucleosome Modifiers

- First, remodeling, and certain modifications, can uncover DNA-binding sites that would otherwise remain inaccessible within the nucleosome.

# Activators Recruit Nucleosome Modifiers

- For example, by removing or increasing the mobility of nucleosomes, remodelers are proposed to free up binding sites for regulators and for the transcriptional machinery.

# Activators Recruit Nucleosome Modifiers

- Similarly, the addition of acetyl groups to histone tails alters the interactions between those tails and adjacent nucleosomes.
- This modification is often said to “loosen” chromatin structure, freeing up sites.

# Activators Recruit Nucleosome Modifiers

- But adding acetyl groups also helps binding of the transcriptional machinery (and other proteins) in another way: it creates specific binding sites on nucleosomes for proteins bearing so-called **bromodomains**.

# Activators Recruit Nucleosome Modifiers

- One component of the TFIID complex bears bromodomains and thus binds to acetylated nucleosomes better than to unacetylated nucleosomes.



# Activators Recruit Nucleosome Modifiers

- Thus, a gene bearing acetylated nucleosomes at its promoter will likely have a higher affinity for the transcriptional machinery than one with unacetylated nucleosomes.

# Activators Recruit Nucleosome Modifiers

- Other proteins contain **chromodomains**.
- These recognize methylated nucleosomes, examples of which we will encounter later.

# Activators Recruit Nucleosome Modifiers

- Some components of the transcriptional machinery are more stringently required at some genes than at others, and the same applies to nucleosome modifiers as well.

# Activators Recruit Nucleosome Modifiers

- Thus, although all genes absolutely require RNA polymerase itself, a given gene may depend on another particular component of the transcriptional machinery, or a nucleosome modifier, or it may not.

# Activators Recruit Nucleosome Modifiers

- In addition, what is needed to activate a given gene can vary depending on circumstances, such as the stage of the cell cycle.
- For example, Gal4 usually activates the *GAL1* gene efficiently in the absence of a histone acetylase.

# Activators Recruit Nucleosome Modifiers

- During mitosis, however, when chromatin is more condensed, activation is eliminated unless that acetylase is recruited to the gene.

# Activators Recruit Nucleosome Modifiers

- In yeast, recent experiments have provided good evidence for particular activator–target interactions at specific genes.



# Activators Recruit Nucleosome Modifiers

- The acidic activator Gcn4 is known to interact with Gal11; it also interacts with the TAF12 subunit of TFIID and other complexes involved in transcription, including the nucleosome remodeler SWI/SNF.

# Activators Recruit Nucleosome Modifiers

- And Gal4 appears to contact at least three components: Mediator, TFIID, and a third complex called SAGA (Spt-Ada-Gcn5-acetyltransferase).

# Activators Recruit Nucleosome Modifiers

- The last of these complexes harbors acetylation activity and seems to be capable of interacting with the transcriptional machinery as well.

# Activators Recruit Nucleosome Modifiers

- The ability of an acidic activator such as Gal4 to work at genes with different requirements can be explained by its ability to interact with multiple targets.

# Activators Recruit Additional Factors

- The elaborate transcriptional machinery of a eukaryotic cell contains numerous proteins required for initiation.
- It also contains some that aid in elongation.

# Activators Recruit Additional Factors

- At some genes, sequences downstream from the promoter cause pausing or stalling of the polymerase soon after initiation.

# Activators Recruit Additional Factors

- At those genes, the presence or absence of certain elongation factors greatly influences the level at which the gene is expressed.



# Activators Recruit Additional Factors

- One example is the *HSP70* gene from *Drosophila*.
- This gene, activated by heat shock, is controlled by two activators working together.

# Activators Recruit Additional Factors

- The GAGA binding factor is believed to recruit enough of the transcription machinery to the promoter for initiation of transcription.

# Activators Recruit Additional Factors

- But, in the absence of a second activator, HSF, most of the initiated polymerases stall some 25–50 bp downstream from the promoter.

# Activators Recruit Additional Factors

- In response to heat shock, HSF binds to specific sites at the promoter and recruits a kinase, P-TEFb (positive transcription elongation factor), to the stalled polymerases.

# Activators Recruit Additional Factors

- The kinase phosphorylates the carboxy-terminal domain (CTD) of the largest subunit of RNA polymerase (the so-called CTD “tail”), freeing the enzyme from the stall and allowing transcription to proceed through the gene.

# Activators Recruit Additional Factors

- Recent studies suggest that P-TEFb is part of a larger complex, the SEC (super elongation complex), which releases paused Pol II from the proximal promoter.

# Activators Recruit Additional Factors

- Phosphorylation of the CTD tail on Ser5 of the heptad repeat is an important step in the early stages of transcription at all genes, and the kinase TFIIF can perform that phosphorylation.
- Whether P-TEFb is also needed at most genes is not clear.



# Activators Recruit Additional Factors

- P-TEFb has been implicated in the phosphorylation of Ser2 of the CTD heptad repeat, and this modification is associated with the release of activated Pol II from promoter sequences.

# Activators Recruit Additional Factors

- A strong acidic activator like Gal4 is able to recruit PTEFb/SEC along with the rest of the machinery.

# Activators Recruit Additional Factors

- It may be that only at certain genes is the recruitment of the machinery partitioned between regulators in the way we see at this *HSP70* gene, allowing an extra layer of control.

# Activators Recruit Additional Factors

- The human immunodeficiency virus (HIV), which causes AIDS, transcribes its genes from a promoter controlled by P-TEFb (and SEC).

# Activators Recruit Additional Factors

- Again, polymerase initiates transcription at that promoter, under the control of the activator SP1, but stalls soon afterward.
- In this case, P-TEFb is brought to the stalled polymerase by an RNA-binding protein called TAT.

# Activators Recruit Additional Factors

- TAT recognizes a specific sequence near the start of the HIV RNA and present in the nascent transcript made by the stalled polymerase.
- Another domain of TAT interacts with P-TEFb and recruits it to the stalled polymerase.



# Activators Recruit Additional Factors

- This results in the release of polymerase, the transcription of the viral genome, and infection of the host cell, typically a T-lymphocyte.
- It is now believed that paused polymerase is more commonly seen, particularly during development.



# Activators Recruit Additional Factors

- Thus, recent studies in human embryonic stem cells and the early *Drosophila* embryo suggest that roughly one-third of all protein-coding genes contain paused Pol II before their activation during development.

# Activators Recruit Additional Factors

- Such genes might be particularly dependent on recruitment of SEC for their expression.

# Activators Recruit Additional Factors

- It is possible that paused Pol II is also away of excluding inhibitory nucleosomes from the promoter region, rendering the promoter “poised” for rapid activation by upstream regulatory sequences.

# Single integration & combinatorial control

- The *lac* genes of *E. coli* are efficiently expressed only when lactose is present and glucose absent.
- The two signals are communicated to the gene through separate regulators - one an activator and the other a repressor.

# Single integration & combinatorial control

- In multicellular organisms signal integration is used extensively.
- In some cases, numerous signals are required to switch a gene on.

# Single integration & combinatorial control

- But just as in bacteria, each signal is transmitted to the gene by a separate regulator, so at many genes multiple activators must work together to switch the gene on.



# Single integration & combinatorial control

- When multiple activators work together, they do so synergistically.
- That is, the effect of, say, two activators working together is usually much greater than the sum of each of them working alone.



# Single integration & combinatorial control

- Synergy can result from multiple activators recruiting a single component of the transcriptional machinery; multiple activators each recruiting a different component;

# Single integration & combinatorial control

- or multiple activators helping each other bind to their sites upstream of the gene they control.
- Let us see all three strategies:-

# Single integration & combinatorial control

- Two activators can recruit a single complex-for example, the Mediator-by touching different parts of it.
- The combined binding energy will have an exponential effect on recruitment.

# Single integration & combinatorial control

- Synergy can also result from activators helping each other bind under conditions where the binding of one depends on binding of the other.

# Single integration & combinatorial control

- This co-operativity can be seen in bacteria, whereby the two activators touch each other when they bind their sites on DNA.

# Single integration & combinatorial control

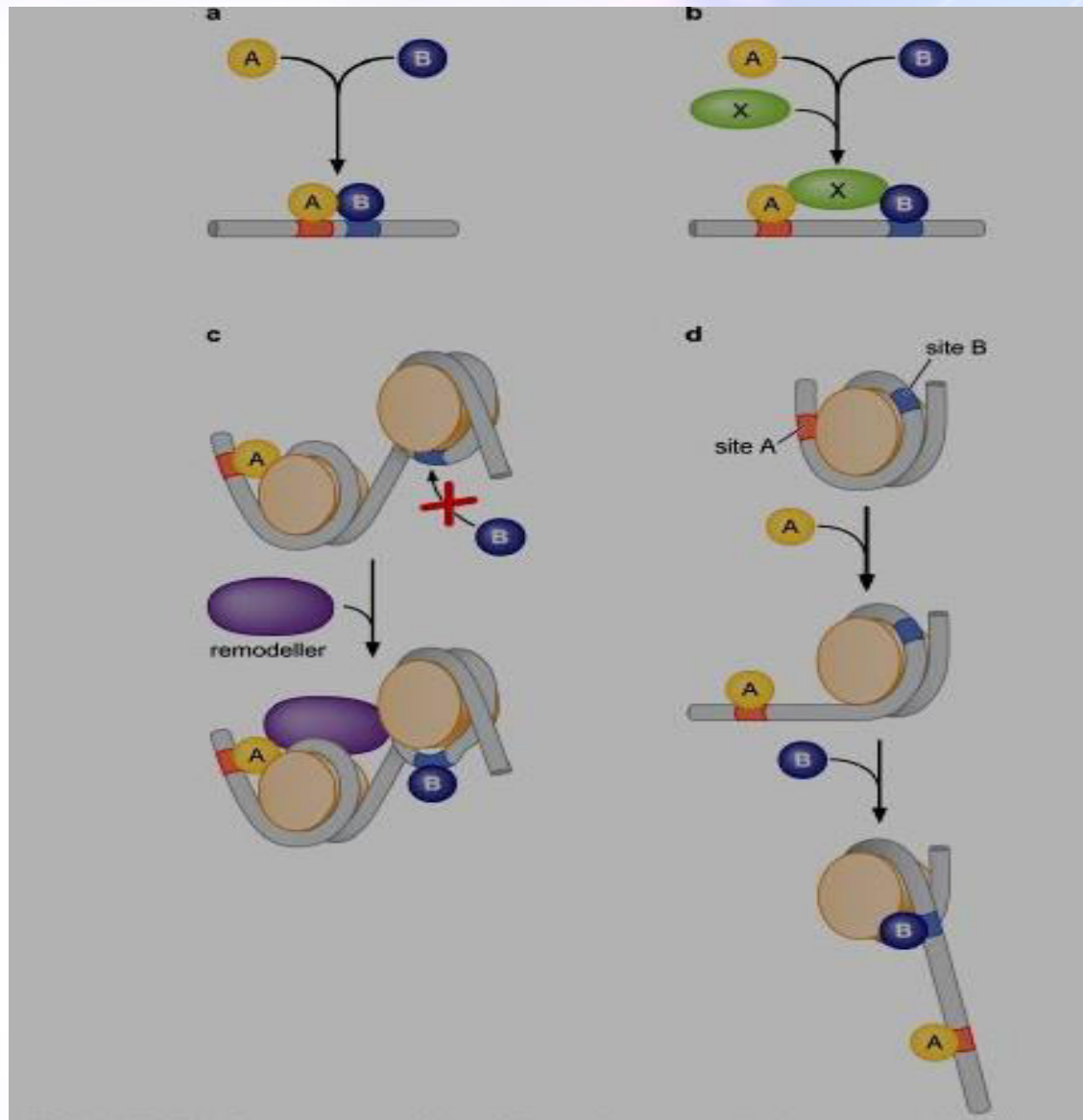
- But it can work in other ways as well: one activator can recruit something that helps the second activator bind.
- The different ways activators help each other bind DNA include "classical" cooperative binding;

# Single integration & combinatorial control

- recruitment of a modifier by one activator to help a second bind; and binding of one activator to nucleosomal DNA uncovering the binding site for another.



# Single integration & combinatorial control



# Single integration & combinatorial control

- Synergy is critical for signal integration by activators.
- Consider a gene whose product is only needed when two signals are received.

# Single integration & combinatorial control

- Each signal is communicated to the gene by a separate activator.
- The gene must be efficiently expressed when both activators are present but be relatively impervious to the action of either activator alone.

# Single integration & combinatorial control

- Each signal is communicated to the gene by a separate activator.
- The gene must be efficiently expressed when both activators are present but be relatively impervious to the action of either activator alone.

# Single integration & combinatorial control

- We have seen simple cases of combinatorial control in bacteria.
- For example, CAP is involved in regulating many genes, in collaboration with other regulators.

# Single integration & combinatorial control

- At the *lac* genes it works with the Lac repressor; at the *gal* genes with the Gal repressor.
- There is also extensive combinatorial control in eukaryotes.



# Single integration & combinatorial control

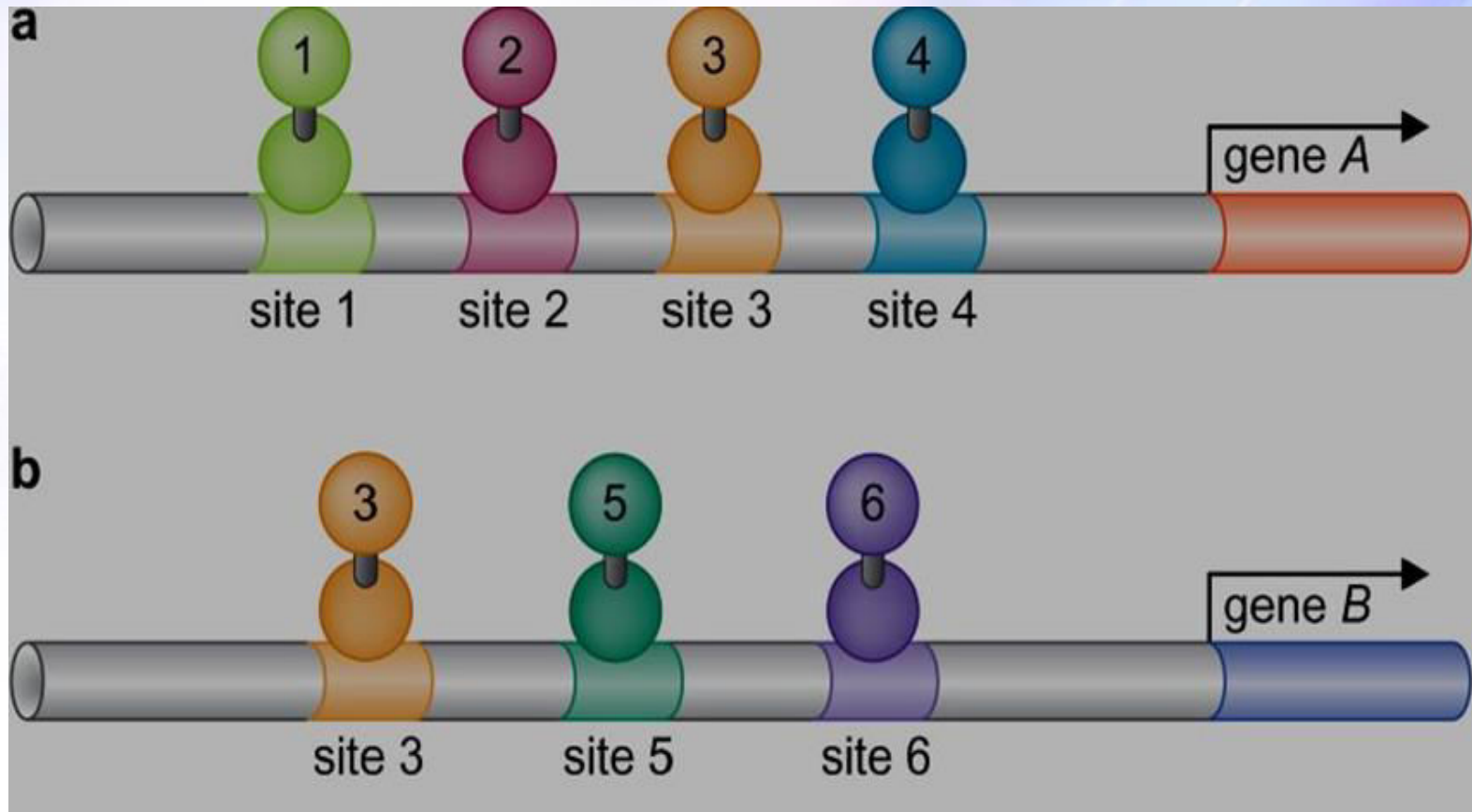
- Let us consider a generic case.
- Gene A is controlled by four signals (1, 2, 3, and 4), each working through a separate activator (activators 1, 2, 3, and 4).
- Gene B is controlled by three signals (3, 5, and 6), working through activators 3, 5, and 6.



# Single integration & combinatorial control

- Note that there is one signal in common between these two cases, and the activator through which that signal works is the same at both genes.

# Single integration & combinatorial control



# Single integration & combinatorial control

- In complex multicellular organisms, such as *Drosophila* and humans, combinatorial control involves many more regulators and genes as shown in this example.

# Single integration & combinatorial control

- How is it that the regulators can intermix so promiscuously?
- As you know, multiple activators work synergistically.

# Single integration & combinatorial control

- In fact, even multiple copies of a single activator work synergistically, suggesting that a given activator can interact with multiple targets.

# Single integration & combinatorial control

- This provides an explanation for why different regulators can work together in so many combinations: because each can use any of an array of targets, the combinations that work together are unrestricted.



# Single integration & combinatorial control

- The examples of signal integration such as the *HO* gene in yeast and the human  $\beta$ -interferon gene- involve activators that also regulate other genes in examples of combinatorial control.



# Transcriptional Repressors

- In bacteria, we saw that many repressors work by binding to sites that overlap the promoter and thus block binding of RNA polymerase.

# Transcriptional Repressors

- But we also saw other ways they can work: they can bind to sites adjacent to promoters and, by interacting with polymerase bound there, inhibit the enzyme from initiating transcription.

# Transcriptional Repressors

- They can also interfere with the action of activators.
- In eukaryotes, we see all of these except the first.
- We also see another form of repression, perhaps the most
- common in eukaryotes, that works as follows.

# Transcriptional Repressors

- As with activators, repressors can recruit nucleosome modifiers, but in this case, the enzymes have effects opposite to those recruited by activators - they compact the chromatin or remove groups recognized by the transcriptional machinery.

# Transcriptional Repressors

- Therefore, for example, histone deacetylases repress transcription by removing acetyl groups from the tails of histones in *S. cerevisiae*; as we have already seen, the presence of acetyl groups helps transcription.

# Transcriptional Repressors

- Paradoxically, the histone deacetylase Rpd3 is also recruited to active genes to ensure transcription fidelity.

# Transcriptional Repressors

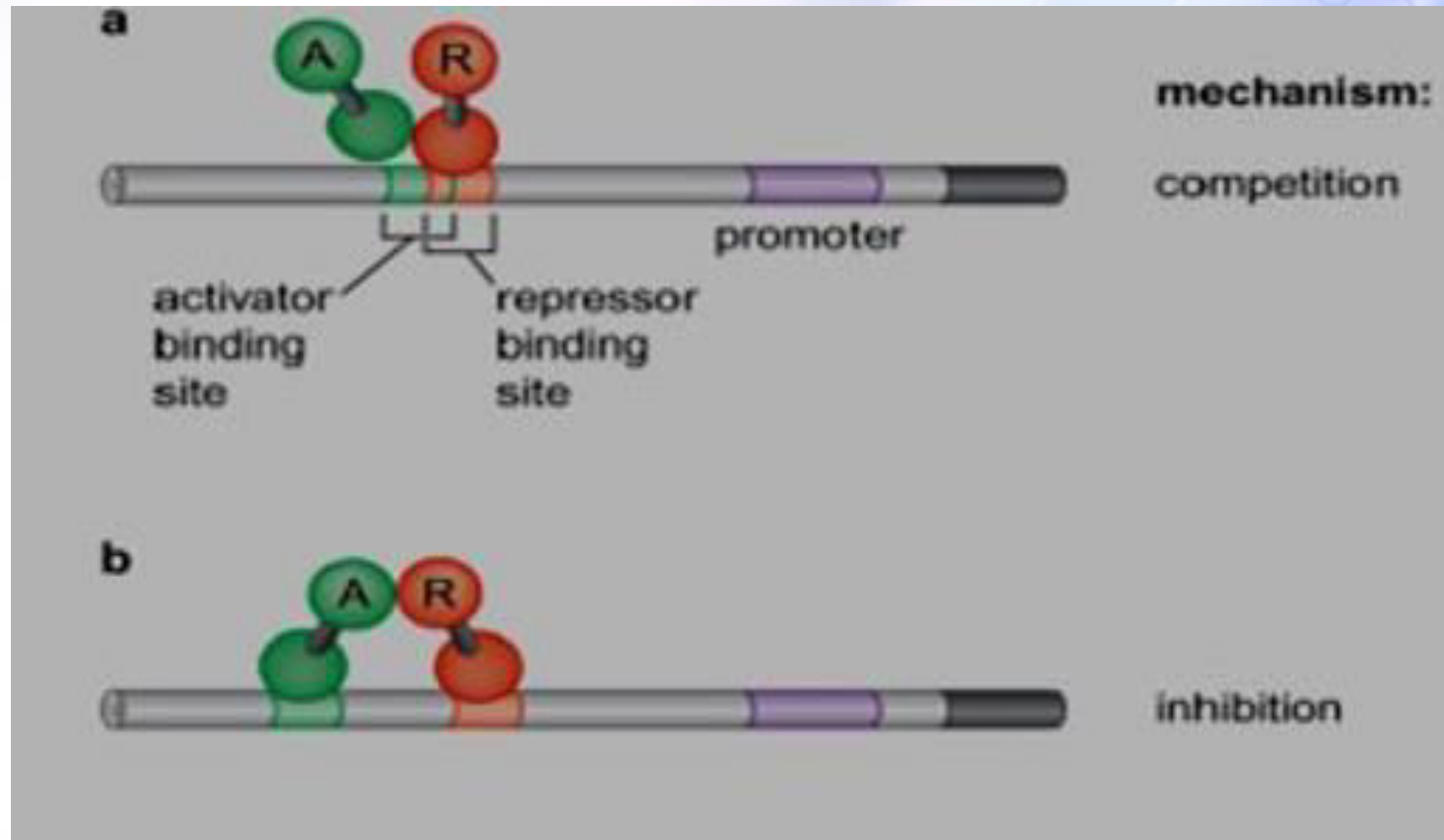
- Nucleosomes are deacetylated behind elongating Pol II to prevent the use of “cryptic” promoters within the transcription unit.



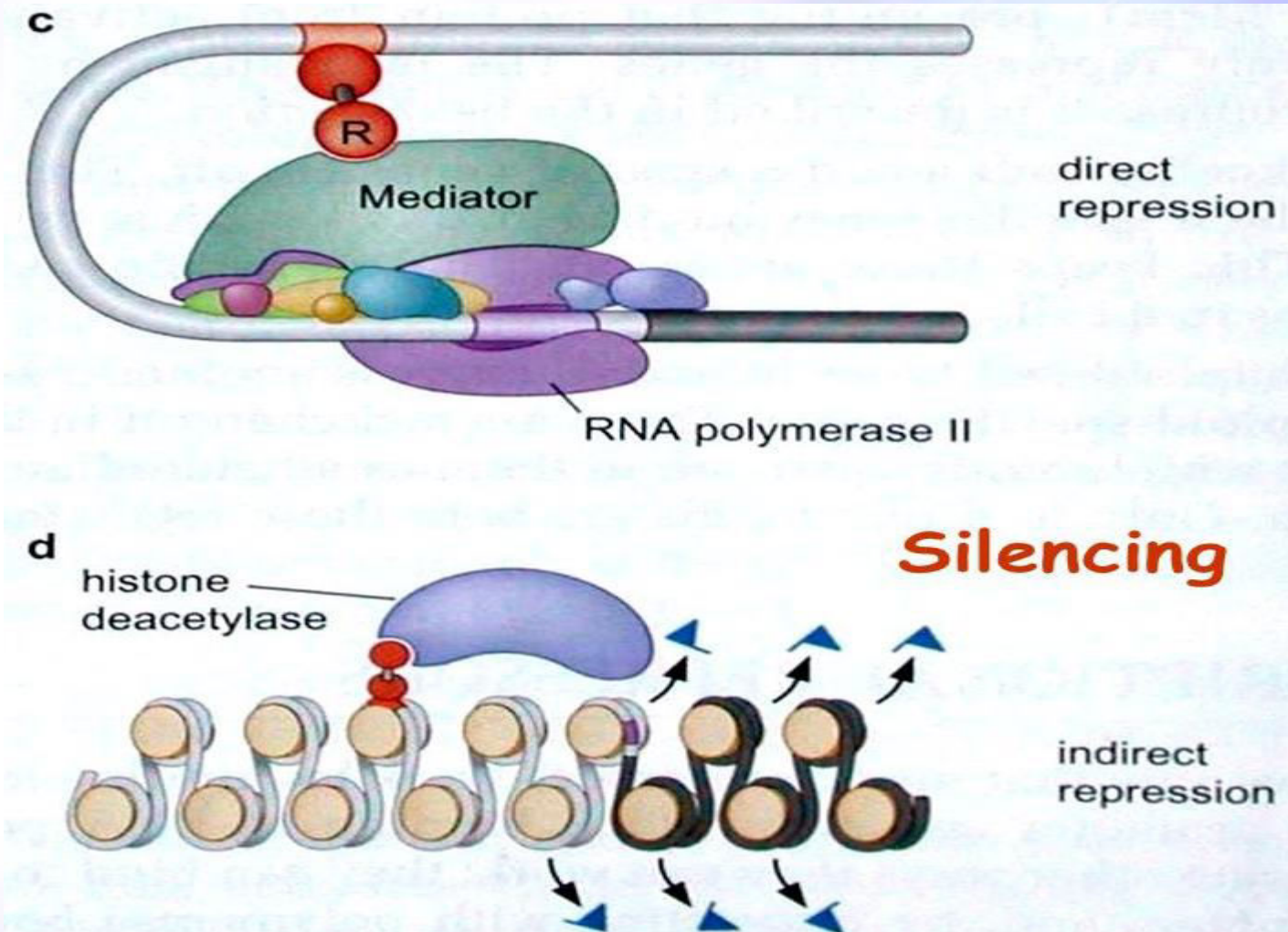
# Transcriptional Repressors

- Other enzymes add methyl groups to histone tails, and this frequently represses transcription, although in some cases it is associated with an actively transcribed gene.

# Transcriptional Repressors



# Transcriptional Repressors



# Transcriptional Repressors

- Histone (and DNA) modifications also form the basis of a type of repression called **silencing**.
- Let us consider one specific example, the repressor called Mig1, which, like Gal4, is involved in controlling the GAL genes of the yeast *S. cerevisiae*.

# Transcriptional Repressors

- The *GAL* genes but with the addition of a site between the Gal4-binding sites and the promoter: this is where, in the presence of glucose, Mig1 binds and switches off the *GAL* genes.



# Transcriptional Repressors

- Thus, just as in *E. coli*, the cell only makes the enzymes needed to metabolize galactose if the preferred energy source, glucose, is not present.
- How does Mig1 repress the *GAL* genes?

# Transcriptional Repressors

- Mig1 recruits a “repressing complex” containing the Tup1 protein.
- This complex is recruited by many yeast DNA-binding proteins that repress transcription, including the  $\alpha 2$  protein involved in controlling the mating-type specific genes.



# Transcriptional Repressors

- Tup1 also has counterparts in mammalian cells.
- Two mechanisms have been proposed to explain the repressing effect of Tup1.

# Transcriptional Repressors

- First, Tup1 acts on nucleosomes either through recruiting histone deacetylases and/or by positioning a nucleosome at or near the transcription start site.

# Transcriptional Repressors

- Second, Tup1 interacts directly with the transcriptional machinery at the promoter and inhibits initiation.

# Signal Transduction Pathways

- As we have seen, whether or not a given gene is expressed very often depends on environmental signals.
- Signals come in many forms:-

# Signal Transduction Pathways

- They can be small molecules such as sugars as in bacteria, but they can also be proteins released by one cell and received by another as in multicellular organisms.

# Signal Transduction Pathways

- There are various ways that signals are detected by a cell and communicated to a gene.
- In bacteria, we saw that signals control the activities of regulators by inducing allosteric changes in those regulators.

# Signal Transduction Pathways

- Often, this effect is direct: a small molecular signal, such as a sugar, enters the cell and binds the transcriptional regulator directly.
- But we saw one example where the effect of the signal is indirect (control of the activator NtrC).



# Signal Transduction Pathways

- In that case, the signal (low ammonia levels) induces a kinase that phosphorylates NtrC.
- This type of indirect signaling is an example of a **signal transduction pathway**.

# Signal Transduction Pathways

- The term **signal** refers to the initiating ligand itself—the sugar or protein, for example.

# Signal Transduction Pathways

- It can also refer to the “information” as it passes from detection of that ligand to the regulators that directly control the genes — that is, as it passes along a signal transduction pathway.

# Signal Transduction Pathways

- In the simplest of bacterial cases, there was no distinction, of course, but once a signal transduction pathway is involved, there is.

# Signal Transduction Pathways

- In addition, in eukaryotes, that most signals are communicated to genes through signal transduction pathways, sometimes very elaborate ones.

# Signal Transduction Pathways

- In a signal transduction pathway, the initiating ligand is typically detected by a specific **cell surface receptor**: the ligand binds to an extracellular domain of the receptor, and this binding is communicated to the intracellular domain.

# Signal Transduction Pathways

- From there, the signal is relayed to the relevant transcriptional regulator, often through a cascade of kinases.



# Signal Transduction Pathways

- How is the binding of ligand to the extracellular domain communicated to the intracellular domain?

# Signal Transduction Pathways

- This can be through an allosteric change in the receptor, whereby binding of ligand alters the shape (and thus activity) of the intracellular domain.

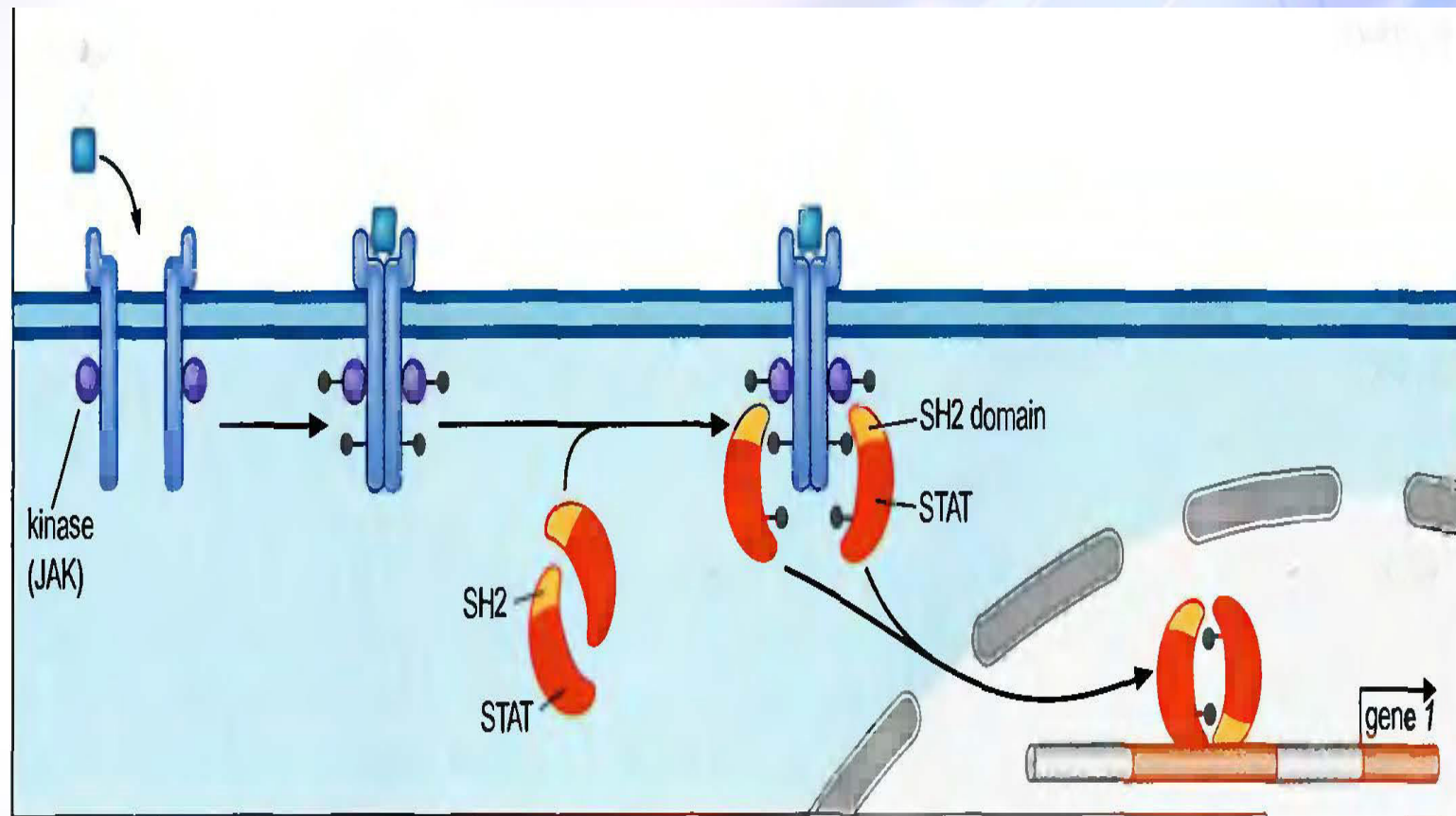
# Signal Transduction Pathways

- Alternatively, the ligand can act simply to bring together two or more receptor chains, allowing interactions between the intracellular domains of those receptors to activate each other.

# Signal Transduction Pathways

- The **STAT** (signal transducer and activator of transcription) pathway is a relatively simple example of signal transduction pathway.
- In this example, a kinase is bound to the intracellular domain of a receptor.

# Signal Transduction Pathways



# Signal Transduction Pathways

- When the receptor is activated by its ligand (a cytokine), it brings together two receptor chains and triggers the kinase in each chain to phosphorylate a particular sequence in the intracellular domain of the opposing receptor.



# Signal Transduction Pathways

- This phosphorylated site is then recognized by a particular STAT protein that, once bound, gets phosphorylated itself.
- Once phosphorylated, the STAT dimerizes, moves to the nucleus, and binds DNA.



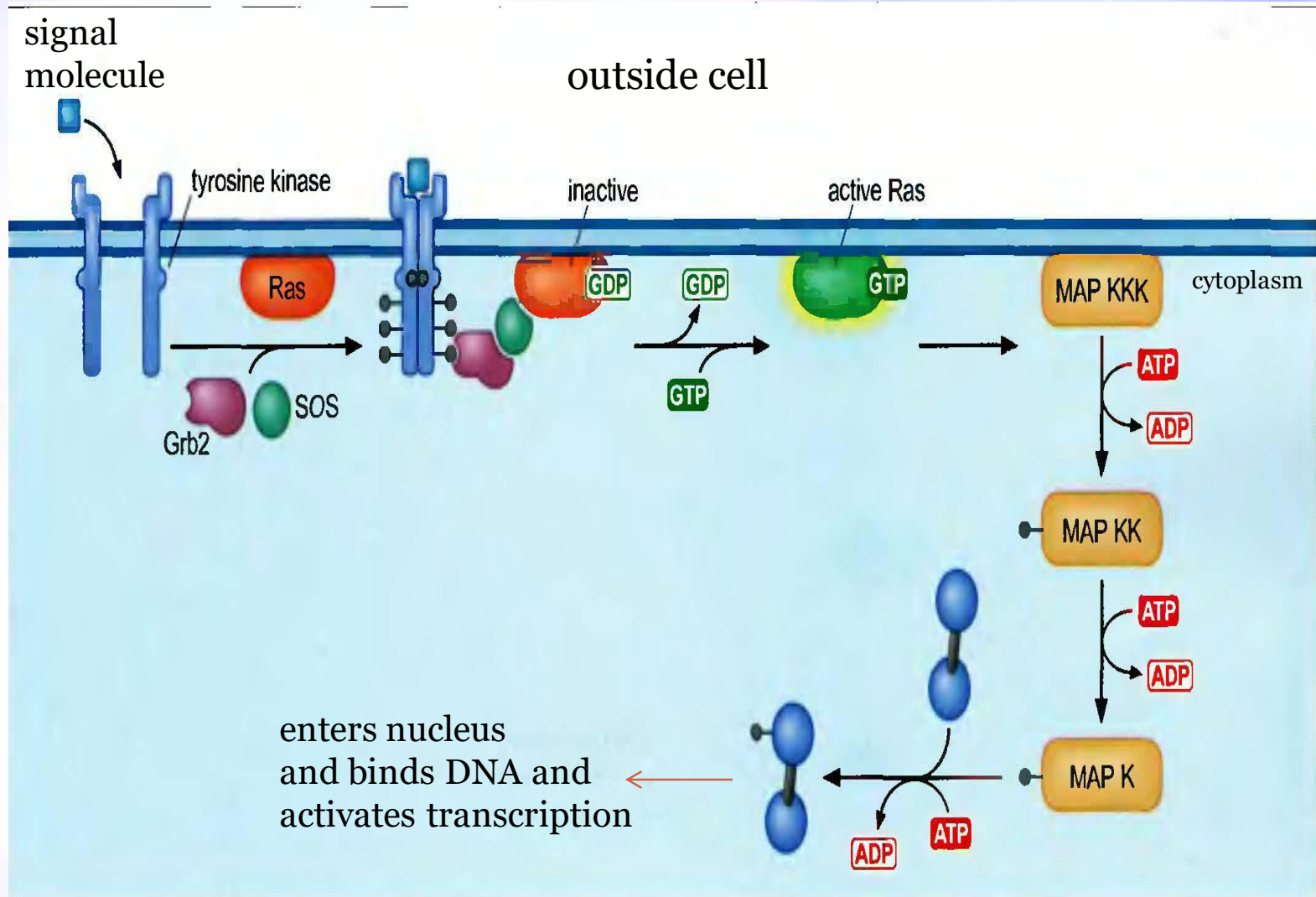
# Signal Transduction Pathways

- Another example is the mitogen-activated protein kinase (MAPK) pathway that controls activators such as Jun, one of the activators that works at the interferon- $\beta$  enhancer.

# Signal Transduction Pathways

- In this case, the activated receptor induces a cascade of signaling events, ending in activation of an MAPK that phosphorylates Jun (and other transcriptional regulators).

# Signal Transduction Pathways



# Signal Transduction Pathways

- The most common way in which information is passed through signal transduction pathways is via phosphorylation, but proteolysis, dephosphorylation, and other modifications are also used.

# Plasmids as Vectors

- **Plasmids** are autonomous self-replicating molecules of DNA (or very rarely RNA).
- They are not chromosomes, although they do reside inside living cells and carry genetic information.

# Plasmids as Vectors

- They are not regarded as part of the cell's genome for two reasons.
- First, a particular plasmid may be found in cells of different species and may move from one host species to another.



# Plasmids as Vectors

- Second, a plasmid may sometimes be present and sometimes absent from the cells of a particular host species.
- They are not needed for cell growth and division under normal conditions.



# Plasmids as Vectors

- The DNA segment to be cloned is first linked to a **vector DNA**, which is a vehicle for carrying foreign DNA into a suitable host cell, such as the bacterium *E. coli*.
- Such carriers are called **vectors**.

# Plasmids as Vectors

- Since the mid-1970s, many vectors have been developed which mainly fall into two major classes:-
  - Plasmids and
  - Phages.

# Plasmids as Vectors

- Bacteriophages lambda and M 13 are commonly used vectors and have the advantage of being able to accept large pieces (up to 40 and 10 kb, respectively) of foreign DNA.

# Plasmids as Vectors

- However, the most commonly used vectors for transferring foreign DNA into cells are bacterial plasmids, which accept smaller pieces (5-10 kb) of foreign DNA.

# Plasmids as Vectors

- A typical plasmid vector is a relatively small circular DNA molecule that contains its own origin of replication (*ori*) so it can replicate inside a host cell independent of the host chromosome.

# Plasmids as Vectors

- Although plasmids occur naturally in bacteria, where they usually confer useful properties on the cell such as antibiotic resistance,

# Plasmids as Vectors

- yet, most plasmid vectors in use today have been specially modified or constructed for the purpose of transferring foreign DNA into cells.



# Plasmids as Vectors

- Some, such as *pUC* 18 and the Bluescript vectors, have been engineered to make many hundreds of copies of themselves, and any foreign DNA they contain, once they are introduced into a host cell.

# Plasmids as Vectors

- This process of producing a large quantity of identical copies of any chosen DNA is referred to as **cloning**, and the cells derived from a host cell containing a recombinant plasmid is referred to as a **cell clone**.

# Plasmids as Vectors

- For cloning to work, the vector must have a cutting site for a restriction enzyme that makes it possible to insert the desired fragment of DNA and, second, the vector must be selectable so that cells that have taken up a vector can be identified.

# Plasmids as Vectors

- To overcome the difficulty of inserting fragments into vectors, most plasmids have been constructed with polylinkers (multiple cloning sites)-pieces of DNA that contain one or more unique restriction enzyme recognition sites.

# Plasmids as Vectors

- Moreover, most plasmids contain antibiotic-resistance genes or other **selectable Markers** that enable cells containing plasmids to be isolated from cells that have not taken up a plasmid.

# Plasmids as Vectors

- In the early years of the cloning era, Boyer and his colleagues developed a set of very popular vectors known as the pBR plasmid series.



# Plasmids as Vectors

- Nowadays, one can choose from many plasmid cloning vectors besides the pBR plasmids.
- One useful, though somewhat dated, class of plasmids is the *pUC* series.



# Plasmids as Vectors

- Nowadays, one can choose from many plasmid cloning vectors besides the pBR plasmids.
- One useful, though somewhat dated, class of plasmids is the *pUC* series.

# Plasmids as Vectors

- One of the most notable plasmids, termed pBR322 after its developers Bolivar and Rodriguez (pBR), was widely adopted and illustrates its desirable features of a cloning vector as indicated below;-

# Plasmids as Vectors

- The plasmid is much smaller than a natural plasmid which makes it more resistant to damage by shearing and increases the efficiency of uptake by bacteria, a process termed transformation.

# Plasmids as Vectors

- A bacterial origin of DNA replication ensures that the plasmid will be replicated by the host cell.
- Most plasmids, including pBR322, have a relaxed origin of replication and which means replication is not tightly linked to cell division.

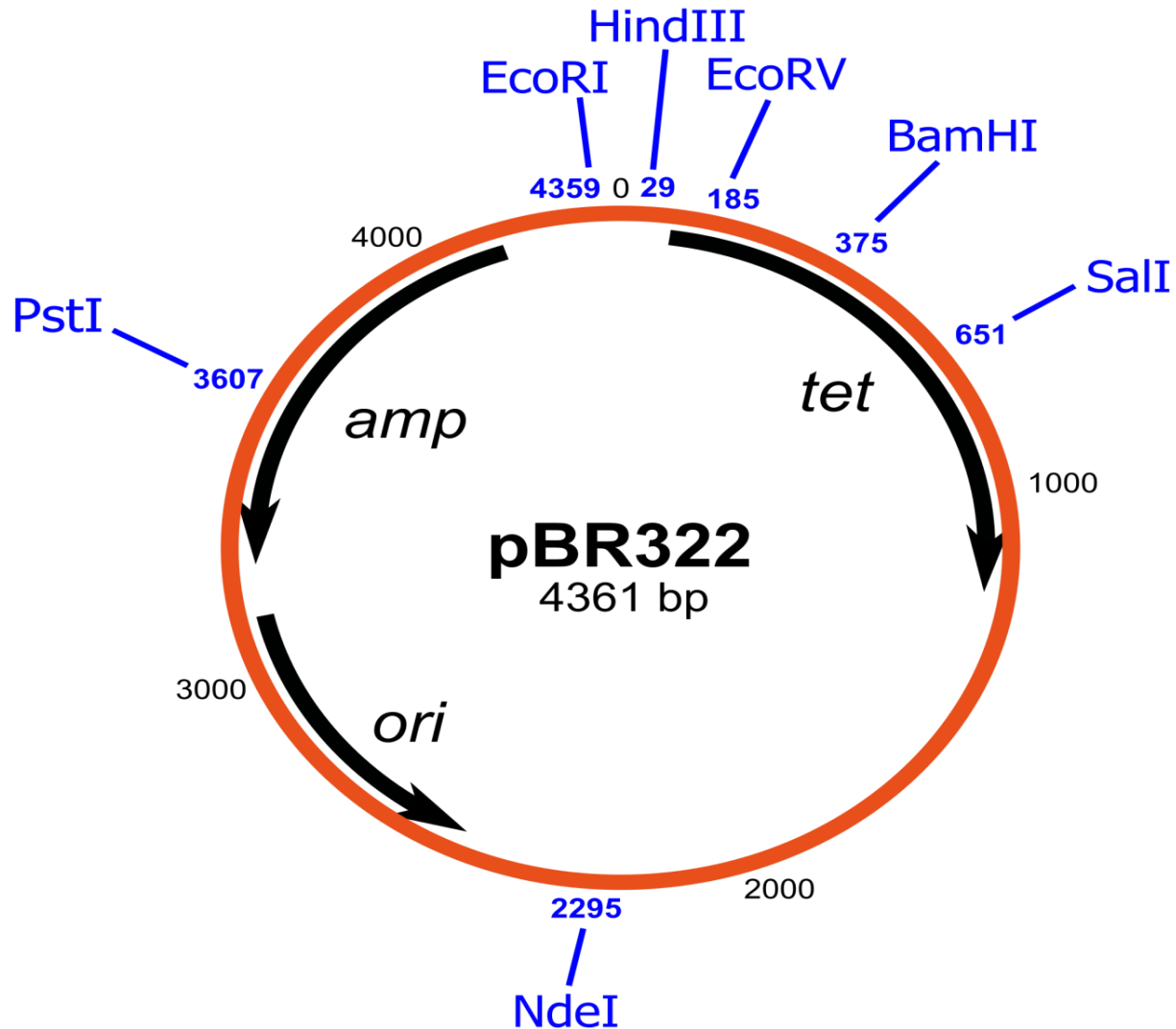
# Plasmids as Vectors

- The genes coding for resistance to antibiotics have been introduced which allow the selection of cells which contain plasmid.

# Plasmids as Vectors

- There are single recognition sites for a number of restriction enzymes at various points around the plasmid which can be used to open or linearize the circular plasmid.

# Plasmids as Vectors





# Plasmid Selection Systems

- Insertional inactivation is a useful selection method for identifying recombinant vectors with inserts. For example, a fragment of chromosomal DNA digested with *Bam*H1 would be isolated and purified.

# Plasmid Selection Systems

- The plasmid pBR322 would also be digested at a single site, using *Bam*H1, after which both samples would be deproteinized to inactivate the restriction enzyme.

# Plasmid Selection Systems

- BamH 1 cleaves to give sticky ends, and so it is possible to obtain ligation between the plasmid and digested DNA fragments in the presence of T4 DNA ligase.

# Plasmid Selection Systems

- The products of this ligation will include plasmid containing a single fragment of the DNA as an insert, but there will also be unwanted products,

# Plasmid Selection Systems

- such as plasmid which has recircularized without an insert, dimers of plasmid, fragments joined to each other, and plasmid with an insert composed of more than one fragment.

# Plasmid Selection Systems

- Most of these unwanted molecules can be eliminated during subsequent steps. The products of such reactions are usually identified by agarose gel electrophoresis.

# Plasmid Selection Systems

- The ligated DNA must now be used to transform *E. coli*. Bacteria do not normally take up DNA from their surroundings, but can be induced to do so by prior treatment with  $\text{Ca}^{2+}$  at 4°C.



# Plasmid Selection Systems

- These cells are then termed **competent**, since DNA added to the suspension of cells will be taken up during a brief increase in temperature, termed a **heat shock**.
- Plasmids can also be introduced into the bacterial cells by electroporation.

# Plasmid Selection Systems

- In this process the cells are subjected to pulses of a high voltage gradient, causing many of them to take up DNA from the surrounding solution.
- This technique has proved to be very useful.

# Plasmid Selection Systems

- After a brief incubation to allow expression of the antibiotic resistance genes the cells are plated onto medium containing the antibiotic, e.g. Ampicillin.

# Plasmid Selection Systems

- Colonies which grow on these plates must be derived from cells which contain plasmid, since this carries the gene for resistance to ampicillin.
- Next the colonies are replica plated, using a sterile velvet pad, onto plates containing tetracycline in their medium.

# Plasmid Selection Systems

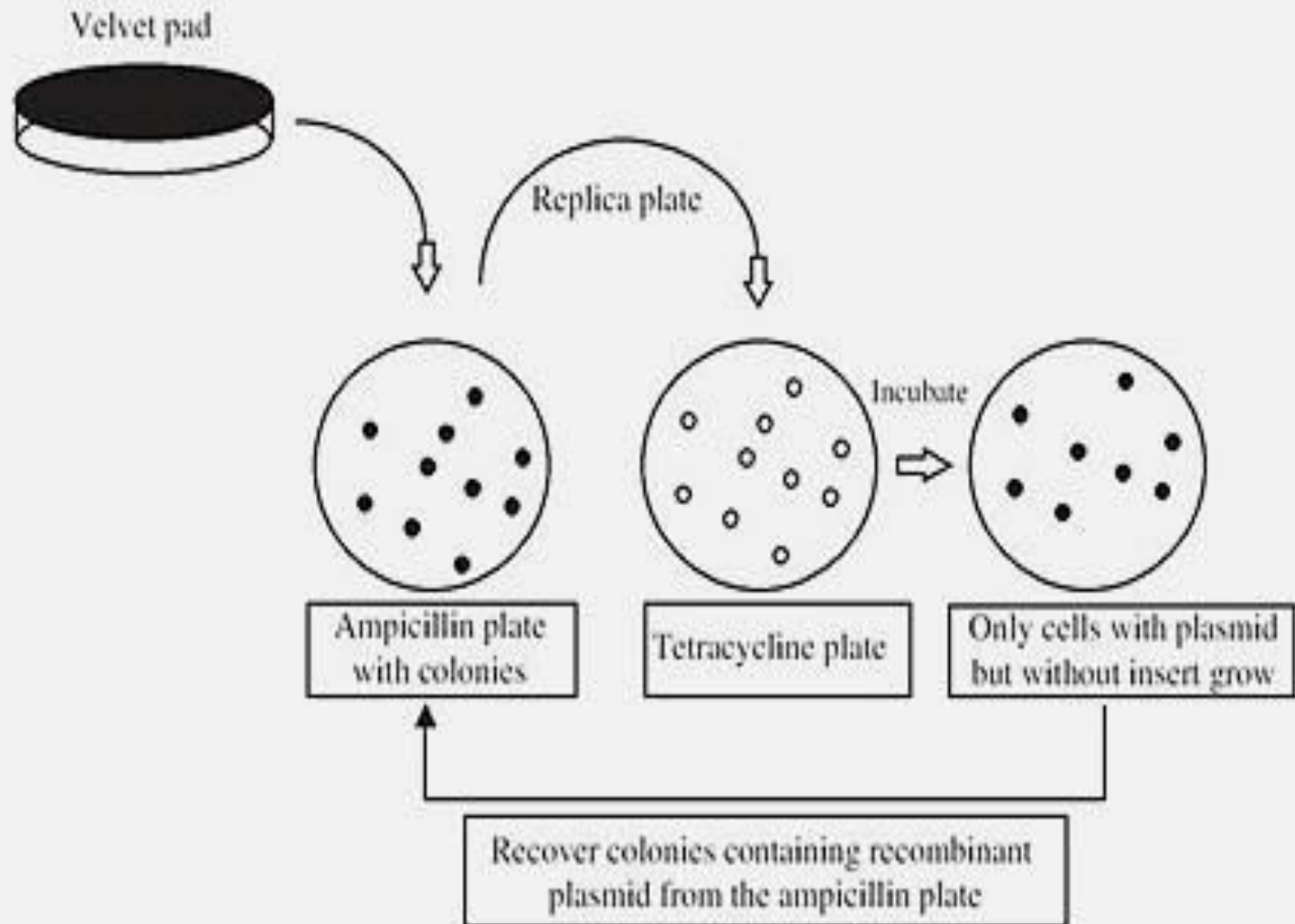
- Since the *Bam*H1 site lies within the tetracycline resistance gene, this gene will be inactivated by the presence of insert, but will be intact in those plasmids which have merely recircularized.

# Plasmid Selection Systems

- Thus colonies which grow on ampicillin but not on tetracycline must contain plasmids with inserts.
- This illustrates the importance of a second gene for antibiotic resistance in a vector.



# Plasmid Selection Systems





# Plasmid Selection Systems

- If the digested plasmid is treated with the enzyme alkaline phosphatase prior to ligation, recircularization will be prevented, since this enzyme removes the 5' phosphate groups which are essential for ligation.

END

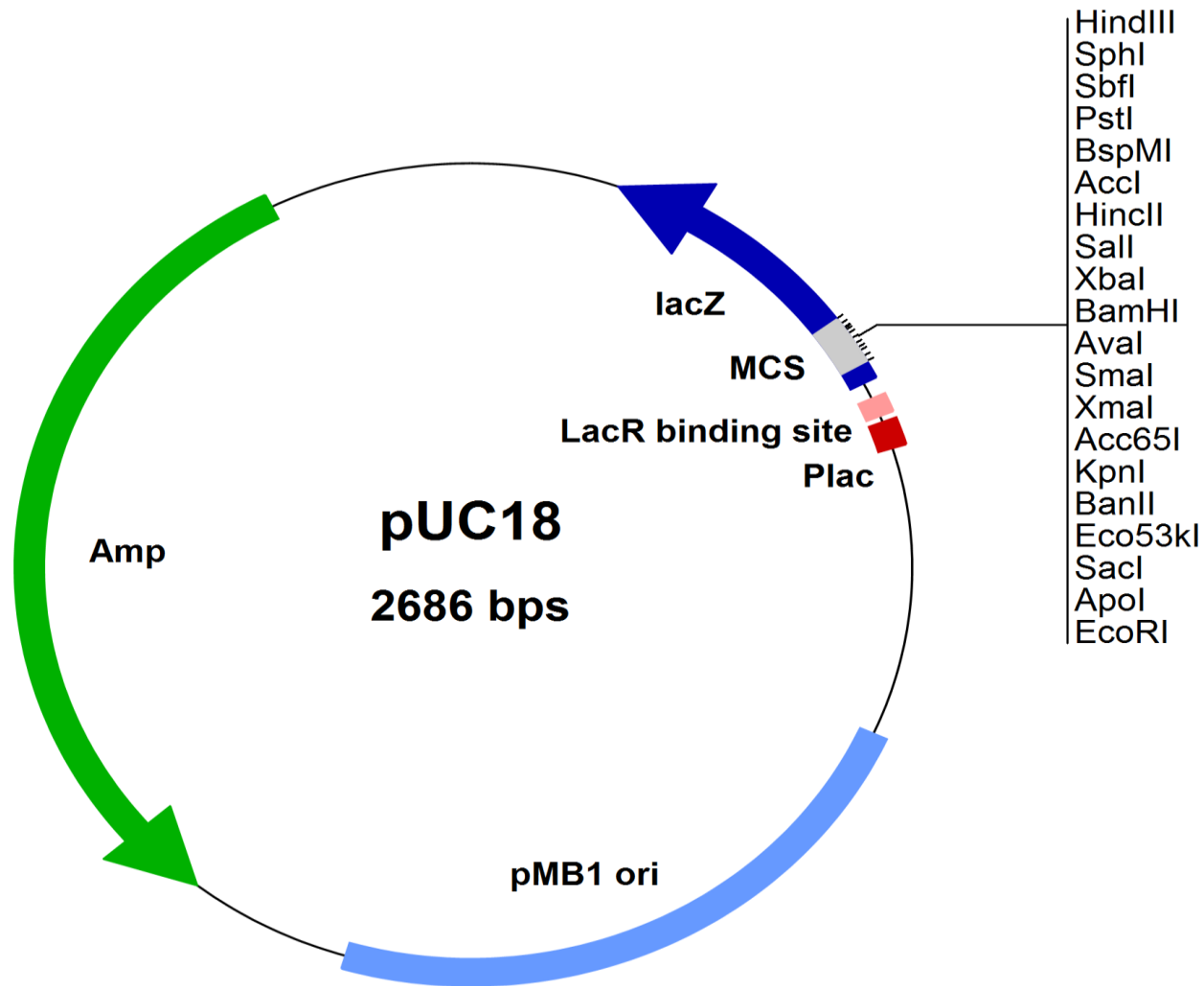
# *pUC* Plasmid Cloning Vectors

- The valuable features of pBR322 have been enhanced by the construction of a series of plasmids termed *pUC* (produced at the University of California)

# ***pUC* Plasmid Cloning Vectors**

- These plasmids are based on pBR322, from which about 40% of the DNA has been deleted.

# pUC Plasmid Cloning Vectors



# *pUC* Plasmid Cloning Vectors

- There is an antibiotic resistance gene for ampicillin and origin of replication for *E. coli*.
- *coli*.
- In addition the most popular restriction sites are concentrated into a region termed the **multiple cloning site** (MCS).

# *pUC* Plasmid Cloning Vectors

- Moreover, the MCS is part of a gene (*LacZ*) in its own right and codes for a portion of a polypeptide called  $\beta$ -galactosidase.

# *pUC* Plasmid Cloning Vectors

- When the *pUC* plasmid has been used to transform the host cell *E. coli* the gene may be switched on by adding the inducer IPTG (isopropyl- $\beta$ -D-thiogalactopyranoside)



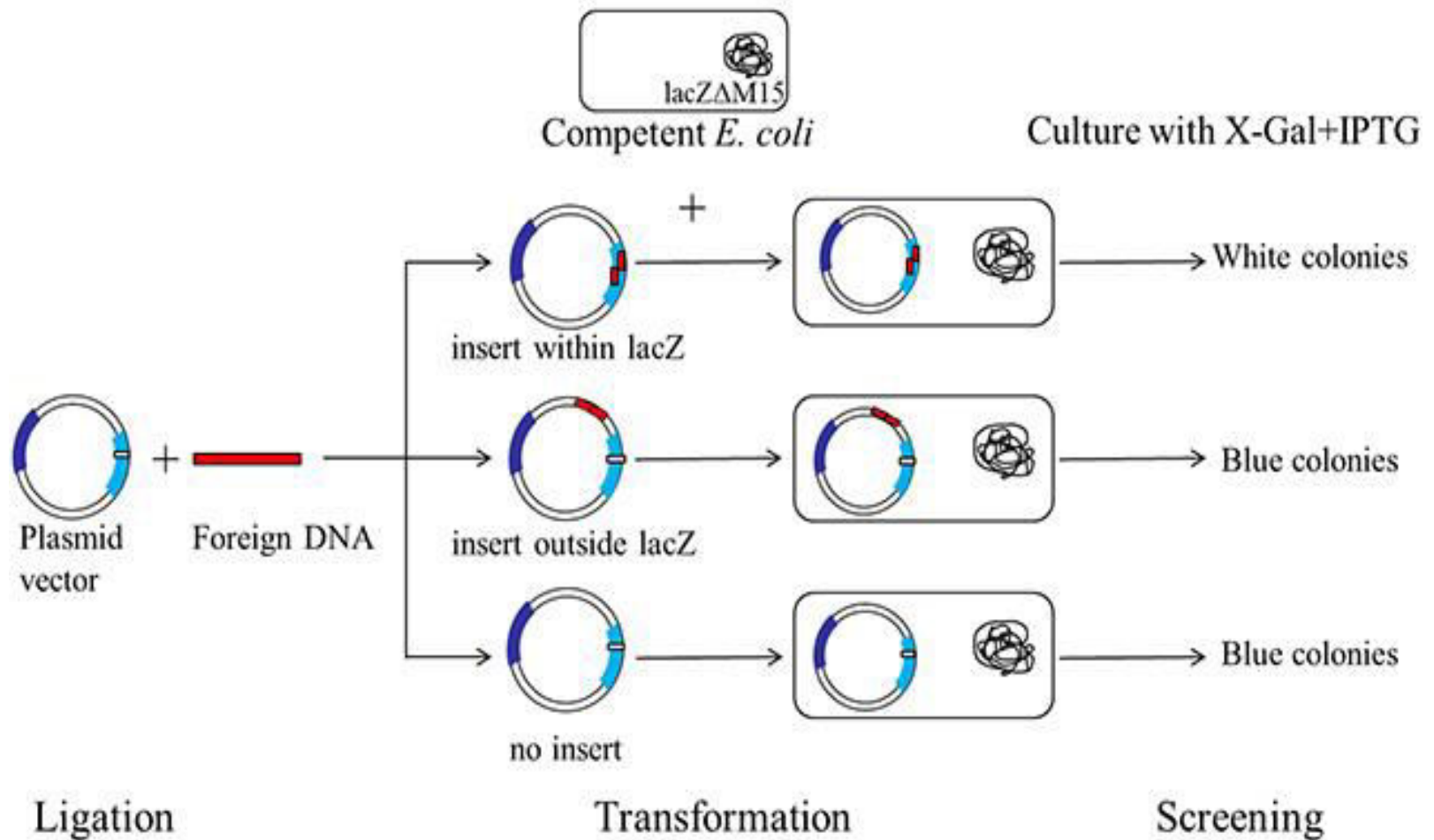
# ***pUC* Plasmid Cloning Vectors**

- Its presence causes the enzyme  $\beta$ -galactosidase to be produced.
- The functional enzyme is able to hydrolyse a colourless substance called X-gal into a blue insoluble material.

# ***pUC* Plasmid Cloning Vectors**

- However if the gene is disrupted by the insertion of a foreign fragment of DNA, a nonfunctional enzyme results which is unable to carry out hydrolysis of Xgal.

# pUC Plasmid Cloning Vectors



# *pUC* Plasmid Cloning Vectors

- Thus, a recombinant *pUC* plasmid may be easily detected since it is white or colourless in the presence of X-gal, whereas an intact non-recombinant *pUC* plasmid will be blue as its gene is fully functional and not disrupted.

# *pUC* Plasmid Cloning Vectors

- This elegant system, termed **blue/white selection**, allows the initial identification of recombinants to be undertaken very quickly and has been included in a number of subsequent vector systems.

# *pUC* Plasmid Cloning Vectors

- This selection method and insertional inactivation of antibiotic resistance genes do not, however, provide any information on the character of the DNA insert, only the status of the vector.

END



# Virus-based Cloning Vectors

- A useful feature of any cloning vector is the amount of DNA it may accept or may be inserted before it becomes unviable.



# Virus-based Cloning Vectors

- Inserts greater than 5 kb in size increase plasmid size to the point at which efficient transformation of bacterial cells decreases markedly.

# Virus-based Cloning Vectors

- Thus, bacteriophages (bacterial viruses) have been adapted as vectors in order to propagate larger fragments of DNA in bacterial cells.

# Virus-based Cloning Vectors

- Cloning vectors derived from one bacteriophage are commonly used since they offer approximately 1 6-fold advantage in cloning efficiency in comparison with the most efficient plasmid cloning vectors.

# Virus-based Cloning Vectors

- Bacteriophage  $\lambda$  is a linear double-stranded phage approximately 49 kb in length. It infects *E. coli* with great efficiency by injecting its DNA through the cell membrane.

# Virus-based Cloning Vectors

- In the wild-type phage  $\lambda$  the DNA follows one of two possible modes of replication.
- Firstly, it may enter the lysogenic life cycle. Alternatively, it may follow a lytic life cycle.

# Virus-based Cloning Vectors

- At the extreme ends of phage DNA, are 12 bp sequences termed **cos (cohesive) sites**. Although they are asymmetrical, they are similar to restriction sites and allow the phage DNA to be circularized.



# Virus-based Cloning Vectors

- Phage may be replicated very efficiently in this way, resulting in concatemers of many phage genomes which are cleaved at the cos sites and inserted into newly formed phage protein heads.



# Virus-based Cloning Vectors

- Phage  $\lambda$  has been used extensively in the production of gene libraries, mainly because of its efficient entry into the *E. coli* cell and the fact that larger fragments of DNA may be stably integrated.

# Virus-based Cloning Vectors

- For the cloning of long DNA fragments, up to approximately 25 kb, much of the nonessential  $\lambda$  DNA that codes for the lysogenic life cycle is removed and replaced by the foreign DNA insert.

# Virus-based Cloning Vectors

- The recombinant phage is then assembled into pre-formed viral protein particles, a process known as **in vitro packaging**.
- These newly formed phages are used to infect bacterial cells which have been plated out on agar.

# Virus-based Cloning Vectors

- Once inside the host cells, the recombinant viral DNA is replicated.
- All the genes needed for normal lytic growth are still present in the phage DNA.

# Virus-based Cloning Vectors

- So multiplication of the virus takes place by cycles of cell lysis and infection of surrounding cells, giving rise to plaques of lysed cells on a background, or lawn, of bacterial cells.

# Virus-based Cloning Vectors

- The viral DNA including the cloned foreign DNA can be recovered from the viruses from these plaques and analysed further by restriction mapping and agarose gel electrophoresis.

END

# $\lambda$ Phage Vectors

- Fred Blattner and his colleagues constructed the first phage vectors by modifying the well-known  $\lambda$  phage.



# $\lambda$ Phage Vectors

- They took out the region in the middle of the phage DNA, but retained the genes needed for phage replication. The missing phage genes could then be replaced with foreign DNA.

# $\lambda$ Phage Vectors

- Blattner named these vectors **Charon phages** after Charon, the boatman on the river Styx in classical mythology.
- In general, two types of  $\lambda$  phage vectors have been developed,  **$\lambda$  insertion vectors** and  **$\lambda$  replacement vectors**.

# $\lambda$ Phage Vectors

- One clear advantage of the  $\lambda$  phages over plasmid vectors is that they can accommodate much more foreign DNA.
- For example, Charon 4 can accept up to about 20 kb of DNA.

# $\lambda$ Phage Vectors

- $\lambda$  insertion vectors accept less DNA than the replacement type since the foreign DNA is simply inserted into a region of the phage genome with appropriate restriction sites; common examples are  $\lambda$  gt10 and  $\lambda$  charon16A.

# $\lambda$ Phage Vectors

- In a replacement vector, a central region of DNA not essential for lytic growth is removed (a stuffer fragment) by a double digestion with, for example, *EcoRI* and *BamHI*.

# $\lambda$ Phage Vectors

- This leaves two DNA fragments, called right and left arms.
- The central stuffer fragment is replaced by inserting foreign DNA between the arms to form a functional recombinant  $\lambda$ phage.



# $\lambda$ Phage Vectors

- The most notable examples of  $\lambda$  replacement vectors are  $\lambda$ EMBL and  $\lambda$ ZAP.
- $\lambda$ ZAP is a commercially produced cloning vector which includes unique cloning sites clustered into a multiple cloning site (MCS).



# $\lambda$ Phage Vectors

- Furthermore, the MCS is located within a *lacZ* region providing a blue/white screening system based on insertional inactivation.
- It is also possible to express foreign cloned DNA from this vector.

# $\lambda$ Phage Vectors

- This is a very useful feature of some  $\lambda$  vectors since it is then possible to screen for protein product rather than the DNA inserted into the vector.
- This screening is undertaken with antibody probes directed against the protein of interest.

# $\lambda$ Phage Vectors

- Another feature that make this a useful cloning vector is the ability to produce RNA transcripts, termed cRNA or **riboprobes**.

# $\lambda$ Phage Vectors

- This is possible because two promoters for RNA polymerase enzymes exist in the vector, a T7 and a T3 promoter, which flank the MCS.

# $\lambda$ Phage Vectors

- One of the most useful features of  $\lambda$ ZAP is that it has been designed to allow automatic in vivo excision of a small 2.9 kb colony-producing vector, the phagemid pBluescript SK.

# $\lambda$ Phage Vectors

- This technique is sometimes called single-stranded DNA rescue and occurs as the result of a ‘superinfection’ process where helper phages are added to the cells which are grown for an additional period of approximately 4 hours.

# $\lambda$ Phage Vectors

- A common use for  $\lambda$  replacement vectors is in constructing genomic libraries.
- Suppose we wanted to clone the entire human genome.



# $\lambda$ Phage Vectors

- A common use for  $\lambda$  replacement vectors is in constructing genomic libraries.
- Suppose we wanted to clone the entire human genome.
- This would obviously require a great many clones.

# $\lambda$ Phage Vectors

- So the larger the insert in each clone, the fewer total clones would be needed.
- In fact, such genomic libraries have been constructed for the human genome and for genomes of a variety of other organisms.

END

# M13 Phage Vectors

- Another phage used as a cloning vector is the filamentous (long, thin, filament-like) phage M13.
- M13 is a filamentous coliphage with a genome composed of single-stranded circular DNA.

# M13 Phage Vectors

## M13 Bacteriophage



# M13 Phage Vectors

- Joachim Messing and his co-workers endowed the phage DNA with the same  $\beta$ -galactosidase gene fragment and multiple cloning sites found in the *pUC* family of vectors.

# M13 Phage Vectors

- In fact, the M13 vectors were engineered first; then the useful cloning sites were simply transferred to the *pUC* plasmids.

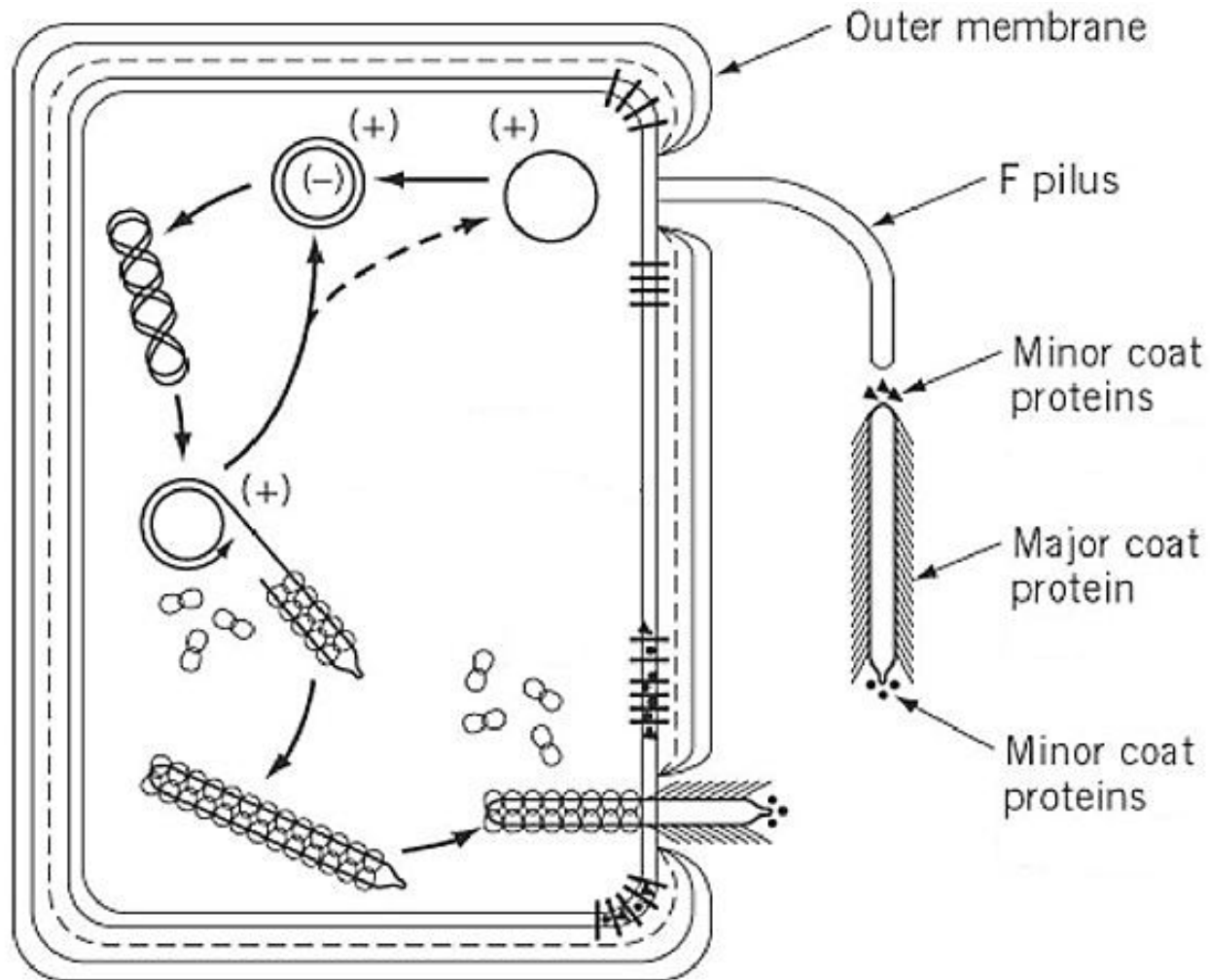


# M13 Phage Vectors

- The DNA in the phage particle itself is single-stranded, but after infecting an *E. coli* cell, the DNA is converted to a double-stranded replicative form (RF).
- This double-stranded replicative form of the phage DNA is used for cloning.



# M13 Phage Vectors



# M13 Phage Vectors

- After it is cut with one or two restriction enzymes at its multiple cloning site, foreign DNA with compatible ends can be inserted.

# M13 Phage Vectors

- This recombinant DNA is then used to transform host cells, giving rise to progeny phages that bear single-stranded recombinant DNA.

# M13 Phage Vectors

- The phage particles, containing phage DNA, are secreted from the transformed cells and can be collected from the growth medium.

# M13 Phage Vectors

- The main factor is that the genome of this phage is a single-stranded DNA, so DNA fragments cloned into this vector can be recovered in single-stranded form.

# M13 Phage Vectors

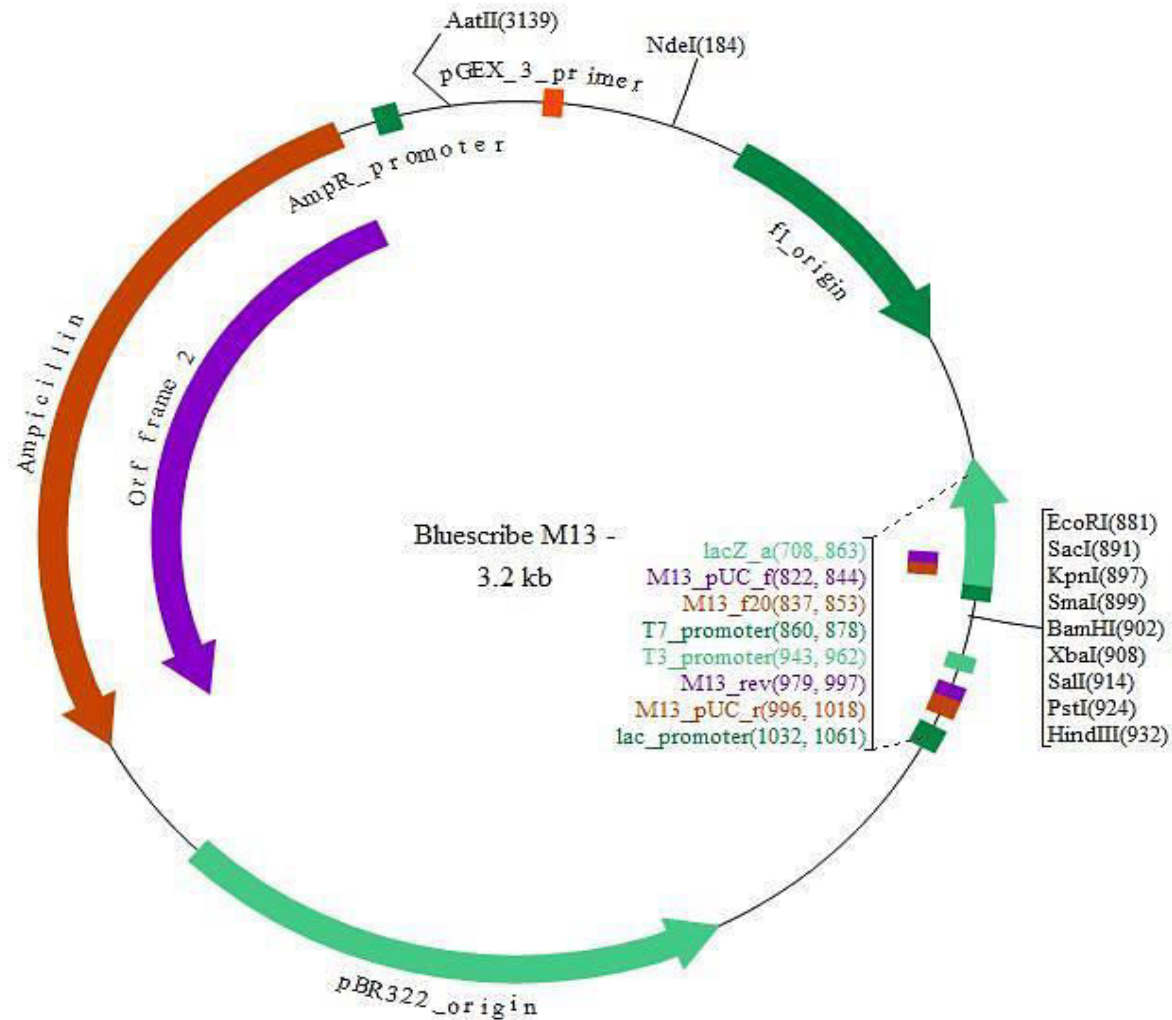
- Single-stranded DNA can be an aid to site-directed mutagenesis, by which we can introduce specific, premeditated alterations into a gene.

# M13 Phage Vectors

- The nature of these vectors makes them ideal for techniques such as chain termination sequencing and *in vitro* mutagenesis, since both methods require single-stranded DNA.



# M13 Phage Vectors



# Cosmids & Phagemids

- Another vector designed especially for cloning large DNA fragments is called a cosmid.
- Cosmids behave both as plasmids and as phages.

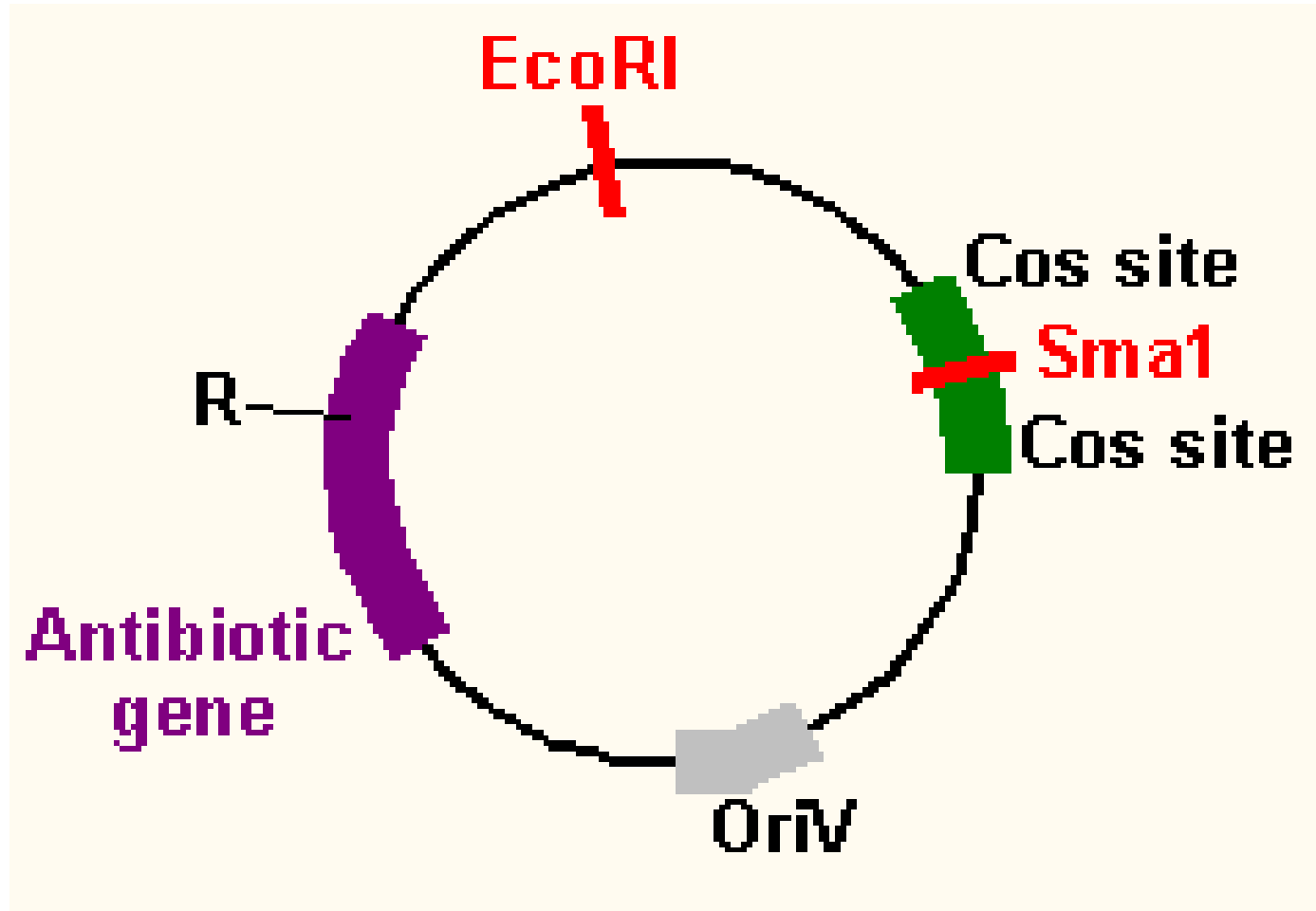
# Cosmids & Phagemids

- They contain the cos sites, or cohesive ends, of  $\lambda$  phage DNA, which allow the DNA to be packaged into  $\lambda$  phage heads (hence the “cos” part of the name “cosmid”).

# Cosmids & Phagemids

- They also contain a plasmid origin of replication, so they can replicate as plasmids in bacteria (hence the “mid” part of the name).

# Cosmids & Phagemids



# Cosmids & Phagemids

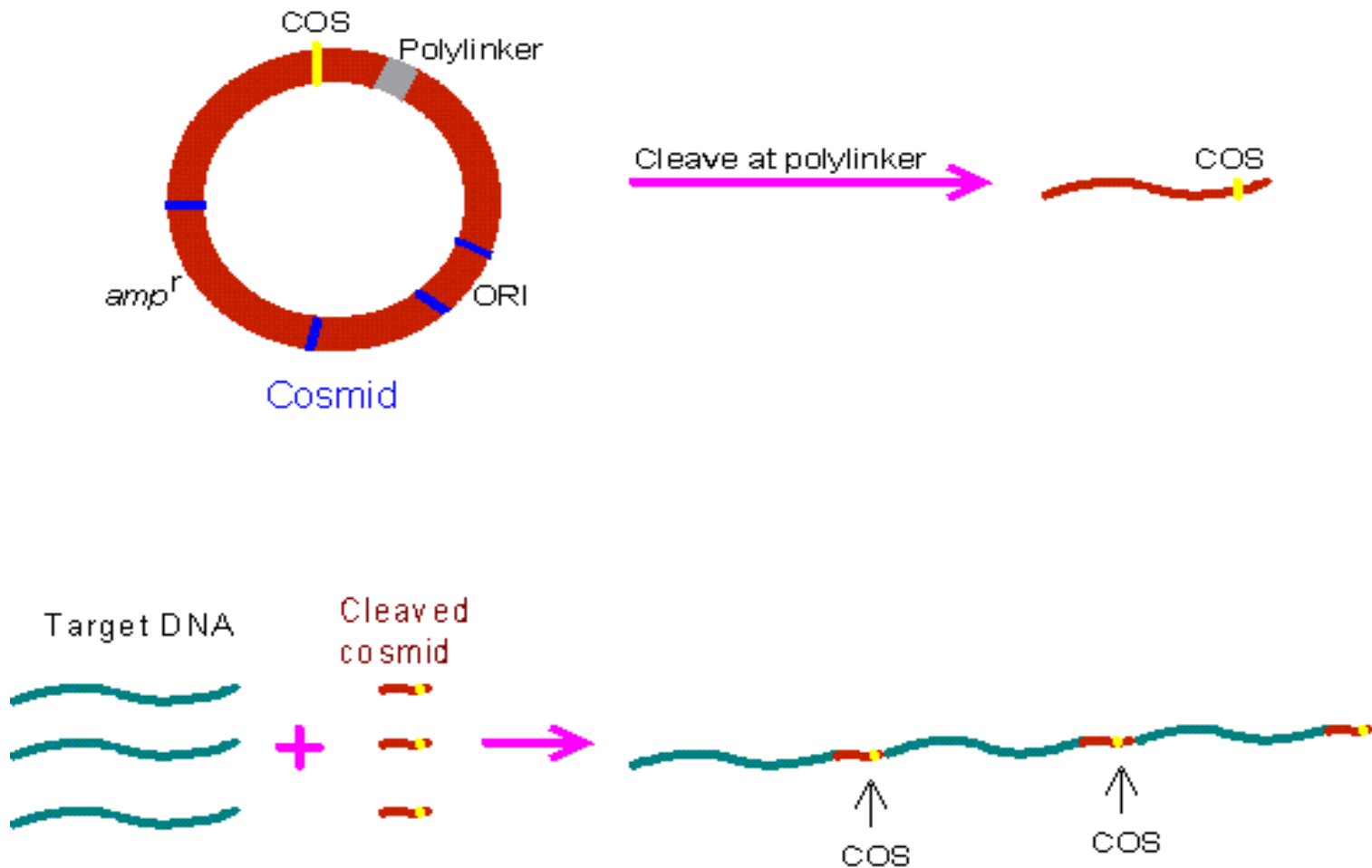
- Because almost the entire  $\lambda$  genome, except for the cos sites, has been removed from the cosmids, they have room for large inserts (40–50 kb).

# Cosmids & Phagemids

- Once these inserts are in place, the recombinant cosmids are packaged into phage particles in vitro.



# Cosmids & Phagemids



# Cosmids & Phagemids

- These particles cannot replicate as phages because they have almost no phage DNA, but they are infectious, so they carry their recombinant DNA into bacterial cells.

# Cosmids & Phagemids

- Once inside the cell, the DNA recircularizes through its cos sites, and from then on behaves exactly like a plasmid and replicate because it has a plasmid origin of replication.

# Cosmids & Phagemids

- Cosmids are especially useful for the analysis of highly complex genomes and are an important part of various genome mapping projects.

# Cosmids & Phagemids

- Another class of vectors that produce single-stranded DNA has also been developed.
- These are like the cosmids in that they have characteristics of both phages and plasmids; thus, they are called **phagemids**.

# Cosmids & Phagemids

- One popular variety goes by the trade name pBluescript (pBS).
- Like the pUC vectors, pBluescript has a multiple cloning site inserted into the *lacZ* gene, so clones with inserts can be distinguished by white versus blue staining with X-gal.

# Cosmids & Phagemids

- This vector also has the origin of replication of the single-stranded phage f1, which is related to M13.



# Cosmids & Phagemids

- This means that a cell harboring a recombinant phagemid, if infected by an f1 helper phage that supplies the single-stranded phage DNA replication machinery, will produce and package single-stranded phagemid DNA.

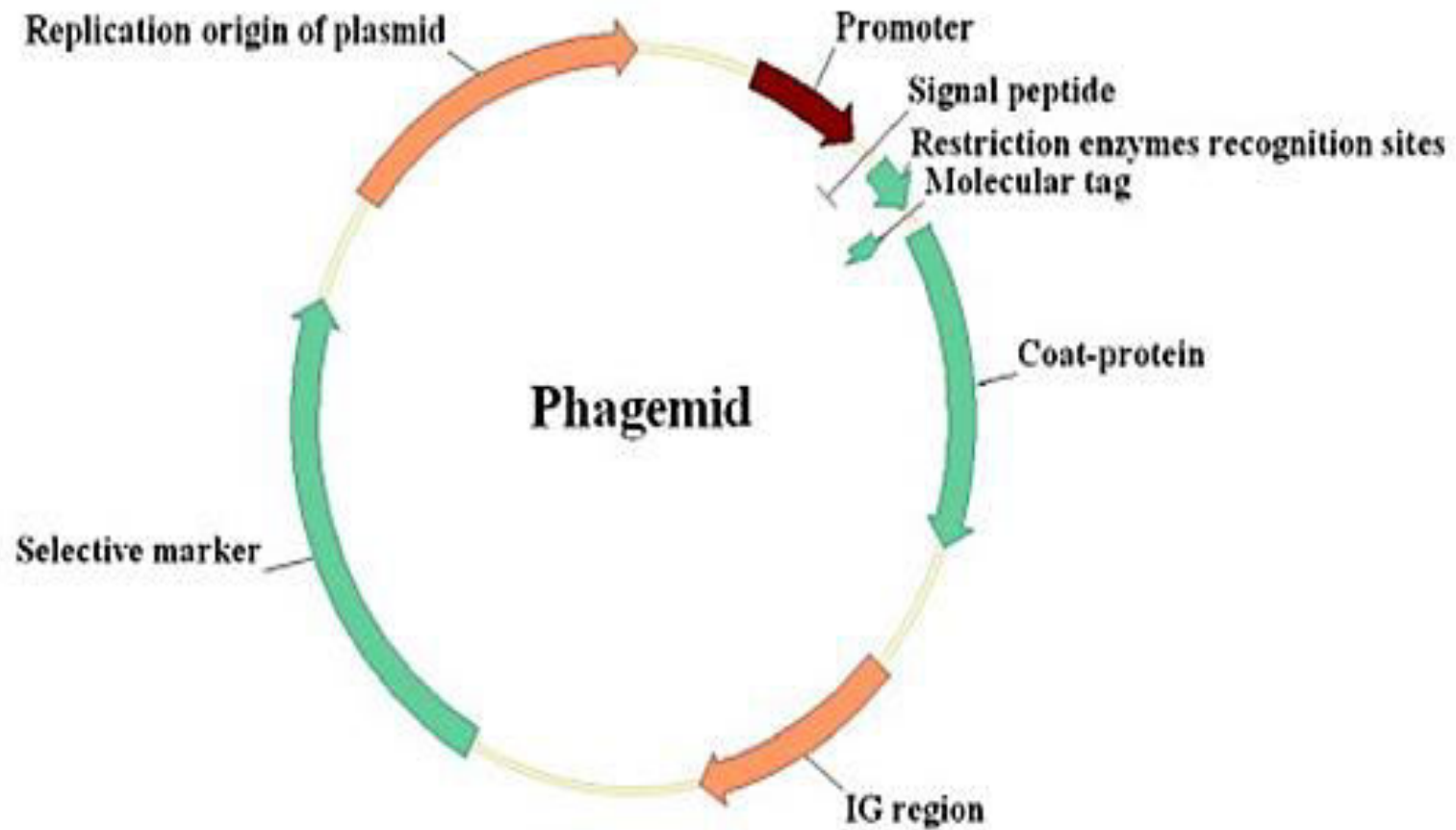
# Cosmids & Phagemids

- A final useful feature of this class of vectors is that the multiple cloning site is flanked by two different phage RNA polymerase promoters.
- For example, pBS has a T3 promoter on one side and a T7 promoter on the other.

# Cosmids & Phagemids

- This allows one to isolate the double-stranded recombinant phagemid DNA and transcribe it in vitro with either of the phage polymerases to produce pure RNA transcripts corresponding to either strand of the insert.

# Cosmids & Phagemids



# Large Insert Capacity Cloning Vectors & YAC vectors

- Vectors which accept larger fragments of DNA than phage A or cosmids have the distinct advantage that fewer clones need to be screened when searching for the foreign DNA of interest.

# Large Insert Capacity Cloning Vectors & YAC vectors

- In addition they have had an enormous impact in the mapping of the genomes of organisms such as the mouse and are used extensively in the human genome mapping project.



# Large Insert Capacity Cloning Vectors & YAC vectors

- Further developments have given rise to the production of large insert capacity vectors based on bacterial and mammalian artificial chromosomes (BACs and MACs) and on the virus P1, P1 artificial chromosomes (PACs).



# Large Insert Capacity Cloning Vectors & YAC vectors

- However, perhaps most significant is the development of vectors based on yeast artificial chromosomes or YACs which could accommodate hundreds of thousands of kilobases each.

# Large Insert Capacity Cloning Vectors & YAC vectors

- YACs containing a megabase or more are known as “**megaYACs**.”
- YACs are linear molecules composed of a centromere, telomere and a replication origin termed an ARS element.

# Large Insert Capacity Cloning Vectors & YAC vectors

- A YAC contains a left and right yeast chromosomal telomere, which are both necessary to protect the chromosome's ends, and a yeast centromere,

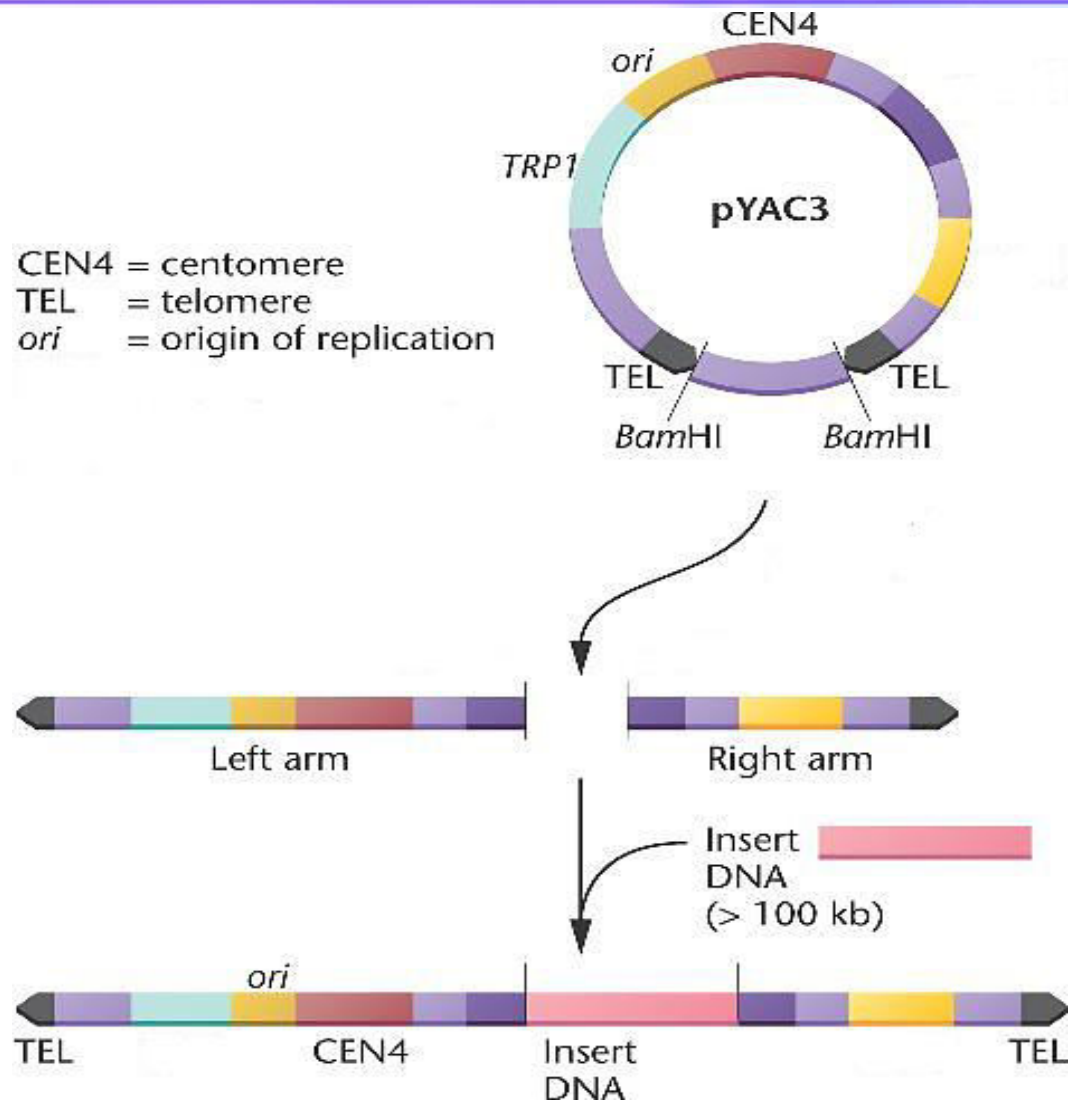
# Large Insert Capacity Cloning Vectors & YAC vectors

- which is necessary for segregation of sister chromatids to opposite poles of the dividing yeast cell.

# Large Insert Capacity Cloning Vectors & YAC vectors

- The centromere is placed adjacent to the left telomere, and a huge piece of human (or any other) DNA can be placed in between the centromere and the right telomere.

# Large Insert Capacity Cloning Vectors & YAC vectors



# Large Insert Capacity Cloning Vectors & YAC vectors

- The large DNA inserts are prepared by slightly digesting long pieces of human DNA with a restriction enzyme.



# Large Insert Capacity Cloning Vectors & YAC vectors

- The YACs, with their huge DNA inserts, can then be introduced into yeast cells, where they will replicate just as if they were normal yeast chromosomes.

# Large Insert Capacity Cloning Vectors & YAC vectors

- Using YACs, geneticists made great strides in the mapping phase of the Human Genome Project.
- They produced a genetic map of the whole genome that provided an average resolution of 0.7 centimorgan.

# Large Insert Capacity Cloning Vectors & YAC vectors

- A centimorgan (cM) is the distance that yields a 1% recombination frequency between two markers and corresponds to an average of about 1 Mb in humans.

# Large Insert Capacity Cloning Vectors & YAC vectors

- These researchers also produced relatively high-resolution physical maps of two of the smallest chromosomes, 21 and Y.

# Large Insert Capacity Cloning Vectors & YAC vectors

- These maps were especially useful in that they represented long stretches of overlapping DNA segments cloned in YACs.

# Large Insert Capacity Cloning Vectors & YAC vectors

- Thus, in the days before the human genome was sequenced, if you were interested in a disease gene that mapped to one of these chromosomes, you had a much simplified task.



# Large Insert Capacity Cloning Vectors & YAC vectors

- You needed only to discover two markers flanking the gene of interest, look on the map to find which YAC or YACs contained these markers, obtain the YACs, and begin your final search for the gene.

END



# Vectors Used in Eukaryotic Cells

- The application of *E. coli* as a general host for cloning and manipulation of DNA is well established.
- However, numerous developments have been made in cloning in eukaryotic cells such as yeasts and plants.

# Vectors Used in Eukaryotic Cells

- Plasmids used for cloning DNA in eukaryotic host cells require a eukaryotic origin of replication and marker genes which will be expressed by these cells.

# Vectors Used in Eukaryotic Cells

- Yeasts, like bacteria, can be grown rapidly, and are therefore well suited for use in cloning.
- These eukaryotic cells also have a natural plasmid, the 2 $\mu$  circle, however this is too large for use in cloning.

# Vectors Used in Eukaryotic Cells

- Plasmids such as the yeast episomal plasmid (YEp) have been produced by manipulation using replication origins from the  $2\mu$  circle, and by incorporating a gene which will complement a gene made defective in the host yeast cell.

# Vectors Used in Eukaryotic Cells

- Thus if a strain of yeast is used which has a defective gene for the biosynthesis of an amino acid, an active copy of that gene on a yeast plasmid can be used as a selectable marker for the presence of that plasmid.

# Vectors Used in Eukaryotic Cells

- Particular use has been made of shuttle vectors which have origins of replication for both yeasts and bacteria such as *E. coli*.



# Vectors Used in Eukaryotic Cells

- This allows the rapid preparation of constructs in *E. coli* which may then be delivered into yeast for expression studies.



# Vectors Used in Eukaryotic Cells

- The bacterium *Agrobacterium tumefaciens* infects plants which have been damaged near soil level, and this infection is often followed by the formation of plant tumours in the vicinity of the infected region.

# Vectors Used in Eukaryotic Cells

- It is well established that *A. tumefaciens* contains the 'Ti' plasmid.
- Part of this plasmid is transferred into the nuclei of plant cells which are infected by the bacterium.

# Vectors Used in Eukaryotic Cells

- Once in the nucleus, this DNA is maintained by integrating with the chromosomal DNA. The integrated DNA carries genes for the synthesis of opines and for tumour induction (hence 'Ti').

# Vectors Used in Eukaryotic Cells

- DNA inserted into the correct region of the Ti plasmid will be transferred to infected plant cells, and in this way it has been possible to clone and express foreign genes in plants.

**END**