Lecture 6 - Mobile Genetic Elements (Learning Objectives)

1. List 4 different mobile genetic elements based on repeats present in the human genome and their characteristics

The overall breakdown of mobile genetic elements is as follows:

1) DNA Transposons: 300,000 copies in the human genome

2) Retrotransposons:
- Non-retroviral elements (**Non-LTR**)
  - SINES:1.5 mil, **Non-autonomous**
  - LINES:850k, **Autonomous**
- Retroviral-like elements (**LTR**): 450k

Mobile Genetic Element	% of Genome	Length	Enzymes Encoded	Moves as
DNA Transposons	3	80-3kb	Transposase	DNA
SINES	13	100-300bp	Does not encode its own functional RT enzyme - rely on other mobile elements (like Line1) for this.	- Requires RNA intermediate - Promoter is within the 5' UTR of the element (and thus is present in the RNA transcript)
LINES	21	6-8kb	Reverse Transcriptase	- RNA intermediate - Promoter is within the 5' UTR of the element (and thus is present in the RNA transcript)
Retroviral-like elements	8	1.5-11kb	Reverse Transcriptase	- RNA intermediate - Promoter lies within LTR repeat

2. Compare the two types of retrotransposons; the **LTR** (long terminal repeats) vs non-LTR

What is characteristic of **retrotransposons** is the need for an **RNA intermediate** during transposition (movement). The **promotor** (P) is responsible for generating the RNA intermediate. The formation of the RNA intermediate requires an enzyme known as **Reverse Transcriptase** - as it must first be transcribed into DNA before integration.

The LTR and the Non-LTR differ in two major ways: 1) Location of the promoter 2) Mechanism of integration

Non-LTR

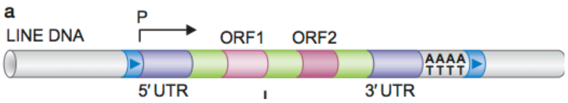
Location of Promotor (P):

Located in the 5’ UTR - which means the promoter is transcribed by the RNA polymerase.

Mechanism of integration:

The ORF1 encodes a 40kDa **RNA binding protein**, which will bind to the retrotransposon, and target DNA where integration is to take place.

The ORF2 will encoded a 150kDa protein, a **RT** with **endonuclease** and **integrase** activities.



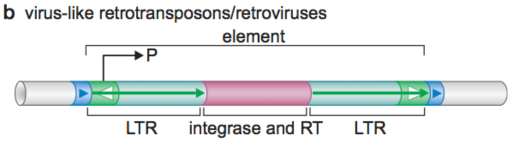
LTR (or Retroviral-like)

Location of Promotor (P):

Is located in the LTR sequence of the element.

Mechanism of integration:

The **RT** enzyme is encoded in the same region as the **integrase**, an enzyme required for DNA integration with the host. No **RNA binding protein**.



So the main difference between the LTR vs non-LTR is the **location of the promoter** as well as the **mechanism of integration**

3. List the sequences included in a LINE1 retrotransposon and briefly define their roles.

Recall that a LINE retrotransposon is part of the Non-LTR retrotransposon. Also remember that retrotransposons use a RNA intermediate for transposition.

5’UTR

This is where transcription will begin. So it is the **promotor** region for LINE1. This promotor has NO TATA box.

ORF1

Encodes a 40kDA **RNA binding protein**: binds to both the RNA intermediate and the host genome to facilitate integration.

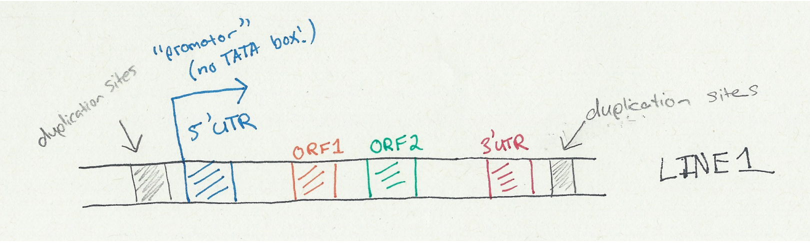
ORF2

Encodes a 150kDA **protein** required for:

- **Reverse Transcriptase**
- **Integrase Activity**
- **Endonuclease Activity**

Duplication sites

Necessary for insertion into the **host genome**.- more info on these.



The transposition is accompanied by a mechanism we call **replicative transposition** - which means the number of copies will be **multiplied** during the event of transposition.

So essentially the RNA is generated from the mobile element from the 5’UTR and this intermediate RNA is then converted into DNA.

4. **List the basic sequences found in a DNA transposon, and indicate a significant difference between retro and DNA transposons.

1) Transposase

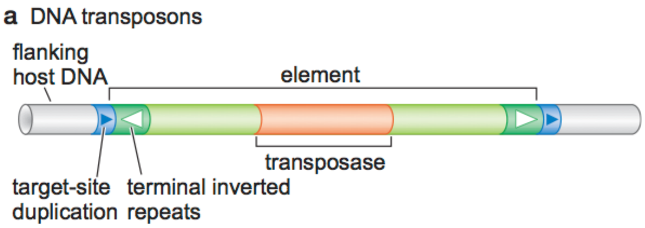
A enzyme coded for within the DNA transposon. Allows for **integration** into a host DNA, by cutting the DNA transposon element out of the donor DNA

2) Two-terminal inverted repeats

Flank the region encoding the transposase. Responsible for **directing** integration into the host DNA.

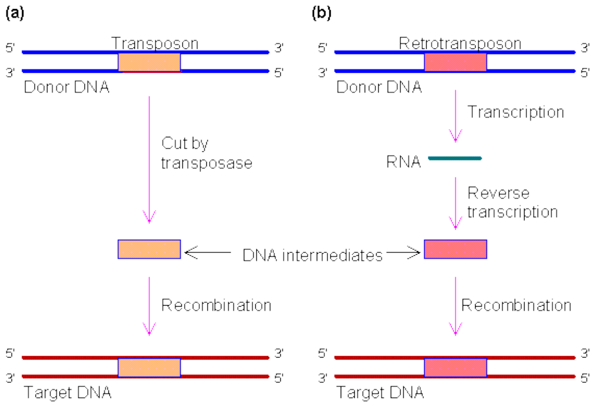
3) Target-site duplication

Flank the element. Important for the **specificity** of **insertion** into the host genome.



The important difference between retro and DNA transposons?

DNA transposons have **no intermediate**. That is, the element itself is **cut** out of the donor DNA by **transposase** and then transposed to a different region. Unlike retrotransposons, which require an **RNA intermediate** *prior to insertion* into the host.



5. The replication of LINE1 retrotransposons uses a target primed reverse transcription mechanism. Explain.

Recall the **structure** of LINE1

First you have the **promotor** located at the beginning of the 5'UTR. This is where RNAPIII will bind and transcribe.

The second important sequences is **ORF1**, and this encodes an RNA binding protein.

The third important sequence is **ORF2**, this encodes for a **reverse transcriptase**, a **endonuclease**, as well as an **integrase**.

Finally a 3'UTR and the poly-A sequence.

Transcription of the **LINE1** element begins at the **promotor** and **ends** at the end of the **poly-A sequence**. Translation of the transcript generates the following proteins:

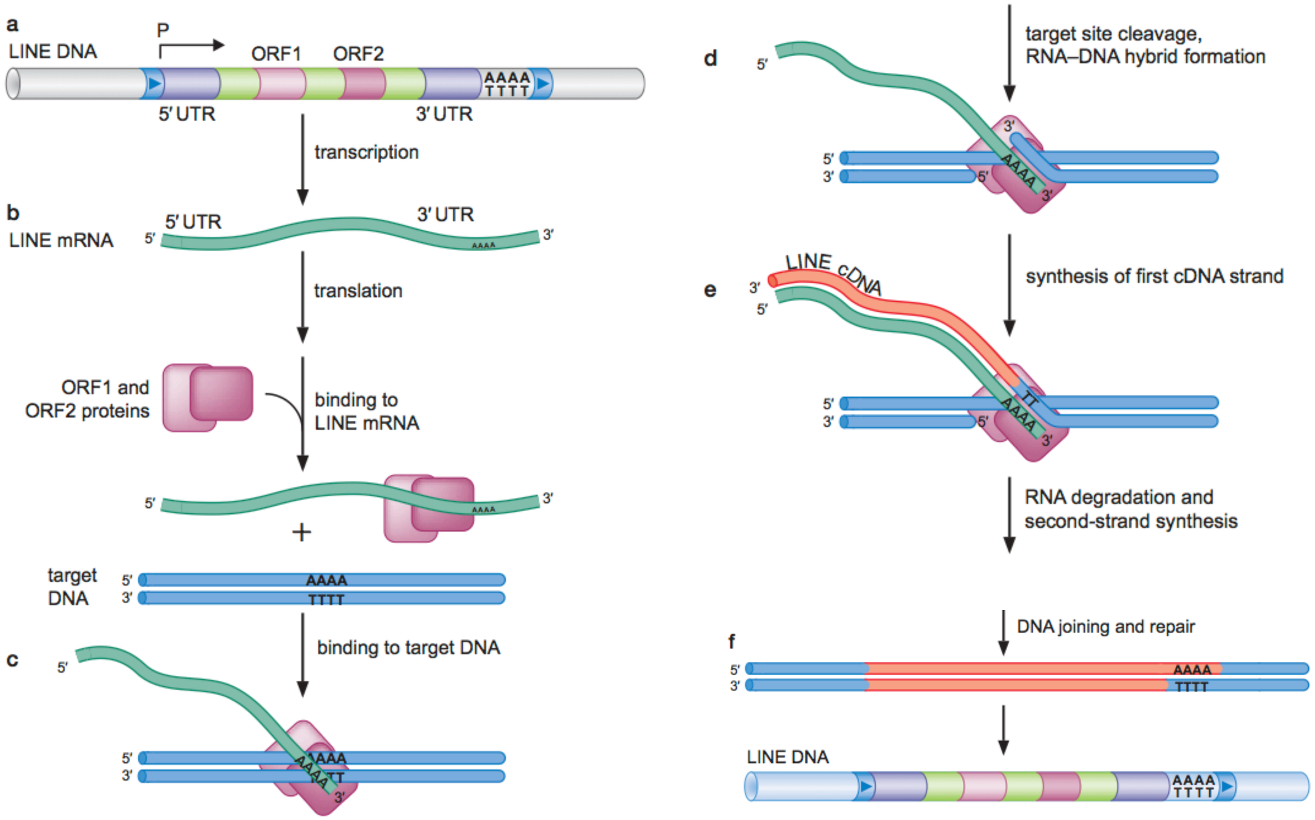
- RNA binding protein (from ORF1)
- RT, integrase, endonuclease from ORF2)

The **RNA binding** and **integrase** proteins form a dimer, which binds near the 3' end of the LINE mRNA. Recognition of the target DNA (the **site of insertion**) is also carried out by this protein dimer.

The targeted DNA contains a poly-T sequence, which will hybridize to the LINE mRNA's poly-A sequence. Upon binding, the DNA is cleaved by **endonuclease** activity and RT will transcribe the LINE mRNA into cDNA.

The RNA is then degraded, the cDNA complement is built, and integrase facilitates the inclusion of the LINE1 element into the host genome at the **site of insertion**.

Note that the LINE sequence inserted in the genome is DNA <u>NOT</u> RNA.



6. How is the full length genome of an RNA virus copied and integrated into the genome of a host cell.

The virus will enter the cell and lose its envelope, releasing its genome - in the form of **ssRNA** - into the host cell. The ssRNA in the cell will be copied into a ssDNA strand via RT activity.

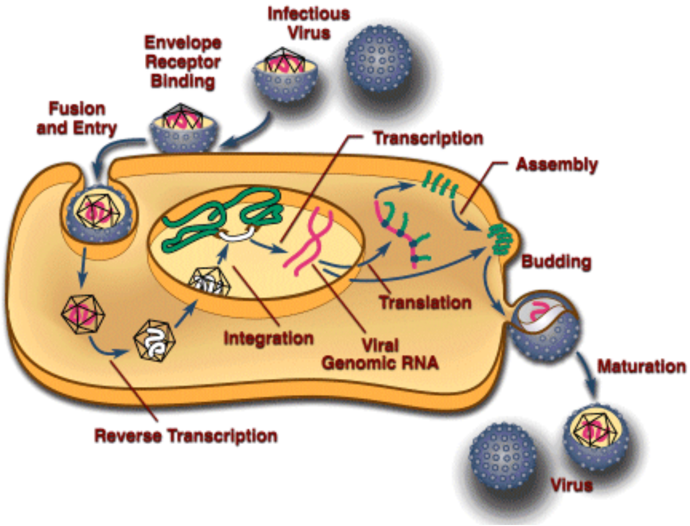
The hybrid RNA/DNA is then **converted** into a **dsDNA** hybrid. This is achieved by dual action of the cells RNase enzyme digesting the RNA of the RNA/DNA complex, and then DNAP synthesizing the complementary DNA strand based on the ssDNA template.

This dsDNA corresponding the the original viral RNA will be integrated into the hosts genome. Upon integration, sequences can then be transcribed and expressed as proteins. The necessary proteins for making viral particles are the **capsid proteins**, the **envelope proteins**, and the **reverse transcriptase enzyme** - necessary for conversion of the viral RNA genome into cDNA before integration into the host genome.

There are 3 (important) genes in the virus genome:

1) **gag**: encodes for RNA binding protein
2) **pol**: RT
3) **env**: encodes “outer-coat” proteins (cover virus and assist its release from the cell)

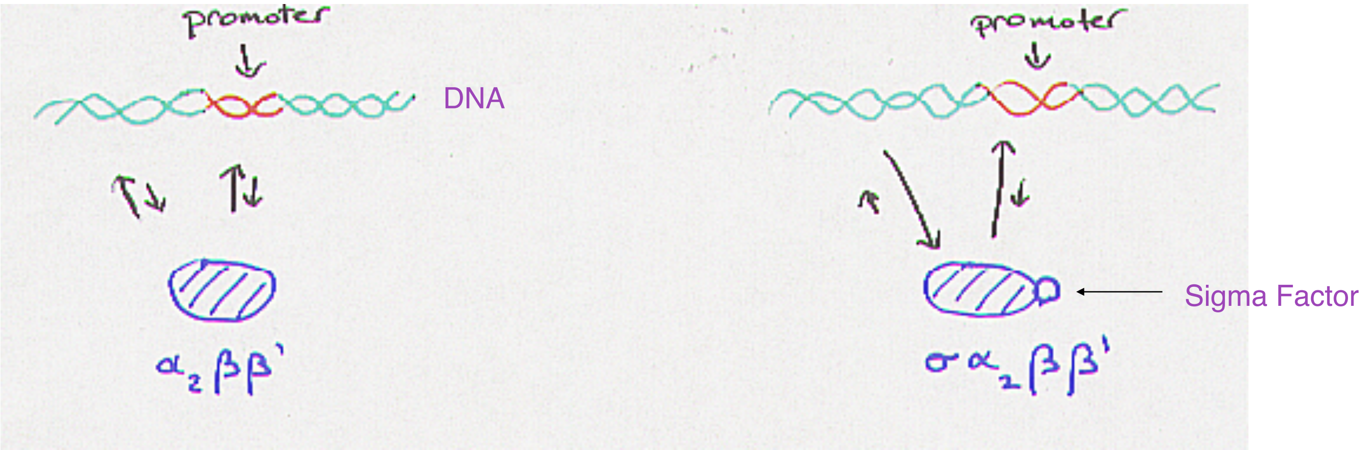
Once the viral DNA is integrated in the host genome, it is known as a **provirus**. This **provirus** must be expressed into a number of proteins which together, form the new virus particles. The transcription and translation reactions responsible for expressing the viral genes are carried out exclusively by the cell's machinery - which cannot distinguish between its own genetic information and that of the virus.



Lecture 7 - Transcription in Prokaryotes (Learning Objectives)

1. How does RNA polymerase find promoters in the genome?

See slides 18 for more info.



Recall: Promoters are 60bp on average, and the size of the E.Coli genome is 4×10^6 bp.

In order for RNAP to find and bind to the promoter, it will associate with a cofactor called **Sigma**. The Sigma factor, when associated with RNAP, is called the **holoenzyme**. The holoenzyme contains the Sigma factor as well as the two **alpha**, **beta**, and **beta prime subunits**. This is in contrast to the **core enzyme**, which is RNAP sans a sigma factor (thus it just contains the alpha, beta and beta prime subunits).

The holoenzyme will recognize promotor sequences due to the fact that it has a higher affinity for these sequences than the rest of the genome.

So the Sigma factor has two roles:

1) **Increase** RNAP's affinity for promotor sequences and
2) **decreases** RNAP's affinity for other regions in the genome.

Without the sigma factor, the core enzyme has approximately the same affinity for promotor and non-promotor regions of the DNA.

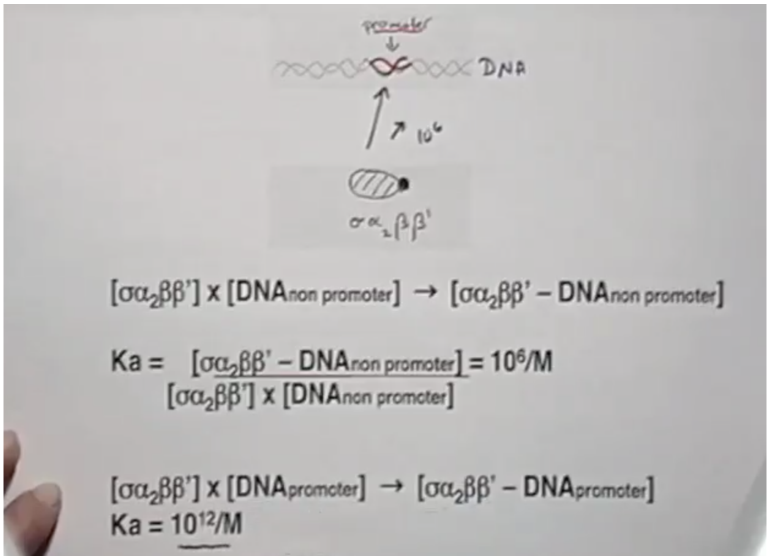
How do we measure RNA polymerase affinity?

Affinity of RNA pol for DNA can be measure by the **Ka** or association constant.

This is measure experimentally, by first starting the reaction between the **holoenzyme** and a DNA region **without** a promoter - then we measure how much of the holoenzyme/DNA complex is formed.

The concentration of the two subtracts subtracted each other divided by this number yields the **Ka** in this example, the value is 10^6

The process is then repeated using DNA promotor sequences, and we generate a knew Ka of 10^{12}, significantly higher.



$$[\sigma\alpha_2\beta\beta'] \text{ x } [\text{DNA}_{\text{non promoter}}] \rightarrow [\sigma\alpha_2\beta\beta' - \text{DNA}_{\text{non promoter}}]$$

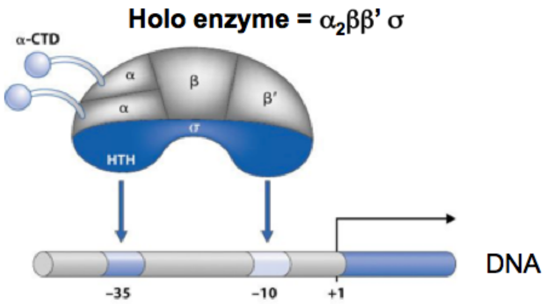
$$\text{Ka} = \frac{[\sigma\alpha_2\beta\beta' - \text{DNA}_{\text{non promoter}}]}{[\sigma\alpha_2\beta\beta'] \text{ x } [\text{DNA}_{\text{non promoter}}]} = 10^6/\text{M}$$

$$[\sigma\alpha_2\beta\beta'] \text{ x } [\text{DNA}_{\text{promoter}}] \rightarrow [\sigma\alpha_2\beta\beta' - \text{DNA}_{\text{promoter}}]$$

$$\text{Ka} = \underline{10^{12}}/\text{M}$$

2. What are the essential elements for a typical strong bacterial promotor?

See slides 15, 16, and 17 for more info.



(A) Sigma Factor binds to the promotor at **two contact points**:

Note that the transcription start site (TSS) is used as a references and located at the +1 position.

1) Sequence **TTGACA** (this can change *depending on the sigma factor bound!)* located precisely **35** nts **upstream** from the TSS (Transcription Start Site) on DNA (-35 site)
2) Sequence **TATAAT** (this can change *depending on the sigma factor bound!)* (the tata box!) located **10** nts **upstream** of the TSS on DNA (-10 site)

Also illustrated in the figure:

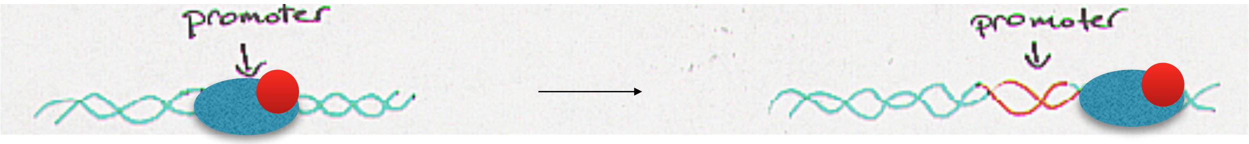
HTH, the *helix-turn-helix*, which is a motif of the **sigma** subunit that will bind to the nucleic acid.

CTD: C-Terminal Domain of the enzyme - important in eukaryotes as phosphorylation is necessary for initiation of transcription to take place (*not as important in prokaryotes!)*

Strong bacterial promoter is recognized by RNA pol holoenzyme (core enzyme plus sigma factor).

3. After the binding of the prokaryote RNA polymerase holoenzyme to a strong promotor, transcription of the gene located downstream from the promoter will take place. Describe the two steps that are necessary for the RNA polymerase to take off from the promotor and to transcribe.

See slides 18 & 19 for more info.



The figure above represents the holoenzyme (RNAP in blue, and sigma factor in red) moving off the promotor.

Two major steps are involved when RNA polymerase leaves the promotor and beings transcription:

1) Release of the sigma factor from the holoenzyme

The **sigma factor** is released from the holoenzyme. The core enzyme will remain attached to the DNA. The **affinity** of core enzyme for non-promotor sequences is now **higher** than for the promotor sequences. Therefore, RNAP (the core enzyme) is now **more attracted** to non-promotor DNA sequences then promotor DNA sequences.

2) Energy for the movement of the RNAP downstream from the promotor (transcription activity)

The **energy** necessary to move RNAP down the DNA helix is derived from the activated **NTP's** that are included in the RNA synthesis reaction (transcription).

Recall that during the addition of the NTP's into the growing RNA chain, pyrophosphate is cleaved - releasing energy. More energy is released during the breakdown of pyrophosphate into two **inorganic phosphates**. The net negative energy of the reaction fuels the movement of RNAP down the helix.

Overall reaction:

1) RNA-OH + NTP —> RNA-NMP + PPi
2) PP —> 2 Pi

4. Why does the reaction of RNA synthesis go to the right; towards the synthesis of the RNA chain (elongation)?

See slides #21 & #22 for more info.

We start with an RNA chain, that includes a free hydroxyl group at the 3' end. To this end, we will add a nucleotide, one at a time, and elongate the chain.

RNA-OH + NTP —> RNA-NMP + PPi

The **incoming** nucleotide is in the **triphosphate** form, but the **added** nucleotide is in the **monophosphate form**.

1) H_2O + NTP —> NMP + PPi

2) RNA-OH + NMP —> RNA-NMP + H_2O

by using a nucleotide **triphosphate**, we are able to form **pyrophosphate** upon elongation, which is then hydrolyzed by the enzyme in **pyrophophatase**, into **2** inorganic **phosphates**. This release of energy allows for the movement of the RNAP along the DNA matrix.

PPi —> 2 Pi

Take home: Movement of RNAP during elongation of a RNA chain (transcription) occurs for the same reasons that DNA synthesis does: The release of energy occurs at two points: 1) The cleavage of **pyrophosphate** during addition of NMP's and 2) The hydrolyzing of pyrophosphate into 2 inorganic phosphates. This release of energy makes the reactions of RNA synthesis favourable, and so elongation continues.

5. Why is it advantageous to use free nucleotides in the form of NTPs rather than NDPs or NMPs or unphosphorylated nucleotides in the reaction of RNA synthesis as catalyzed by the enzyme RNA polymerase?

See slides #21 & #22 for more info.

Overall reaction of the addition of one nucleotide to RNA chain

RNA-OH + NTP —> RNA-NMP + pyrophosphate (PPi)

The overall reaction is **energetically favourable** because of the **net negative** energy production, in other words energy is released (quite a bit too):

First Release of Energy: the hydrolysis of the phopshodiester bond between phosphate alpha and the sugar of the NTP, which generates energy.

Second Release of Energy: the hydrolysis of pyrophosphate into two inorganic phosphates.

Note that: the phophdiester bond which forms between the 3' free hydroxyl group of the RNA and the incoming NTP does NOT require energy, it is a simply nucleophilic attack.

Addition of incoming NTP's (said to be "activated") to the growing chain occurs at the 3' end, requiring a free hydroxyl group, with the NTP based on the template strands base.

See above for first and second releases of energy

If we have **NTP**

Recall: RNA-OH + NTP —> RNA-NMP + pyrophosphate
PPi —> (hydrolysis by **pyrophosphatase**) 2 Pi

This is energetically favourable for the reasons listed above. It is driven to the right (product formation). We have both the **first** and **second** releases of energy.

If we have **NDP**:

RNA-OH + NDP —> RNA-NMP + inorganic phosphate (Pi)

In this case, we do not have the **second** release of energy in the form of pyrophosphate being cleaved to form 2 inorganic phosphates. Thus, the reaction is not driven to products nearly as much as if NTP was used. Products can be formed but at a much slower rate.

If we have **NMP**:

RNA-OH + NMP —> RNA-NMP

In this case, we do not have the **first** or **second** release. Thus, the reaction is not driven to products nearly as much as if NTP or NMP was used. Products can be formed but at a *very very* slow rate.

If we have **unphosphorylated NT**:

RNA-OH + Nt's —> No reaction

Simply put, we do not have a phosphate which is required to form a phophodiester bond with the 3' free hydroxyl and elongate the strand.

important:
- use the term **Le Chatelier's principle** on the midterm, she is looking for this as a mark!
- Also note that because she asked virtually the same question on the last midterm, she probably wont ask this one again.

6. Consider the following DNA sequence corresponding to the coding strand (or **sense** strand). What would be the RNA sequence transcribed from this DNA sequence: 5' - TTAGCAGTCTAGTAACAT - 3'

See slides #2 for more info.

Two methods to solve this question:

1) Because we are giving the **sense strand**, the RNA sequence will be identical, **except for** the thymine residues will be replaced by **uracil** in the transcribed RNA, thus

RNA Sequence Transcribed: 5’ - UUAGCAGUCUAGUAACAU - 3’

2) The second and more complicated method involves coming up with the complementary DNA strand (**antisense**) and **transcribing this**. *Remember to get your 5’-3’ directions right.*

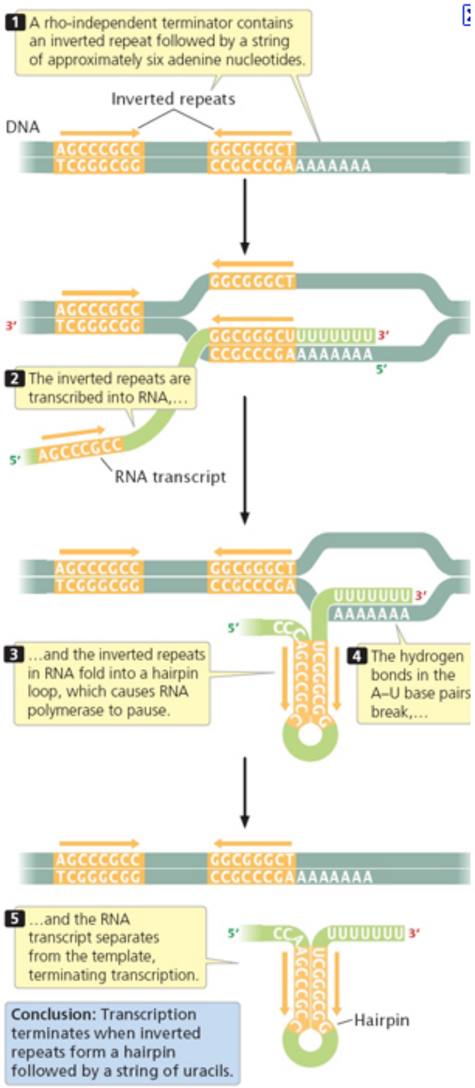
Both methods result in the same answer, but method two has an extra unnecessary step.

7. Name and give the principal characteristics of the 2 modes of termination of transcription for prokaryotes.

Rho-**independent** termination

Takes place because there is, in the first part, a sequence that is transcribed in the RNA and able to fold into a **stem-loop** structure. That is a section of the RNA that forms a dsRNA chain with a ssRNA loop at then end.

This stem loop structure is followed by a series of U residues at the 3’ end of the RNA. This series of U residues will hybridize with the DNA matrix. When RNAP encounters this structure, it will pause. This allows for the dislocation of the RNA from the DNA matrix strand - thus terminating synthesis.



Rho-**dependent** termination

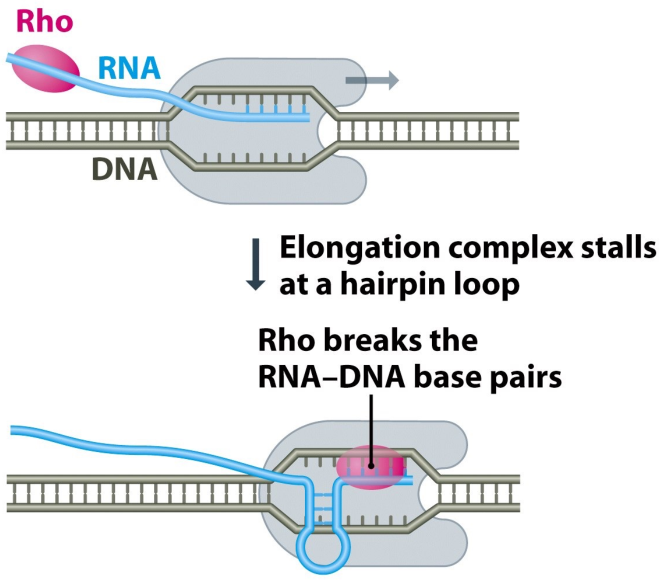


Figure 12-7 Genomes 3 (© Garland Science 2007)

Whenever there is the binding of **Rho** at the 5' end of the RNA, *termination will occur*.

For this to be completed, **Rho protein** (which is a hexamer of 6 identical subunits, shaped like a donut). The 5' end of the RNA enters in the centre of the Rho protein, and because of its helicase activity, will migrate toward the transcription bubble.

Rho will then unzip the DNA/RNA hybrid, and the RNA will be released from the transcription bubble - at this point, transcription has been terminated.

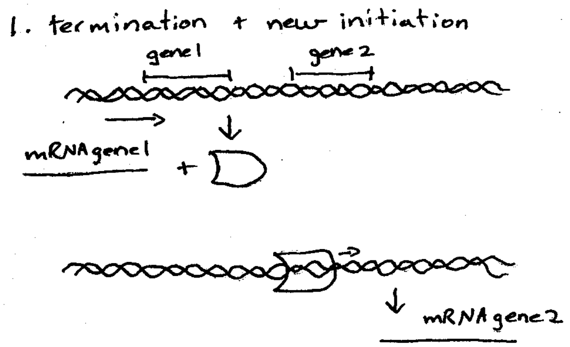
So the Rho-dependent termination requires an **input of energy** because the Rho protein must migrate from the 5' end of the RNA toward the transcription bubble - at a speed faster than RNAP so that it can catch up to the transcription bubble and undo the DNA Matrix/RNA hybrid.

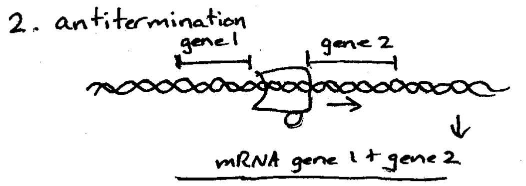
8. Explain anti-termination used by prokaryotes to end the transcription of a gene.

See slides #26 for more info.

This process is **unique to prokaryotes**, it does not occur in **eukaryotes**.

Scenario #1: **Termination + new initiation**





DNA of a prokaryote with two genes, gene1 and gene2. RNAP will bind to the promotor of gene, transcribe, and at the end find a termination sequence, dissociate, and RNA from gene1 will be released.

In order to transcribe gene2, RNAP must reassociate to DNA, that is to the promotor of gene2, in order to produce its RNA transcript.

So in this particular scenario, we have termination of transcription in gene1, and a new initiation of transcription prior to transcription of gene2. **Two**

distinct RNAs are formed.

Scenario #2: **Antitermination**

In the scenario where we have anti-termination (which is **unique** to prokaryotes) the structure of DNA is identical. However, the RNAP, after reaching the termination site at the end of gene1, will remain associated to DNA - it does not dissociate as is the case with scenario 1. It will remain associated to the DNA, and transcribe through the promotor of gene2.

In order for RNAP to remain associated to the DNA, it must contain an accessory protein - an example of which is **protein N** - which we find in phage lamda. The binding of protein N to RNA polymerase will allow the RNAP to remain associated to DNA when it reaches the termination sequence of gene1. In contrast to the above, the product of the transcription reaction is a **long, SINGLE RNA** which contains the sequence for both **gene1** and **gene2** - known as *polycistronic* or *bicistronic* in this case!

Lecture 8 - Gene expression in Prokaryotes (Learning Objectives)

1. Identify 3 factors controlling the initiation step of transcription in prokaryotes?

See slides #3 and 4 for more info.

First: the initiation rate of transcription is largely controlled by the **rate of formation** of the **complex RNA polymerase** with dsDNA (specifically, the promotor region).

1. Two consensus sequences of nucleotides in the **strong bacterial** promotor region
    A. sequence at -35: **TTGACA**
    B. sequence at -10: **TATAAT**

2. **Structural Changes** modifying the affinity of the RNA polymerase for the promoter:

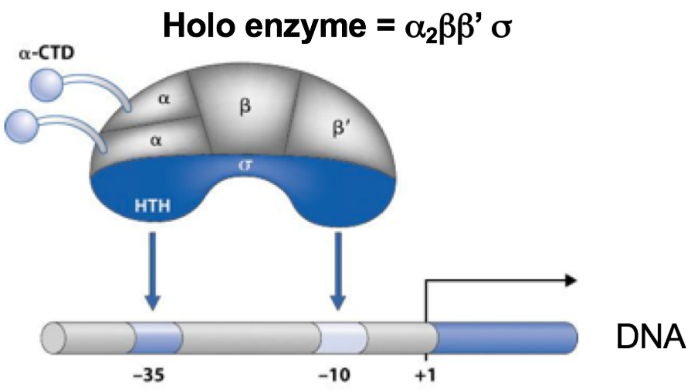
These structural changes take place within the enzyme RNA polymerase, and they occur in the **presence** of accessory proteins (either activators or repressors)

3. **Proteins** can change the affinity of the RNA polymerase for the promoter. If the proteins increase the affinity, they are **activators**, and if they decrease the affinity , they are **repressors**.

The presence of accessory proteins will influence the binding of RNA polymerase to DNA. Illustrated in the diagram is the **holoenzyme**, composed of two alpha subunits, and the ß and ß' subunits *as well as* the sigma factor, σ (our accessory protein in this case).

In this case, the σ subunit increases the affinity of RNAP for the bacterial promotor by binding directly to the sequences at positions -10 and -35 (from the TSS).

Figure:



Holo enzyme = $\alpha_2\beta\beta'\ \sigma$

2. What are the differences between positive and negative controls of gene expression

See slide 5 for more info.

Negative Controls

Genes under negative controls are expressed unless a **repressor** protein is present and **blocks** their transcription.

In this depiction, the yellow bands represent the -**10** and -**35** sites (upstream from the TSS) and the binding site for the repressor protein - known as the **operator** - in green. This makes up the **promotor**. The presence of a **repressor protein** bound to operator in the promotor physically blocks RNAP from binding to the promotor and transcribing the downstream gene under its control.

The binding of the **repressor** is regulated by a **ligand** (in red):

> The **ligand** can either **increase** the affinity of the **repressor** for the operator or it can **decrease** it.

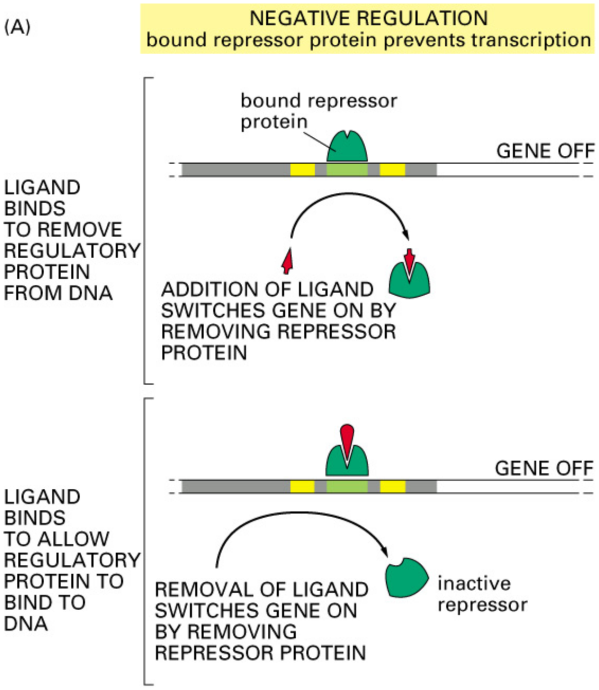


Figure 7–36 part 1 of 2. Molecular Biology of the Cell, 4th Edition.

Positive Controls

Genes under positive control will be expressed at **maximal** level when a **activator** protein (CAP-cAMP protein complex for the lac operon) is present and at a **basal**. In this case, the binding site for the activator protein is actually **upstream** from the -35 site of the promotor (as opposed to being between the -10 and -35 sites as is the case with negative controls) - known as the **CAP binding site**.

The binding of an **activator** protein to its binding site in the promotor will **increase** the initiation rate of transcription: that is, the **likelihood** of RNAP and the promotor forming a complex.

Again, the binding of the **activator** is also regulated by a **ligand** (in red):

> The **ligand** can either **increase** the affinity of the **activator** for its binding site on the polymerase, or it can **decrease** it.

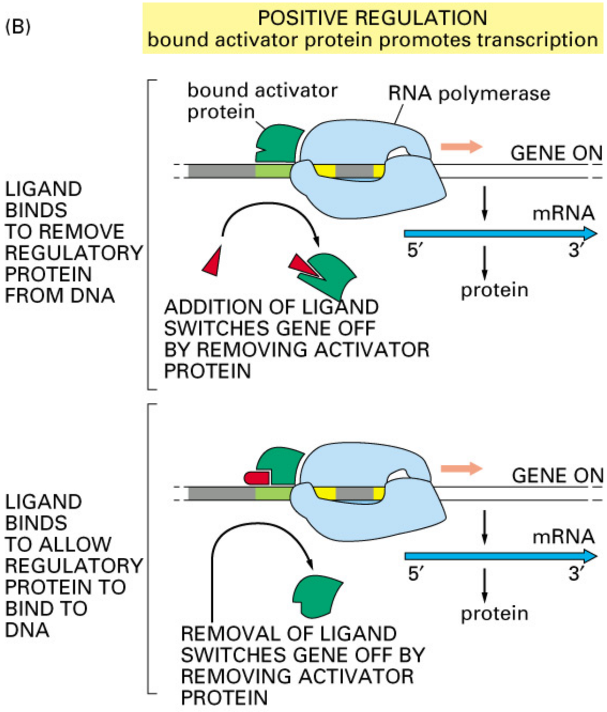


Figure 7–36 part 2 of 2. Molecular Biology of the Cell, 4th Edition.

3. Define the term operon and illustrate key components

NEED TO ADD IN HERE WHAT EACH STRUCTURAL GENE ENCODES PLUS THAT PROTEINS FUNCTION

See slides 8 & 10 for more info.

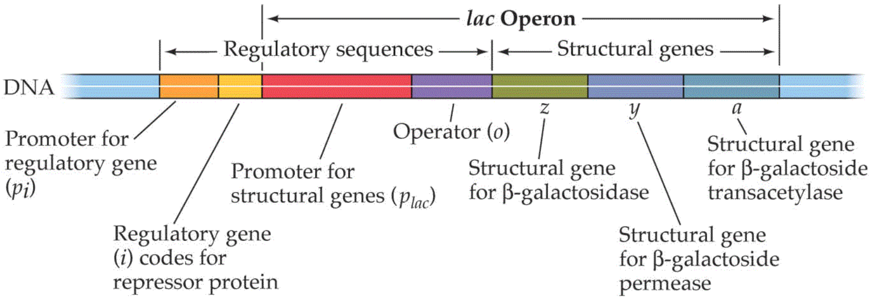
Defintion

An **operon** is a group of genes located in close proximity and which are regulated by the same regulatory region (contains promotor, operator, regulatory genes).

Key Components (using **Lac Operon** as example)

The lac operon contains **three** structural genes, **LacZ**, **LacY**, and **LacA**. Immediately upstream from those genes is a regulatory region which includes a **promotor** - to which RNA polymerase will bind - and an **operator**.

The *occupation* of the operator will influence the binding of RNA polymerase to the promotor and thus the level of gene expression or transcription.



4. Explain the mechanism used by the repressor (*lac repressor)* of the lactose operon to regulate gene expression. How is the binding of the repressor to the lactose operon regulated?

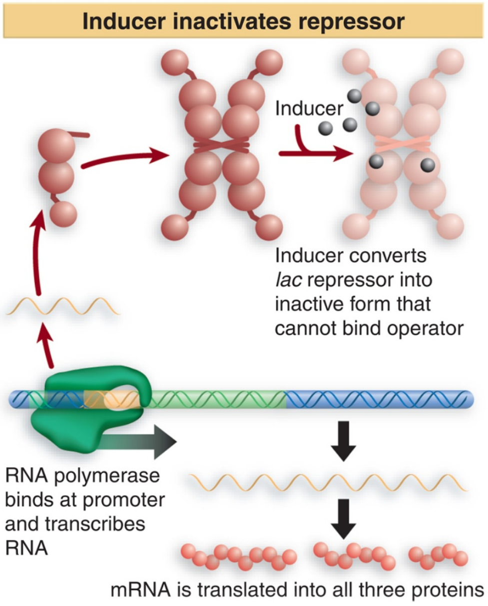
See slides 12 & 13 for more info.

Upstream from the regulatory region of the Lac operon is a gene encoding for a **repressor protein**, known as the **LacI** gene.

The protein will form a tetramer, this tetramer is the active form of the **repressor** protein - known as the *lac repressor*. It will bind to the **operator** thus preventing access to **RNAP,** preventing transcription.

The **repressor proteins** affinity for the operator region of the Lac operon can be reduced by **allolactose**, the **inducer**. When allolactose is bound to the repressor, the repressor will not bind to the operator and RNAP will be able to bind and transcription proceed.

So the binding activity of the repressor to the operator is controlled by the inducer, allolactose, which reduces its affinity for the operator.



Copyright © 2006 Pearson Prentice Hall, Inc.

5. Explain: when the repressor of the lactose operon is inactivated, transcription is **still** not optimum. What is the mechanism involved to reach maximal expression?

See slides 14 & 15 for more info.

The mechanism is that of **positive control**, and this is mediated by an accessory protein called **CAP**: catabolic activator protein.

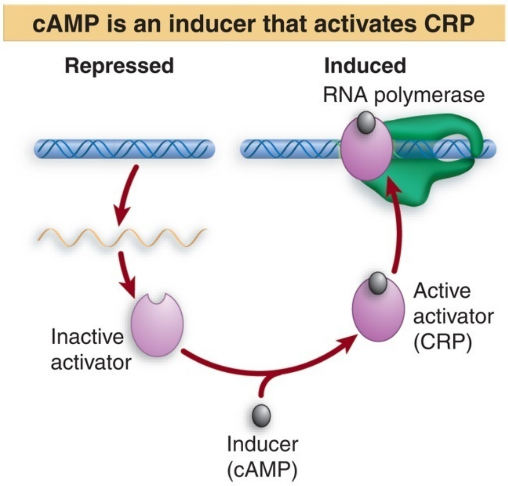
The binding of **CAP** to its binding site on the promotor of the lac operon is mediated by a ligand, **cAMP** (cyclic AMP). When the complex CAP-cAMP s formed, it will have a high affinity for the CAP **binding site**, and will increase the affinity of RNAP for the promotor - thus increasing transcription of the structural genes of the LAC operon.

What regulates the levels of cAMP?

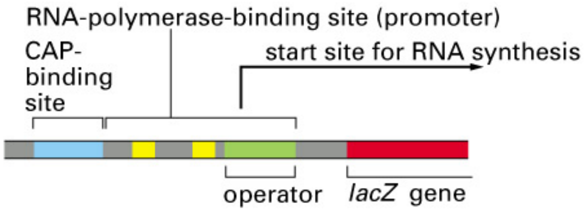
When glucose levels are **high**, the cAMP levels are **low**:
Low levels of cAMP, will not allow for the formation of CAP-cAMP complexes, CAP will not bind to the CAP binding site of the promotor, and the affinity of RNAP for the promotor will not be increased. Transcription will occur at its **basal rate**.

When glucose levels are **low**, the cAMP levels are **high**:

High levels of cAMP, will lead to the formation of CAP-cAMP complexes, CAP will bind to the CAP binding site of the promotor, and the affinity of RNAP for the promotor will be increased. Transcription will occur at its **optimum rate**.



Here we can see the structure of the Lac operon: The structural genes (starting with lacZ) are downstream from the regulatory regions, which includes the **operator**, to which the **repressor** (lacI) **binds**, the **promotor** to which RNAP binds, and the **CAP binding site**, to with the activator (CAP) binds.

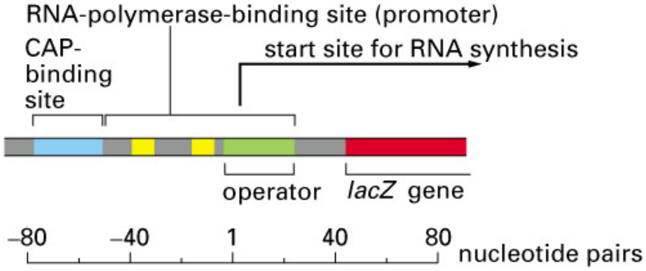


6. The activity of the lactose operon is also controlled by glucose concentration in the cell. Describe the state of transcription when:

a) both glucose and lactose are present
b) both glucose and lactose are absent
c) glucose is present and lactose is absent
d) lactose is present and glucose is absent

See slides 16 for more info.

First Recall the structure of the lac operon



1) both **glucose** and **lactose** are **present**

The presence of lactose will allow for the formation of **allolactose**, a ligand for the repressor. The complex **allolactose**-**repressor** has <u>no affinity</u> for the operator site, leaving it free. Access to the promotor by RNAP is <u>not inhibited</u>.

When **glucose** is present is **high** concentration, cAMP is present in **low** concentration. the **CAP-cAMP** complex will <u>not form</u>, and thus the activator protein CAP has no affinity for the CAP binding site.

Therefore transcription occurs at its **basal rate**.

2) **glucose** is **present** and **lactose** is **absent**

If **lactose** is absent, there is no formation of **allolactose**. Thus, the **repressor protein** (lacI) will have a high affinity for the **operator site**, preventing binding of RNAP to the promotor.

When **glucose** is present is **high** concentration, cAMP is present in **low** concentration. the **CAP-cAMP** complex will <u>not form</u>, and thus the activator protein CAP has no affinity for the CAP binding site.

Therefore transcription does **not occur** (we say the **operon is “off”**)

3) both **glucose** and **lactose** are **absent**

Even though glucose is **absent**, and the activator CAP-cAMP complex’s will form and bind to the CAP binding site increasing affinity for RNAP to bind to the promotor, RNAP **cannot bind to** the promotor because of the presence of the **repressor protein** bound to the **operator**.

Therefore transcription does **not occur** (we say the **operon is “off”**)

4) **glucose** is **absent** and **lactose** is **present**

The presence of lactose will allow for the formation of **allolactose**, a ligand for the repressor. The complex **allolactose**-**repressor** has no affinity for the operator site, leaving it free. Access to the promotor by RNAP is not inhibited.

When glucose is **absent**, the activator CAP-cAMP complex’s will form and bind to the CAP binding site increasing affinity for RNAP to bind to the promotor.

Therefore transcription does **occurs** at its **optimal rate** (we say the **operon is “on”**)

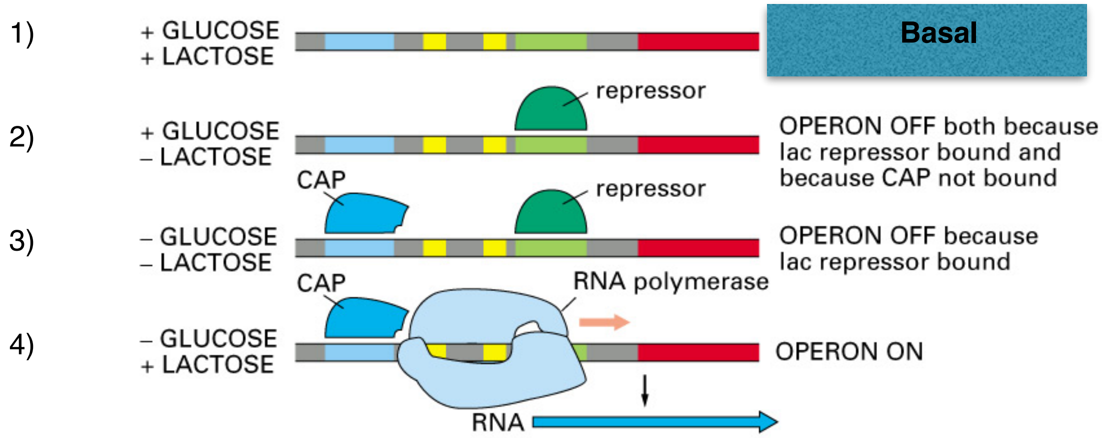


Figure 7–38. Molecular Biology of the Cell, 4th Edition.

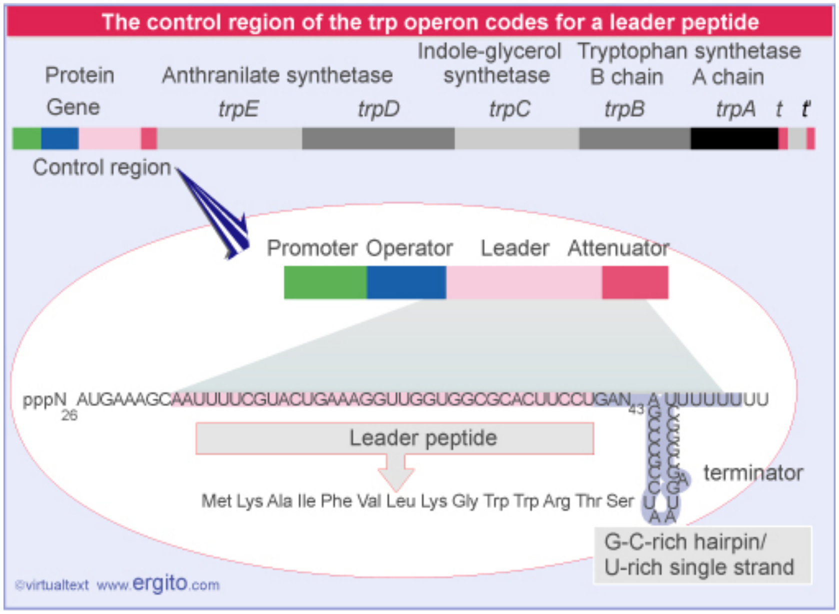
7. Name and illustrate the important elements of the regulatory region of the tryptophan (**trp**) operon

See slides 18 for more info

It is a bit more complex than the lactose operon. The **trp operon** includes **5 structural genes**: *trpE*, *trpD*, *trpC*, *trpB*, *trpA* which encode **3 enzymes** required to synthesize **tryptophan** from the substrate **chorismate**.

The regulatory region for the trp operon includes a **promotor** - to which RNAP will bind, and **operator** - to which the repressor will bind and in addition: a **leader** and **attenuator** sequence.

The transcription of the trp operon will start **immediately after** the **operator** sequence, and so as a result, both the **leader** and **attenuator** sequences will be included in the RNA.



8. How is the binding activity of the repressor of the trp operon regulated?

See slides 20 for more info

+1 is the TSS, in yellow are the promotor binding sites for RNAP (-10 and -35), and in between those sites is the **operator** sequence - this is where the **repressor** binds.

trp, is known as a **co-repressor**. It forms a **complex** with the **repressor protein**, and this complex has a high affinity for the operator sequence - preventing RNAP from binding to the promotor.

In absence of trp, the **repressor protein** has a low affinity for the operator sequence, and so it does not bind (we say it is **inactive**). This allows RNAP to bind to the promotor and transcribe the gene.

In this way, the bacteria only transcribes the ***5 genes*** *needed to produce the* ***3 enzymes*** *with synthesize trp from* ***chorismate*** *when trp levels in the cell are low.*

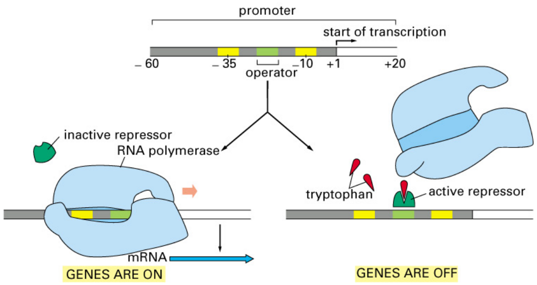


Figure 7–34. Molecular Biology of the Cell, 4th Edition.

9. Explain: allolactose acts as an inducer of the lactose operon while tryptophan acts as a co-repressor of the trp operon.

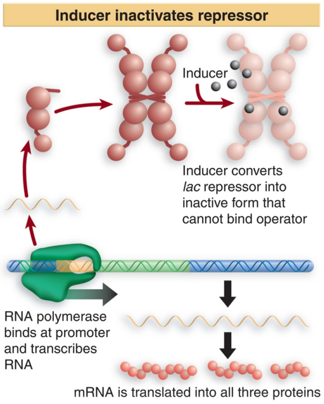
See slides 12 and 20 for more info.

Allolactose: inducer

The binding activity of the repressor is under the control of the ligand **allolactose**, and allolactose will **reduce the affinity** of the repressor for the **operator** sequence.

If the repressor is not bound onto the operator site, RNAP has access to the promotor and **can transcribe** the structural genes of the operon.

For this reason, allolactose is called an **inducer**, because in its presence the expression of the lactose operon takes place.



Copyright © 2006 Pearson Prentice Hall, Inc.

trp: co-repressor

trp is said to be a **co-repressor**, because when present trp will bind to the **repressor**, and this **repressor-trp** complex has a **high affinity** for the operator, blocking access by RNAP.

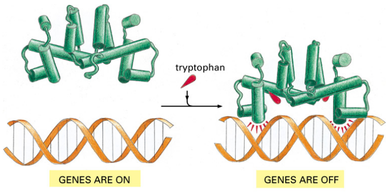


Figure 7–35. Molecular Biology of the Cell, 4th Edition.

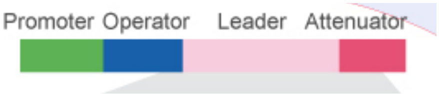
For this reason, trp is said to be an **repressor** (co-repressor to be exact) because in its present the trp operon is **not transcribed**.

She continuously mentions that allolactose and trp are effectively opposites.

10. Explain the mechanism of attenuation of the trp operon

See slides 21 and 24 for more info.

Recall: in addition to the **promotor** and **operator** regions of the trp operon, there are the **leader** and **attenuator** sequences. Remember also that transcription starts immediately after the **operator**, so both the leader & attenuator sequences are included in the transcribed RNA. This process is **unique to prokaryotes**.

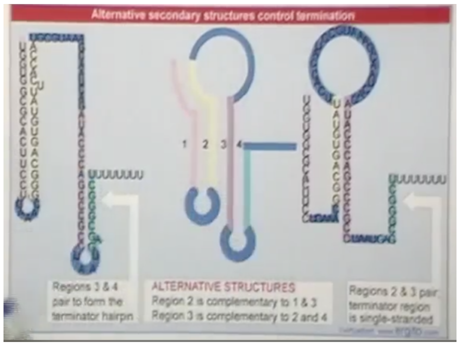


What is particular about this transcribed RNA, is that it can **two possible** sets of **secondary** structures (hairpin-loop).

The RNA transcribed from the **leader** sequence can be divided into 4 regions, 1, 2, 3 and 4.

The **first** secondary structure that can be formed is a result of structure **1** annealing to structure **2**, and structure **3** annealing to structure **4**. So the formation of the 3-4 structures, will serve as a terminator signal for RNAP, because the region **4** is followed by a series of **U**-**Residues**. (This is Rho independent termination!)

The **second** secondary structure that can be formed is the 2-3 (region 2 hybridizes to region 3) hairpin loop structure - forming a **single** hairpin loop. Region 4 remains single stranded, and this structure does **not** serve as a termination signal for RNAP.



Alternative secondary structures control termination

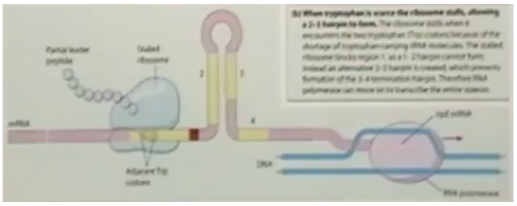
Regions 3 & 4 pair to form the terminator hairpin

ALTERNATIVE STRUCTURES
Region 2 is complementary to 1 & 3
Region 3 is complementary to 2 and 4

Regions 2 & 3 pair terminator region is single-stranded

In this example, we have a region 2-3 hybridization - which **does not** form a termination signal for RNAP (hairpin loop is not followed by U residues, so Rho-independent termination does not occur)

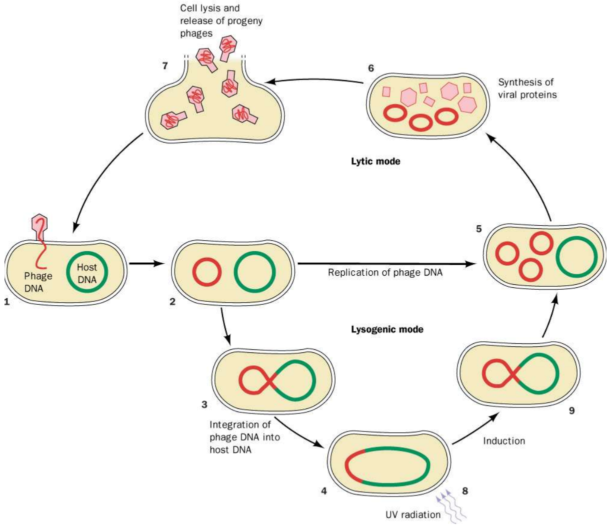
The reason why the 2-3 hybrid forms and not the 1-2 and 3-4 structures in this case is because a **ribosome** is covering (and paused over) **region 1**. Its paused is caused by **two** consecutive **trp codons**. Because levels of trp are low in the cell, the ribosome must pause and wait for tRNA to bring in trp residues to add to the peptide chain. *Thus, if trp levels climb, there is no pause by the ribosome over region 1, 1-2, 3-4 hairpin loops form and translation is terminated.*



Lecture 9 - Regulation of Gene Expression in Bacteriophages
(Learning Objectives)

Note that Phage Lambda is a bacterial virus, or Bacteriophage, which infect *E. coli*

1. What are the difference between the lytic and lysogenic modes in terms of bacteriophage - bacteria interaction?



See slides 2 for more info.

Phage lamda (illustrated in pink) will infect a bacteria by injection of the viral genome into the bacterium. The remaining structure of the phage particle will remain attached **to** <u>but</u>

outside of *the bacterium*. Phage DNA will then become circular (Similar to the bacterium DNA).

At this point there are two choices, the bacterium enters
1) The **lytic** mode OR
2) **Lysogenic** mode

Basically, the **lytic mode** or **lytic cycle** will allow the phage to replicate its genome and generate multiple new phage particles - and at the end of the cycle will cause lysis of the bacterial cell, thus **killing** it.

In the **lysogenic** mode, the phage DNA will be integrated into the bacterial genome, and the bacterium will divide and replicate both its own genome and the integrated viral genome. In the lysogenic mode there are **no expression of phage genes**.

To summarize:
- In lytic mode, phage DNA is expressed, viral particles are produced and the host bacterium is lysed.
- In lysogneic mode, phage DNA is not expressed but rather replicated alongside the bacterial DNA, and the host bacterium remains alive!.

2. Name two repressors and two anti-terminators important for the expression of phage λ genes and briefly explain the mechanisms of repression and anti-termination.

See slides 20 of Reg. Of Gene Exp in PRo, 26 of Transc. in prokaryotes, and 6 of lec 9 for more info

Recall: **Repression** is the binding of a repressor protein to a DNA sequence adjacent to the promotor (the **operator**), this blocks access of RNA polymerase to the promotor and prevents transcription of genes downstream under its control.

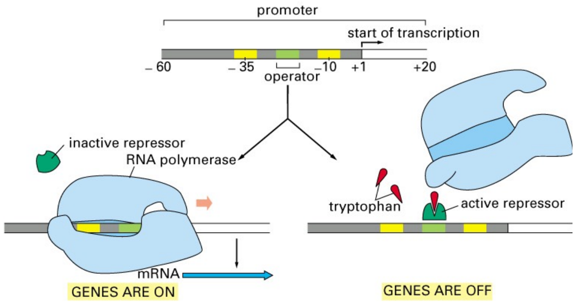
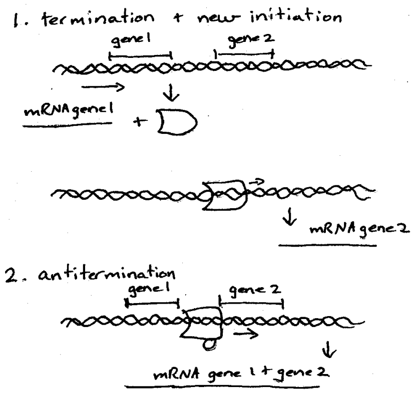


Figure 7–34. Molecular Biology of the Cell, 4th Edition.

Fig.12



Recall: **Anti-termination**: mechanism by which the RNAP does not end transcription of a gene when it runs into a termination sequence

- but rather *reads through* the promotor of another adjacent gene also transcribing it. This is achieved by binding of an accessory protein to the RNAP.

This allows transcription to continue onto adjacent genes and produces a **polycistronic** mRNA that results in additional proteins being expressed.

Two important **repressors** *in the expression of phage λ genes are:*

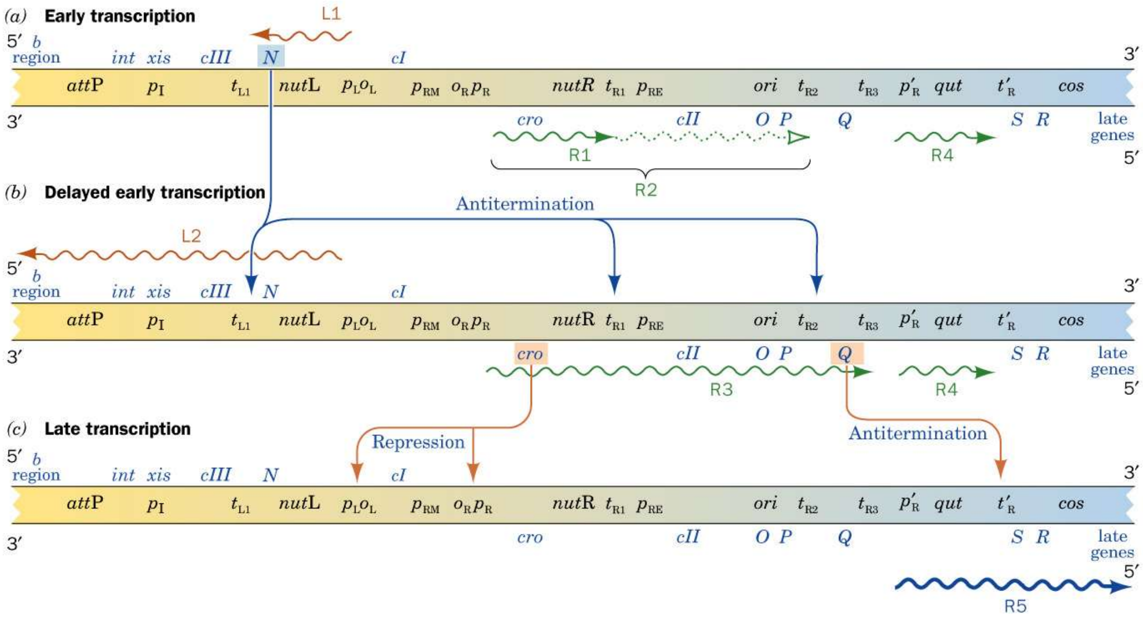
1) **Cro**: represses the expression of c1 gene
2) **c1**: represses the expression of **early viral genes** - known as *"the" repressor protein/ gene*

Two important **anti-terminators** *in the expression of phage λ genes are*
Note these are the accessory proteins which bind to RNAP!

1) **N**: generates transcripts L2 from L1, R3 from R1 or R2
2) **Q**: generates transcripts R5 from R4

The following is a map of the genome of phage λ. You will recall that in-between the two strands are listed the **regulatory sites** (P for promotor, O for operator). Above and below the two DNA strands are indicated the names of the phage genes (Example **N**, and **Cro**). The arrows indicate the transcripts produced from the transcriptional reactions. The two strongest promotors for phage lamda are **P_R** and **P_L**. R stands for "right" which means transcription occurs to the right, and L stands for "left".

The two important repressors are **Cro** - which is regulated by the P_R promotor and the O_R operator. And the second important repressor is **c1**, regulated by P_R and P_L promoters.



The two important anti-terminators for phage λ gene expression are **N** - which is expressed very early on after infection of the bacterium. As protein N accumulates, it will allow for extension of transcript L1 into L2. Similarly, it will extend transcript R1 or R2 into R3. Translation of R3, includes the gene encoding protein **Q** - then allows for the extension of transcript R4 into R5.

3. For each phase of the lytic cycle, list the products of transcription, name the transcripts produced, orientation (left or right), promotor and terminator sites used and location on the chromosome.

See slides 5 & 6 for more info

Basically, the early event starts from the strongest promotors: P_R and P_L. Soon after infection or induction, *E.coli RNA pol* initiates transcription

Early Transcription

- **L1** transcript generated from **P_L**, it contains the anti-terminator protein **N**.
- **R1** generated from the **P_R** promotor. This transcript includes the sequence for the gene **cro**. In the presence of anti-terminator N, R1 will be extended to R2 (and then R3 in **delayed early transcription** phase) and include genes cII, O and P.
- **R4** generated from at **P'_R** and includes the genes *qut*.

The remaining stages to not see any transcription reactions producing *new transcripts*, but strictly extended those produced during the early transcription phase.

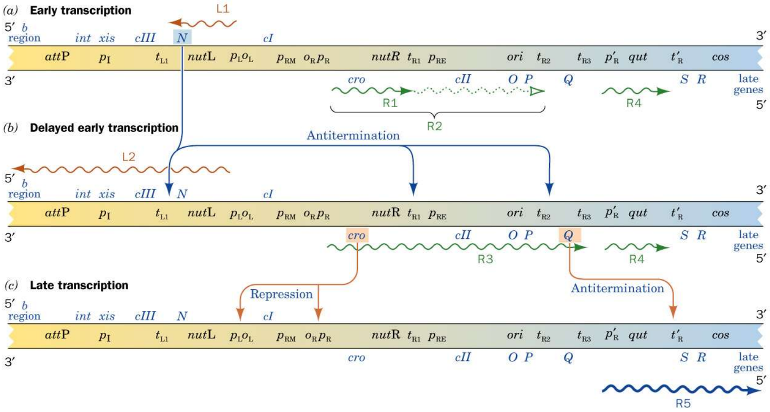
Delayed Early Transcription

- **L1** is extended to **L2** via action of RNAP and anti-termination protein **N**
- **R1** or **R2** are extended to **R3** via RNAP and anti-termination protein **N**: R3 contains cII, O, P and in addition, anti-terminator protein **Q**, which will elongate R4 in late transcription.
-

Late Transcription

- R4 is extended to R5 via action of RNAP and anti-termination protein Q. It includes "**late phage genes**"

Thus, the last two phases of the lytic cycle (delayed early transcription and late transcription) are characterized by anti-termination reactions, with initiation of transcription occurring in first phase (early transcription)

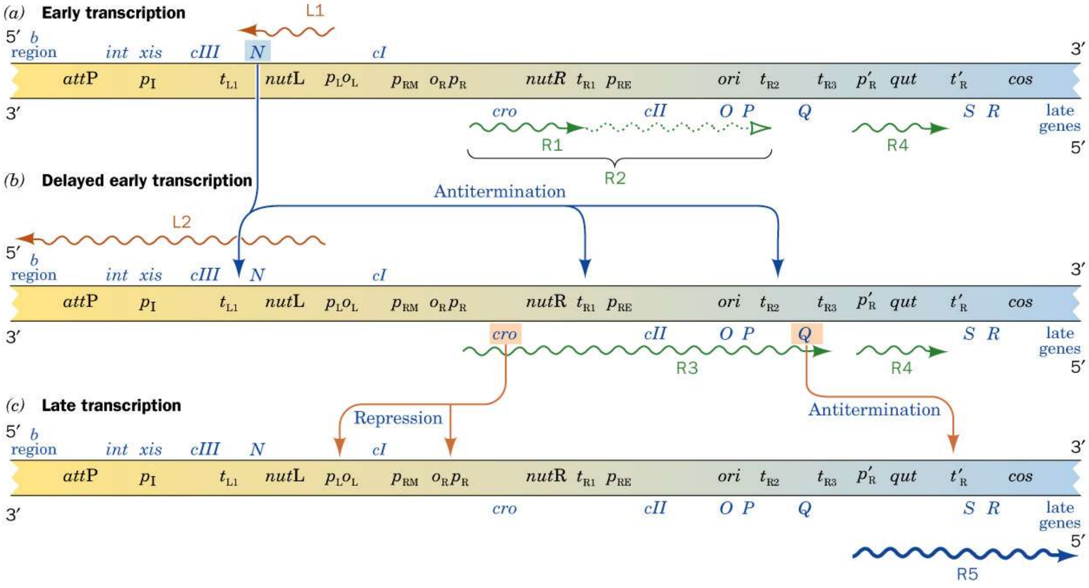


4. Explain the meaning of: the lytic cycle is regulated by a cascade of gene expression

See slides 6 for more info

Cascade of expression: gene expression in each of the latter 2 phases (B - delayed early transcription and C - late transcription) is regulated by proteins (**anti**-**terminators**) synthesized in the preceding phase (early transcription).

Transcription occurring during phases B&C of the lytic cycle occur only because **protein N** (an anti-terminator) has been produced in sufficient amounts in phase A - and thus can induce anti-termination. This timing in expression of genes is essential for the **regulation** of the **lytic mode**.



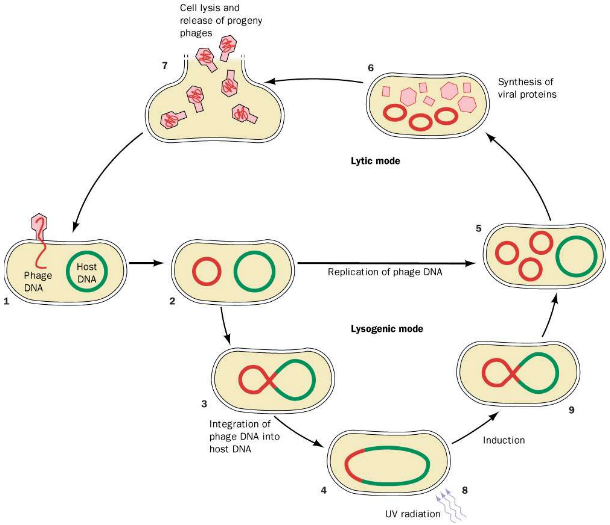
Similarly, the expression of protein **Q** (anti-terminator) must occur in the delayed early transcription phase in order for transcription of R4 to be elongated into R5 in late transcription.

So there is a cascade of events which must occur in a certain order in order for expression of phage λ genes to be timed correctly.

5. Name the protein that triggers lysogeny. Detail the transcripts produced during lysogeny and the role of proteins encoded by those transcripts. What is the role of c1 repressor?

See slides 12 for more info

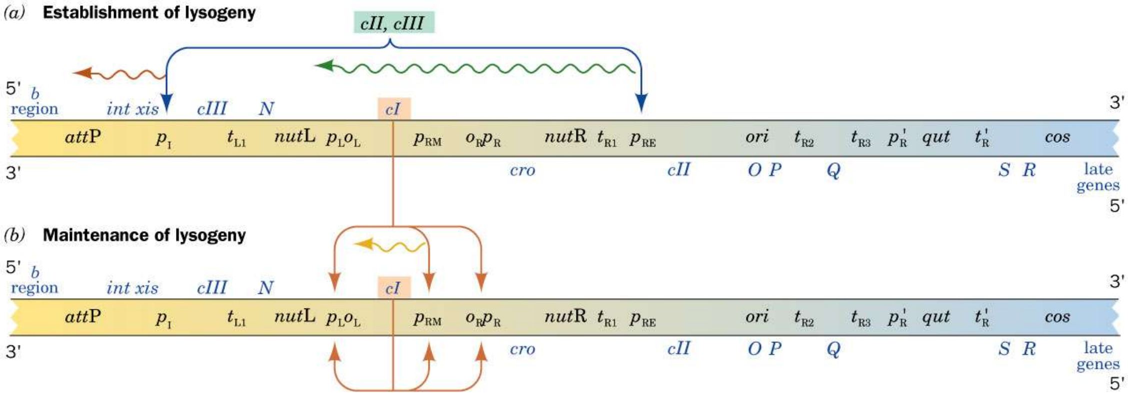
Recall: **Lysogeny** is the integration of the phage genome into the bacterial host genome and replicated alongside it - to be transmitted through progenies.



Lysogeny is triggered by high concentration of **cII** protein. cII stimulates transcription leftward of **2 transcripts**:

1) from P_{RE} to generate a transcript containing the gene for cI, a **repressor**.
2) from P_1 to produce the **integrase** gene - allowing for phage DNA to integrate into the host bacterium DNA.

So where does cII come from?



It is encoded by the gene that is under control of a **weak promotor**, P_{RE}. Because it is under control of a weak promotor, it will **not** accumulate fast initially - *remember that the **proteins** under control of **strong** promotor (Like N from P_L) are the first to be produced and accumulate after infection.* Once cII protein does begin to accumulate in the bacterium, it will induce transcription (so its an activator) to the left of the **2 transcripts** described above - and lysogeny is established.

Once lysogeny is established - because cII protein has accumulated - the **cI repressor** protein will begin to **accumulate** (translated from the transcript under control of P_{RE}). cI repressor will "turn off" expression of all genes in phage lamda by binding to the O_L and O_R operator sites preventing transcription of all **early gene products** under the control of the two strong promotors P_L and P_R. But it will also **induce** its own expression from P_{RM}, in order to **maintain** lysogeny.

6. Cro and cI repressors both shut down the synthesis of **early gene products**, how can this same activity by regulated by two different proteins?

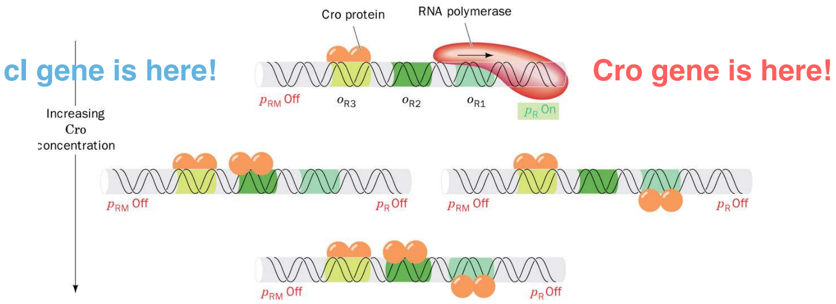
See slides 14 and 16 for more info

Cro and cI bind to (one of!) the same regulatory sequence - the **O_R** sites - influencing the binding of RNA pol to **P_R** and therefore the transcriptional activity of all **early viral genes** under its control.

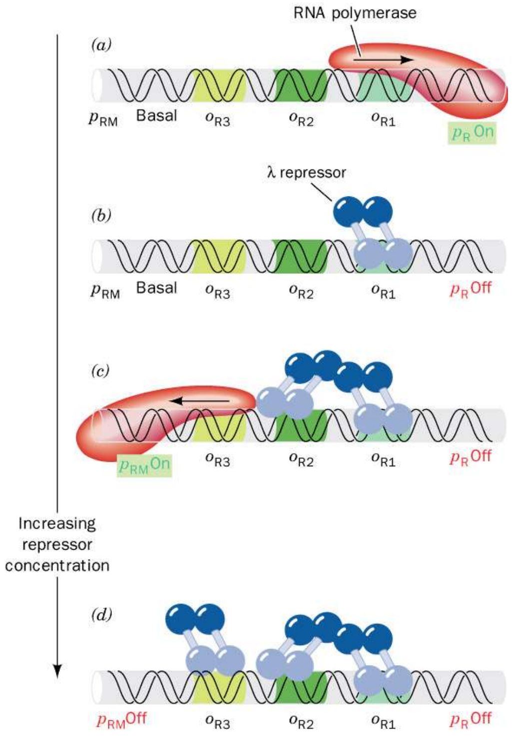
O_R is divided into three sites: O_{R1}, O_{R2} and O_{R3}. Near the O_{R1} site, is the P_R promoter. Right next to this is **cro gene**. Near the O_{R3} gene is the P_{RM} promoter, and right next to this is the **cI** gene ("the repressor gene").

Cro Protein

Cro and **cI** will fight for those same regulatory sites - but with *essentially* **opposite affinities.** Cro has the highest affinity for the **O_{R3}** site, and will bind here **first**. Once bound, RNAP will be **prevented** from transcribing genes under control of the P_{RM} promotor, and thus, **cI** is repressed. **O_{R1}** remains free, and so cro is expressed from **P_R**. As concentration of cro accumulates, it will begin to occupy O_{R1} and O_{R2} sites - with identical affinity, thus **repressing** its own expression by turning of the **P_R** promotor. *Thus, accumulation of cro protein gradually represses it own expression*.



cI Protein



cI protein also binds to the O_{R1}, O_{R2} and O_{R3} regulatory sites, but will *essentially opposite affinity* than that of cro protein.

Firstly, it will bind with highest affinity for the O_{R1} site, and then through **cooperative binding**, to the O_{R2} site. Once both the **O_{R1}** and **O_{R2}** sites are occupied, the P_{RM} promotor is switched "on" (notice this is different than cro, which does not induce its own expression) and cI protein is effectively inducing its own expression.

Further accumulation of cI protein leads to the occupation of the **O_{R3}** site, which because of its proximity to P_{RM}, switches off the promotor and thus represses cI protein.

Notice the major differences here: cro protein strictly represses its own expression(and cI's expression) as its concentration increases. It first occupies the **O_{R3}** site with the highest affinity, then the **O_{R1}** and **O_{R2}** sites with equal affinity.

cI protein cooperatively binds to the **O_{R1}** and then **O_{R2}** sites, repressing P_R and expression of cro and inducing its own expression from P_{RM}. Accumulation leads to occupation of the **O_{R3}** site, and repression of its own expression via P_{RM}.

This needs a big touch up

7. How can cI repressor stimulate transcription of its own gene while repressing all other mRNA synthesis. Does Cro regulate its own synthesis?

See slides 14 and 16 for more info

Overview:

Absence of repressor: RNA pol initiates Transcription at high level from p_R and at basal level at p_{RM}.

Repressor [CONCENTRATION] increases: affinity of cI for o_R:

$o_{R1} > o_{R2} > o_{R3}$ binding to o_{R1} first and transcription from p_R is off.

Repressor increases higher: binding to o_{R1} and to o_{R2} (cooperative binding). Stimulates transcription from p_{RM}. Correct balance of repressor for the maintenance of lysogeny.

Very high concentration of repressor: all o_R sites are occupied and transcription from both p_R and p_{RM} is off.

Thus, this is how cI promotes is own expression. It promotes it when O_{R1} and O_{R2} are occupied by cI repressor protein (at which point P_R is turned **off**, all other transcription is repressed and binding of RNAP to P_{RM} is facilitated, so cI is produced) - but also represses it by binding to O_{R3} and turning P_{RM} off when cI concentration becomes very high.

Does the same apply to cro?

We know that cro binds to the same O_R sites ($O_{R1,2 \text{ and } 3}$) but the affinity for each is *essentially* **opposite**. It is highest for O_{R3}, with no difference between O_{R2} and O_{R1}. So at low concentration of cro, O_{R3} sites are filled first. Once concentration builds, **cro** may bind to either O_{R2} or O_{R1} - but cooperative binding is **not observed** (contrast to the above scenario). At which point, cro has effectively suppressed its own repression as P_R is "off". Once the concentration builds even higher, all three sites are bound to (although this doesn't change expression - P_R is off either way).

Important point here: notice that cI protein can both increase its own expression **and** repress it (all explained above). In the case of cro protein, increasing concentration only serves to **repress** its own expression.

8. How does the binding of increasing concentration of Cro to the O_R affect transcription?

See slides 16 for more info

Basically a mechanism to regulate high concentration of **cro**. Binding occurs with highest affinity to O_{R3} sites, turning off expression of cI gene from P_{RM}, then binding of O_{R2} and O_{R1} sites will occur as [cro] increases. All three O_R sites are eventually occupied, turning off expression of genes under the control of P_{RM} and P_R simultaneously (this includes **cro** itself).

So a **high** concentration of **cro** will **repress** its own **expression**.

Note this is basically the same answer/explanation as the second part to the question above.

9. The lytic switch is induced by cleavage of the cI repressor protein to allow transcription of which gene? How is the transcription of cI then blocked? - *THIS IS BASICALLY ASKING HOW WE GO FROM LYSOGENY TO LYTIC CYCLE.*

See slides 16&17 for more info

We will now see how an infected bacterium in the lysogenic cycle can switch to the lytic mode. Note: a bacterium **cannot** go from lytic to lysogenic modes - because the host bacteria will lys and die in the lytic cycle. If the bacterium is in the lysogenic mode however, it can be induced to switch to the lytic mode.

So how does this occur?

Recall: cI is “the repressor protein” as it repressor **early viral genes** and prevents the host bacterium from entering lytic mode.

Cleavage of cI protein will free the O_{R3} site and allow for RNA pol to bind DNA and transcribe cro gene. Cro protein accumulates and acts as a repressor for the P_L and P_R promotors preventing expression of cI gene.

So we are in the lysogenic mode, and cI expression is induced, and that occurs because the O_{R1} and O_{R2} sites are occupied by cI protein - which is a positive signal to transcribe from P_{RM} (cI protein is under its control). R and and M stand for “repressor maintenance”. So in order to turn off P_{RM}, one must destroy cI - this occurs through UV radiation <u>OR</u> a signal cascade within the bacterium leading to the cleavage of the protein. Once cI protein is destroyed, then the transcription of cI gene is no longer up-regulated, and it will stop accumulating.

If it stops accumulating, the P_R is freed. Cro, which is under its control, will then start to accumulate. It will first bind to O_{R3}, turning off expression of cI by blocking P_{RM}. So now it is the cro repressor that is dictating expression. Because we don't wont cro to accumulate in excess, it will regulate its own expression when concentrations rise to high (see above).

So at this point, cro is being expressed and cI is repressed. The P_R strong promotor is now free (O_R not occupied by cI) as well as P_L which allows for the transcription of **early viral genes**, and the production of **anti-terminator** proteins which leads to the cascade of reactions of the lytic cycle.

Summary: cleavage of cI repressor allows for expression of cro, which will then allow for transcription of R1/R2 transcription, expression of **anti-terminator** proteins, extension of transcripts required to produced **viral proteins**, and eventually production of viral particles and lysis of the cell (the lytic cycle).