CHAPTER 13 – PROKARYOTE GENES: *E. COLI* LAC OPERON

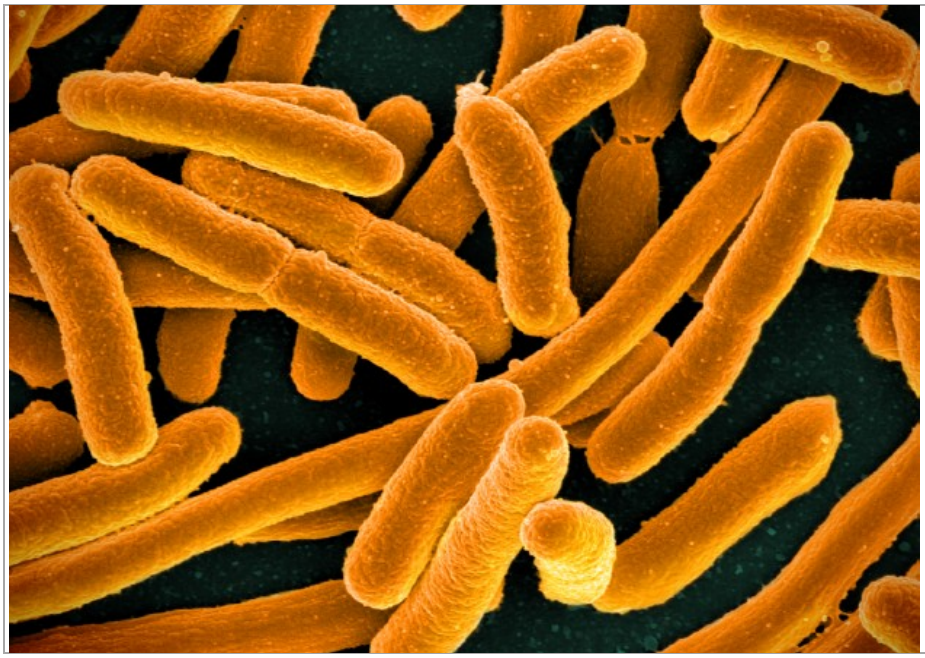

Figure 1.
Electron micrograph of growing *E. coli.* Some show the constriction at the location where daughter cells separate. The colouring is false.
(Flickr-NIAID-CC BY 2.0)

INTRODUCTION

With most organisms, every cell contains essentially the same genomic sequence. How then do cells develop and function differently from each other? The answer lies in the regulation of **gene expression**. Only a subset of all the genes is expressed (i.e. are functionally active) in any given cell participating in a particular biological process. Gene expression is regulated at many different steps along the process that converts DNA information into active proteins. In the first stage, transcript abundance can be controlled by regulating the rate of transcription initiation and processing, as well as the degradation of transcripts. In many cases, higher abundance of a gene's transcripts is correlated with its increased expression. We will focus on **transcriptional regulation** in *E. coli* (**Figure 1**). Be aware, however, that cells also regulate the overall activity of genes in other ways. For example, by controlling the rate of mRNA translation, processing, and degradation, as well as the post-translational modification of proteins and protein complexes.

1. THE *LAC* OPERON – A MODEL PROKARYOTE GENE

Early insights into mechanisms of transcriptional regulation came from studies of *E. coli* by researchers Francois Jacob & Jacques Monod (see Section 4 on page 4). In *E. coli,* and many other bacteria, genes encoding several different polypeptides may be located in a single transcription unit called an **operon**. The genes in an operon share the same transcriptional regulation, but are translated individually into separate polypeptides. Most prokaryote genes are not organized as operons, but are transcribed as single polypeptide units.

Eukaryotes do not group genes together as operons (an exception is *C. elegans* and a few other species).

1.1. BASIC LAC OPERON STRUCTURE

E. coli encounters many different sugars in its environment. These sugars, such as **lactose** and **glucose**, require different enzymes for their metabolism. Three of the enzymes for lactose metabolism are grouped in the ***lac* operon**: ***lacZ***,

lacY, and ***lacA*** (**Figure 2**). *LacZ* encodes an enzyme called **β-galactosidase**, which digests lactose into its two constituent sugars: glucose and galactose. *lacY* is a **permease** that helps to transfer lactose into the

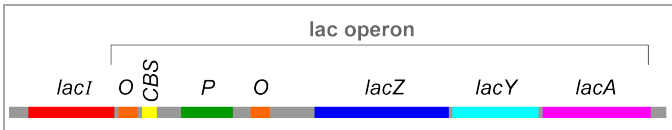

Figure 2.
Diagram of a segment of an E. coli chromosome containing the lac operon, as well as the lacI coding region. The various genes and cis-elements are not drawn to scale. (Original-Deyholos-CC BY-NC 3.0)

cell. Finally, *lacA* is a **trans-acetylase**; the relevance of which in lactose metabolism is not entirely clear. Transcription of the *lac* operon normally occurs only when lactose is available for it to digest. Presumably, this avoids wasting energy in the synthesis of enzymes for which no substrate is present. In the *lac* operon, there is a single mRNA transcript that includes coding sequences for all three enzymes and is called a polycistronic mRNA. A cistron in this context is equivalent to a gene.

1.2. *CIS*- AND *TRANS*- REGULATORS

In addition to these three protein-coding genes, the *lac* operon contains several short DNA sequences that do not encode proteins, but instead act as binding sites for proteins involved in transcriptional regulation of the operon. In the *lac* operon, these sequences are called **P (promoter)**, **O (operator)**, and **CBS (CAP-binding site).** Collectively, sequence elements such as these are called ***cis*-elements** because they must be located adjacently to the same piece of DNA in order to perform correctly. On the other hand, elements outside from the target DNA (such as the proteins that bind to these *cis*-elements) are called ***trans*-regulators** because (as diffusible molecules) they do not necessarily need to be encoded on the same piece of DNA as the genes they regulate.

2. NEGATIVE REGULATION – INDUCERS AND REPRESSORS

2.1. *lacI* ENCODES AN ALLOSTERICALLY REGULATED REPRESSOR

One of the major *trans*-regulators of the *lac* operon is encoded by ***lacI***, a gene located just upstream from the lac operon (**Figure 2**). Four identical molecules of *lacI* proteins assemble together to form a **homotetramer** called a **repressor** (**Figure 3**). This repressor is **trans-acting** and binds to two cis-acting operator sequences adjacent to the promoter of the *lac* operon. Binding of the repressor prevents RNA polymerase from binding to the promoter (**Figure 2**, **Figure 5.**). Therefore, the operon is not transcribed when the operator sequence is occupied by a repressor.

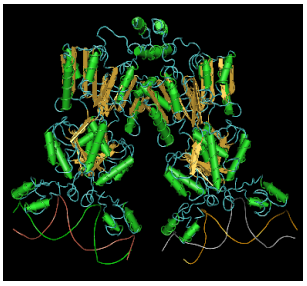

Figure 3.
Structure of *lacI* homotetramer bound to DNA.
(Original-Deyholos- CC BY-NC 3.0)

2.2. THE REPRESSOR ALSO BINDS LACTOSE (ALLOLACTOSE)

Besides its ability to bind to specific DNA sequences at the at the operator, another important property of the *lacI* protein *lacI* protein is its ability to bind to allolactose. If lactose is lactose is present, **β-galactosidase** (β-gal) enzymes convert a convert a few of the lactose molecules into allolactose (

Figure 4.).

This allolactose can then bind to the *lacI* protein. This alters the shape of the protein in a way that prevents it from binding to the operator. Proteins which change their shape and functional properties after binding to a ligand are said to be regulated through an **allosteric** mechanism.

Therefore, in the presence of lactose (which β-gal converts to allolatose), the repressor doesn't bind the operator sequence and thus RNA polymerase is

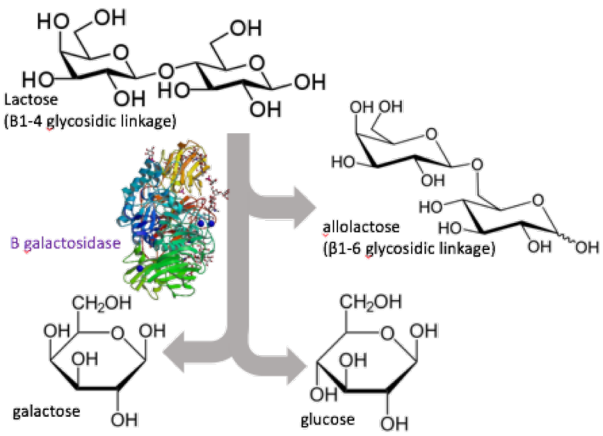

Figure 4.
β-galactosidase catalyzes two reactions: one converts lactose to allolactose; the other splits lactose into galactose and glucose. (Original-Nickle-CC BY-NC 4.0)

able to bind to the promoter and transcribe the *lac* operon. This leads to a moderate level of expression of the mRNA encoding the *lacZ*, *lacY*, and *lacA* genes. This kind of secondary molecule that binds to either activator or repressor and induces the production of specific enzyme is called an **inducer**. The role of *lacI* in regulating the *lac* operon is summarized in **Figure 5**.

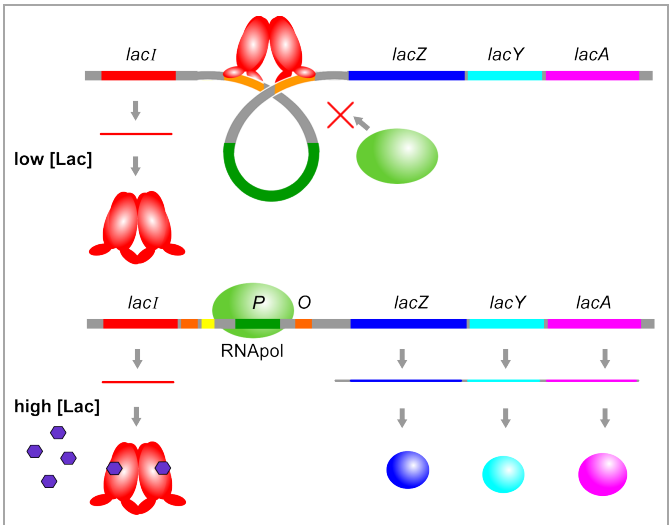

Figure 5.
When the concentration of lactose [Lac] is low, *lacI* tetramers bind to operator sequences (O), thereby blocking binding of RNApol (green) to the promoter (P). Alternatively, when [Lac] is high, lactose binds to *lacI*, preventing the repressor from binding to O, and allowing transcription by RNApol.
(Original-Deyholos-CC BY-NC 3.0)

3. POSITIVE REGULATION – *CAP*, cAMP & POLYMERASE

A second aspect of *lac* operon regulation is conferred by a *trans*-acting factor called **cAMP binding protein** (**CAP**, **Figure 6**). CAP is another example of an allosterically regulated *trans*-factor. Only when the CAP protein is bound to cAMP can another part of the protein bind to a specific *cis*-element within the *lac* promoter called the **CAP binding sequence (CBS)**. CBS is immediately in front of the promoter (P), and thus is a *cis-acting* element. When CAP is bound to at the CBS, RNA polymerase is better able to bind to the promoter and initiate transcription. Thus, the presence of cAMP ultimately leads to a further increase in *lac* operon transcription.

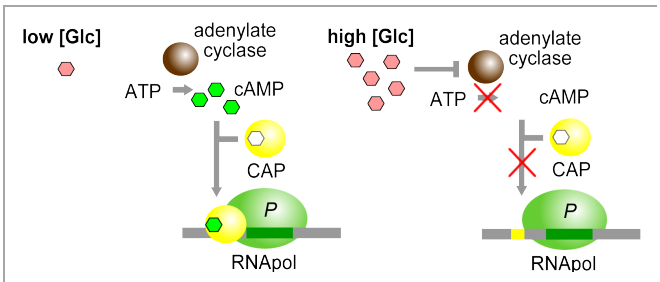

Figure 6.
CAP, when bound to cAMP, increases RNApol binding to the lac operon promoter. cAMP is produced only when glucose [Glc] is low. (Original-Deyholos-CC BY-NC 3.0)

The physiological significance of regulation by cAMP becomes more obvious in the context of the following information. The concentration of cAMP is inversely proportional to the abundance of glucose. When glucose concentrations are low, an enzyme called **adenylate cyclase** is able to produce cAMP from ATP. Evidently, *E. coli* prefers glucose over lactose, and so expresses the *lac* operon at high levels only when glucose is absent and lactose is present. This provides another layer of adaptive control of *lac* operon expression: only in the presence of lactose **and** in the absence of glucose is the operon expressed at its highest levels.

4. THE USE OF MUTANTS TO STUDY THE *LAC* OPERON

4.1. SINGLE MUTANTS OF THE *LAC* OPERON

The *lac* operon and its regulators were first characterized by studying mutants of *E. coli* that exhibited various abnormalities in lactose metabolism. Mutations can occur in any of the *lacZ*, *lacY*, and *lacA* genes. Such mutations result in altered protein sequences, and cause non-functional products. These are mutations in the protein coding sequences (non-regulatory).

Other mutants can cause the lac operon to be expressed constitutively, meaning the operon was transcribed whether or not lactose was present in the medium. Remember that normally the operon is only transcribed if lactose is present. Such mutants are called **constitutive** mutants. Constitutive mutants are always on and are unregulated by inducers. These include *lacO* and *lacI* genes.

4.2. OPERATOR MUTATIONS

The operator locus (*lacO*) - One example is ***O***$^-$, in which a mutation in an operator sequence reduces or precludes the repressor (the *lacI* gene product) from recognizing and binding to the operator sequence. Thus, in O^- mutants, *lacZ*, *lacY*, and *lacA* are expressed whether or not lactose is present. Note that this mutation is **cis dominant** (only affects the genes on the same chromosome) but not in trans (other DNA molecule). Another common name for a defective operator is O^c to refer to its constitutive expression. O^- and O^c are synonyms.

Note that constitutively expressed O^- mutants may not be maximally expressed, and the extent of the mutation can also affect the level of expression. (**Table 1**)

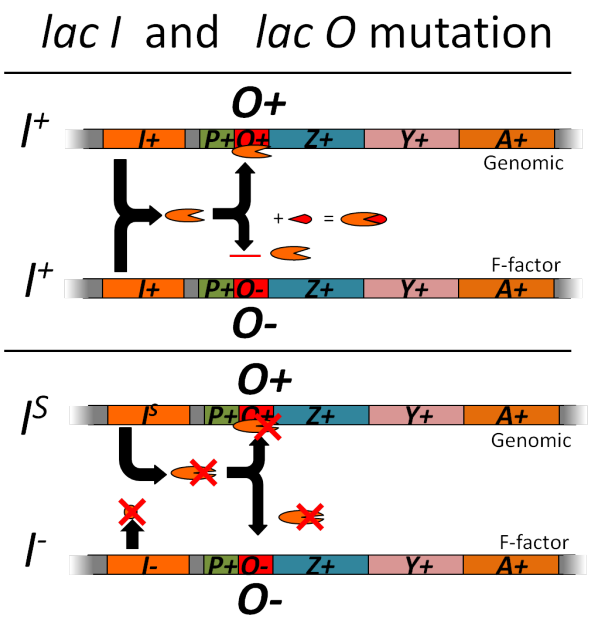

Figure 7.
Both *E. coli* strands with genotypes $I^+O^+Z^-Y^+A^+/F\ I^+\mathbf{O}^-Z^+Y^-A^+$ and $I^SO^+Z^+Y^+A^+/F\ I^-\mathbf{O}^-Z^+Y^+A^+$ will induce all the *lac* genes because repressor cannot bind to the O^- sequence on the F-factor and cannot prevent transcription.
(Original-Locke-CC BY-NC 3.0)

Table 1. Constitutively expressed O^c mutants may not be maximally expressed and have various levels of expression.

Level	Genotype	Explanation
100%	lac I^- O^c	no repressor
10-20%	lac I^+ O^c	repressor fails to bind tightly
~1%	P^+ O^c, high glucose	basal transcription, constitutive
0%	P^- or Z^-	no transcription

4.3. INDUCER MUTATIONS (*lacI* LOCUS)

The *lacI* locus has two types of mutations: ***I***$^-$ and ***I***^{S}.

One class of mutant allele for *lacI* (called ***I***$^-$) either (1) prevents the production of a repressor polypeptide, or (2) produces a polypeptide that cannot bind to the operator sequence. Note that these two alleles would have different genetic sequences, but the phenotype is the same. Theoretically, we should have better locus names to properly identify specific alleles, but for now we'll group them under one name. With this form of mutation, the repressor *cannot* bind and transcription *can* occur without the presence of inducer (allolactose). This can also be referred to as a "constitutive expresser" of the *lac* operon: the absence of repressor binding permits transcription.

Note that I^+ is dominant over I^-. For example, in *E. coli* strain with $I^+Z^-Y^+A^+/F\ I^-Z^+Y^-A^+$, the *lac* genes will not be inducible because the I^+ allele will still

produce functional repressors that bind to operator sequence, preventing transcription. (**Figure 8**)

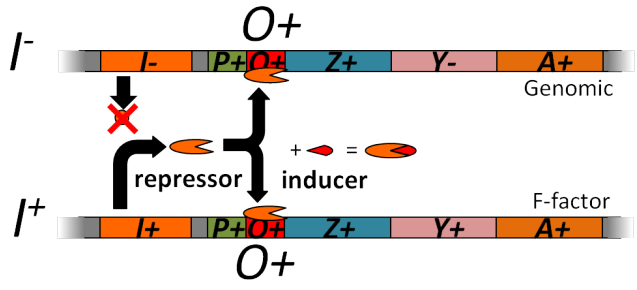

Figure 8.
E. coli strain with genotype $I^+Z^-Y^+A^+/F\ I^-Z^+Y^-A^+$ will not produce *lac Z, lac Y* and *lac A* products.
(Original-Locke-CC BY-NC 3.0)

The other class of mutant alleles for *lacI* are called I^s. The altered amino sequence of their proteins remove the "allosteric site". This means the repressor polypeptide cannot bind allolactose – and therefore it cannot change its shape. Even in the presence of lactose, it remains attached to the operator, so the *lac Z, lacY* and *lac A* genes cannot be expressed (no RNA polymerase, no protein). This mutant constitutively represses the *lac* operon whether lactose is present or not. The lac operon is not expressed at all and this mutant is called a "super-repressor". I^s is therefore dominant to both

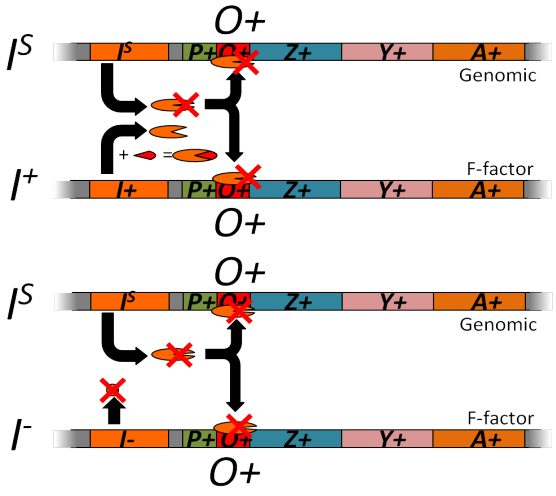

Figure 9.
E. coli strains with genotypes 1) $I^SZ^+Y^+A^+/F\ I^+Z^+Y^-A^+$ and 2) $I^SZ^+Y^+A^+/F\ I^-Z^+Y^+A^+$ will not produce *lac Z, lac Y* and *lac A* products. (Original-Locke-CC BY-NC 3.0)

I^+ and I^- in trans. Therefore, *E. coli* strains with the genotypes 1) $I^SZ^+Y^+A^+/F\ I^+Z^+Y^-A^+$ and 2) $I^SZ^+Y^+A^+/F\ I^-Z^+Y^+A^+$, the *lac Z*, *lac Y* and *lac A* genes will not be inducible (**Figure 9**).

The repressor protein encoded by *lacI* gene has at least two independent functional domains. This is the reason why it different mutations can give different classes of mutant phenotypes. **(Figure 10)**

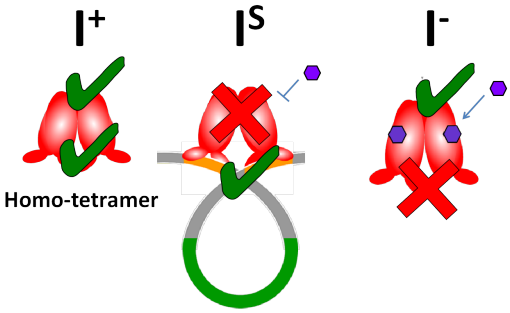

Figure 10.
Because the repressor encoded by lac I gene has independent domains, mutations can also occur independently.
(Original-Deyholos/Locke-CC BY-NC 3.0)

4.4. THE F-FACTOR AND TWO *lac* OPERONS IN A SINGLE CELL – PARTIAL DIPLOID IN *E. COLI*

More can be learned about the regulation of the lac operon when two different copies (each containing mutations) are present in one cell. This can be accomplished by using an **F-factor** (also known as a plasmid) to carry one copy, while the other is on the genomic *E. coli* chromosome. This results in a partial diploid in *E. coli* that contains two independent copies (alleles) of the lac operon and *lacI*.

An F-factor (named so because it creates "fertility" in the cell which contains it) is an **episome**. This is a self-replicating extrachromosomal piece of DNA: it is outside of the large, circular bacterial chromosome but has its own origin of replication. Jacob and Monod explored the regulation of the *lac* operon by introducing different genotypes into bacterial cells and noting how they responded in the presence of absence of sugars, namely glucose and lactose using the mutant classes discussed above.

For example, the genotype of a host bacterium that has a *lacI*$^-$ gene that is supplied with F factor containing *lacI*$^+$ can be written as *lacI*$^-$/*F*$^+$*lacI*$^+$. In fact, it doesn't matter which piece of DNA (genomic or episome) has which operon, so we can leave off the "F" altogether – equivalent notation is *lacI*$^-$/*lacI*$^+$.

Thus, cells with two copies of a gene sequence are partial diploids, called **merozygotes** (**Figure 11**). These allow researchers to test the regulation with different combinations of mutations in one cell. For example, the F-factor copy may have a I^S mutation while the genomic copy might have an O^C mutation. How would this cell respond to the presence/absence of lactose (or glucose)? This partial diploid can be used to determine that I^S is dominant to I^+, which in turn is dominant to I^-. It can also be used to show the O^C mutation only acts in *cis*- while the *lacI* mutation can act in *trans*- .

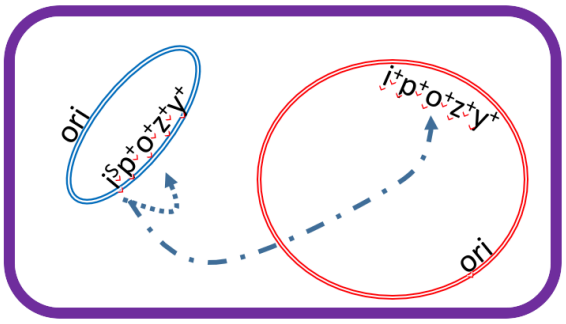

Figure 11.
A diagram showing a merozygote. This is a form of bacterial cell where a plasmid (blue) has a normal operon but a *lacI*^{S} allele for the repressor. Not all bacterial cells have plasmids. The chromosome (red) has a normal repressor and the operon has a *lacY*$^-$ allele. The super-repressor will inactivate both operons because it will bind to both operators (a trans-acting effect). Note that both the plasmid and chromosome have their own origins of replication. (Original-Nickle-CC BY-NC 4.0)

5. SUMMARY

In positive regulation, **low** levels of glucose (inducer) allow adenylate cyclase to produce cAMP from ATP, which binds to CAP protein. CAP protein can bind to DNA and increase the level of transcription. **High** glucose level halts adenylate cyclase from producing cAMP from ATP. Hence, cAMP will not bind to CAP protein and in turn, CAP will not bind to DNA and the level of transcription would be low.

In negative regulation, the repressor protein acts to prevent transcription. Inducer binds to repressor to alter conformation so it no longer binds to the operator sequence and transcription can take place. **High** levels of lactose (inducer) would allosterically inhibit repressor and therefore would not prevent transcription. **Low** levels of lactose would not cause inhibition to the repressor, so transcription would be prevented. Various forms of regulation in the lac operon are found in **Figure 12**.

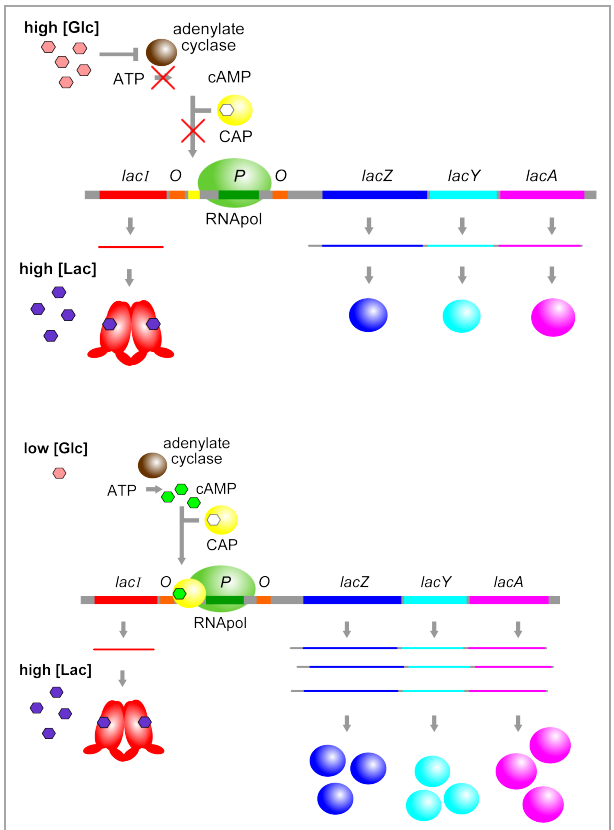

Figure 12.
When glucose [Glc] and lactose [Lac] are both high, the lac operon is transcribed at a moderate level, because CAP (in the absence of cAMP) is unable to bind to its corresponding cis-element (yellow) and therefore cannot help to stabilize binding of RNApol at the promoter. Alternatively, when [Glc] is low, and [Lac] is high, CAP and cAMP can bind near the promoter and increase further the transcription of the lac operon. (Original-Deyholos-CC BY-NC 3.0)

SUMMARY:

- Regulation of gene expression is essential to the normal development and efficient functioning of cells
- Gene expression may be regulated by many mechanisms, including those affecting transcript abundance, protein abundance, and post-translational modifications
- Regulation of transcript abundance may involve controlling the rate of initiation and elongation of transcription, as well as transcript splicing, stability, and turnover
- The rate of initiation of transcription is related to the presence of RNA polymerase and associated proteins at the promoter.
- RNApol may be blocked from the promoter by repressors, or may be recruited or stabilized at the promoter by other proteins including transcription factors
- The *lac* operon is a classic, fundamental paradigm demonstrating both positive and negative regulation through allosteric effects on *trans*-factors.

KEY TERMS:

gene expression
transcriptional regulation
operon
lactose
glucose
lac operon
lacZ
lacY
lacA
β-galactosidase
permease
trans-acetylase
P / promoter
O / operator
CBS
CAP-binding site
cis-elements
trans-regulators
lacI
homotetramer
repressor
inducer
allosteric
cAMP binding protein
CAP
CAP binding sequence
CBS
adenylate cyclase
constitutive
O^c / I^- / I^s
cis dominant
cis-acting factors
trans dominant
trans-acting factors
F-factor / episome
merozygotes

STUDY QUESTIONS:

1) List all points in the "central dogma" of gene action the mechanisms that can be used to regulate gene expression.

2) With respect to the expression of β-galactosidase, what would be the phenotype of each of the following strains of *E. coli*?

Use the symbols:
+++ Lots of β-galactosidase activity (100%)
+ Moderate β-galactosidase activity (10-20%)
− No β-galactosidase activity (0-1%)

a) I^+, O^+, Z^+, Y^+ (no glucose, no lactose)
b) I^+, O^+, Z^+, Y^+ (no glucose, high lactose)
c) I^+, O^+, Z^+, Y^+ (high glucose, no lactose)
d) I^+, O^+, Z^+, Y^+ (high glucose, high lactose)
e) I^+, O^+, Z^-, Y^+ (no glucose, no lactose)
f) I^+, O^+, Z^-, Y^+ (high glucose, high lactose)
g) I^+, O^+, Z^+, Y^- (high glucose, high lactose)
h) I^+, O^c, Z^+, Y^+ (no glucose, no lactose)
i) I^+, O^c, Z^+, Y^+ (no glucose, high lactose)
j) I^+, O^c, Z^+, Y^+ (high glucose, no lactose)
k) I^+, O^c, Z^+, Y^+ (high glucose, high lactose)
l) I^-, O^+, Z^+, Y^+ (no glucose, no lactose)
m) I^-, O^+, Z^+, Y^+ (no glucose, high lactose)
n) I^-, O^+, Z^+, Y^+ (high glucose, no lactose)
o) I^-, O^+, Z^+, Y^+ (high glucose, high lactose)
p) I^s, O^+, Z^+, Y^+ (no glucose, no lactose)
q) I^s, O^+, Z^+, Y^+ (no glucose, high lactose)
r) I^s, O^+, Z^+, Y^+ (high glucose, no lactose)
s) I^s, O^+, Z^+, Y^+ (high glucose, high lactose)

3) In the *E. coli* strains listed below, some genes are present on both the chromosome, and the extrachromosomal F^-factor episome. The genotypes of the chromosome and episome are separated by a slash. What will be the β^- galactosidase phenotype of these strains? All of the strains are grown in media that lacks glucose.

Use the symbols:
+++ Lots of β-galactosidase activity (100%)
+ Moderate β-galactosidase activity (10-20%)
− No β-galactosidase activity (0-1%)

a) $I^+, O^+, Z^+, Y^+ / I^+, O^-, Z^-, Y^-$ *(high lactose)*
b) $I^+, O^+, Z^+, Y^+ / I^+, O^-, Z^-, Y^-$ (no lactose)
c) $I^+, O^+, Z^-, Y^+ / I^+, O^-, Z^+, Y^+$ (high lactose)
d) $I^+, O^+, Z^-, Y^+ / I^+, O^-, Z^+, Y^+$ (no lactose)
e) $I^+, O^+, Z^-, Y^+ / I^-, O^+, Z^+, Y^+$ (high lactose)
f) $I^+, O^+, Z^-, Y^+ / I^-, O^+, Z^+, Y^+$ (no lactose)
g) $I^-, O^+, Z^+, Y^+ / I^+, O^+, Z^-, Y^+$ (high lactose)
h) $I^-, O^+, Z^+, Y^+ / I^+, O^+, Z^-, Y^+$ (no lactose)
i) $I^+, O^c, Z^+, Y^+ / I^+, O^+, Z^-, Y^+$ (high lactose)
j) $I^+, O^c, Z^+, Y^+ / I^+, O^+, Z^-, Y^+$ (no lactose)
k) $I^+, O^+, Z^-, Y^+ / I^+, O^c, Z^+, Y^+$ (high lactose)
l) $I^+, O^+, Z^-, Y^+ / I^+, O^c, Z^+, Y^+$ (no lactose)
m) $I^+, O^+, Z^-, Y^+ / I^s, O^+, Z^+, Y^+$ (high lactose)
n) $I^+, O^+, Z^-, Y^+ / I^s, O^+, Z^+, Y^+$ (no lactose)
o) $I^s, O^+, Z^+, Y^+ / I^+, O^+, Z^-, Y^+$ (high lactose)
p) $I^s, O^+, Z^+, Y^+ / I^+, O^+, Z^-, Y^+$ (no lactose)

4) What genotypes of *E. coli* would be most useful in demonstrating that the *lacO* operator is a *cis*-acting regulatory factor?

5) What genotypes of *E. coli* would be useful in demonstrating that the *lacI* repressor is a *trans*-acting regulatory factor?

6) What would be the effect of the following loss-of-function mutations on the expression of the *lac* operon?
a) loss-of-function of adenylate cyclase
b) loss of DNA binding ability of CAP
c) loss of cAMP binding ability of CAP
d) mutation of CAP binding site (CBS) *cis*-element so that CAP could not bind