Cellular & Molecular Biology
SQBS 1143

Chapter 3:
Transcription

Prof. Dr. Noor Aini Abdul Rashid
Dr. Chan Giek Far

How is the Genetic Information Used?

- During the day-to-day life of a cell, **working copies** of the genes are used.
- Genetic information can be carried by two kinds of molecules, **DNA** or **deoxyribonucleic acid**. The working copies of genes are made of **RNA** or **ribonucleic acid**, which is very similar in chemical structure to DNA.
- The particular type of RNA molecule that carries genetic information from the genes into the rest of the cell is known as **messenger RNA** usually abbreviated to **mRNA**.

- The transfer of information from DNA to messenger RNA is known as **transcription**.

- For a gene to be transcribed, DNA, which is double-stranded must first be **pulled apart temporarily**. Then a molecule of single stranded RNA is made.

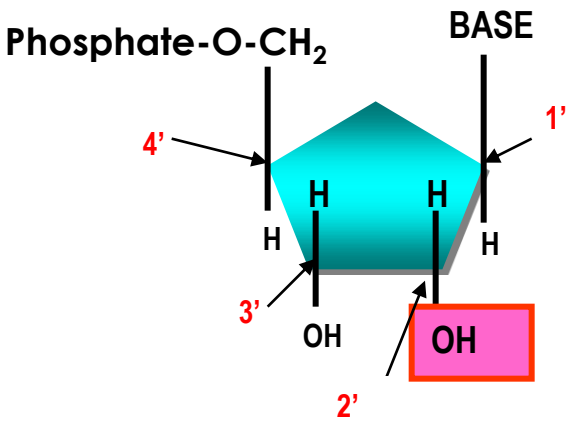
- This is the **messenger RNA** and it has base sequence complementary to that of the DNA strand used as template.

What is the Chemical Difference between DNA and RNA?

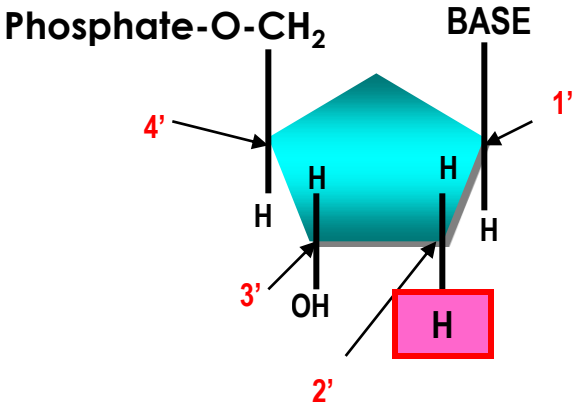
- There are two related kinds of nucleic acid, **deoxyribonucleic acid (DNA)** and **ribonucleic acid (RNA)**.

1. The first difference between them is that in DNA the sugar is always **deoxyribose**, whereas in RNA the sugar is **ribose**. As its name suggests, **deoxyribose** has **one less oxygen atom than ribose**. It is this initial difference which gives the D in DNA versus the R in RNA.

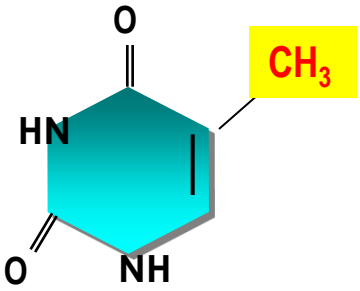
Differences between DNA and RNA



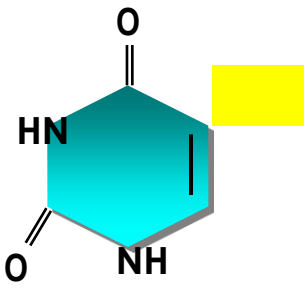
RNA



DNA



THYMINE (DNA)



URACIL (RNA)

2. The second difference is that in RNA, the **base thymine** (T) is replaced by the closely related base **uracil** (U) . Wherever you find thymine in DNA, you get uracil in RNA.

Hence uracil in RNA and thymine in DNA, convey the same genetic information. So, if you include RNA with DNA, the genetic alphabet has five letters (**A, C, G, T and U**) .

3. The third and final difference between DNA and RNA is that **DNA** is **double stranded** (ds), whereas **RNA** is **normally single stranded** (ss).

Which strand to copy ???

- Thus, when a gene made of DNA is transcribed into an RNA message, only **one of the strands of DNA is copied**. The sequence of the RNA message is complementary to the **template strand** of the DNA upon which it is synthesized.

- Apart from the replacement of thymine in DNA with uracil in RNA, this means that the **sequence of the new RNA molecule** is **identical** to the **sequence of the DNA**, the one not actually used during transcription.

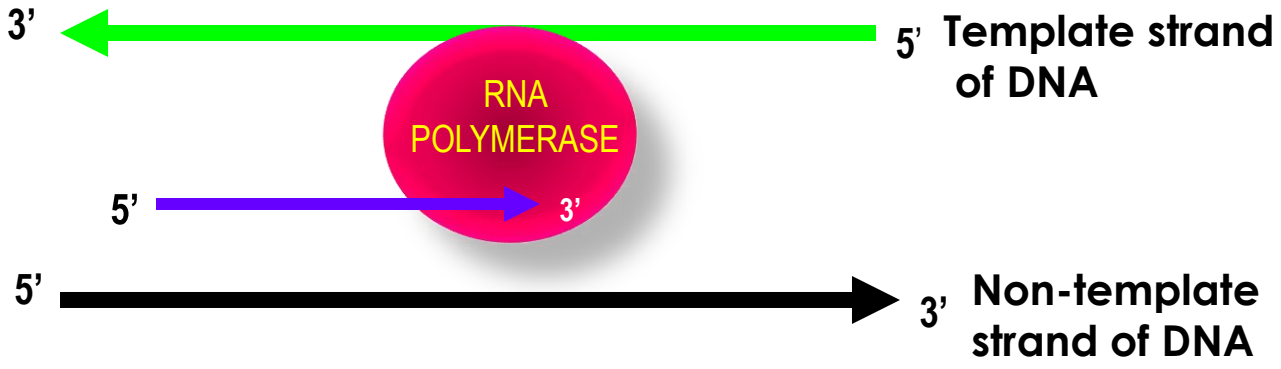
Short segments of the chromosome are turned into messages.

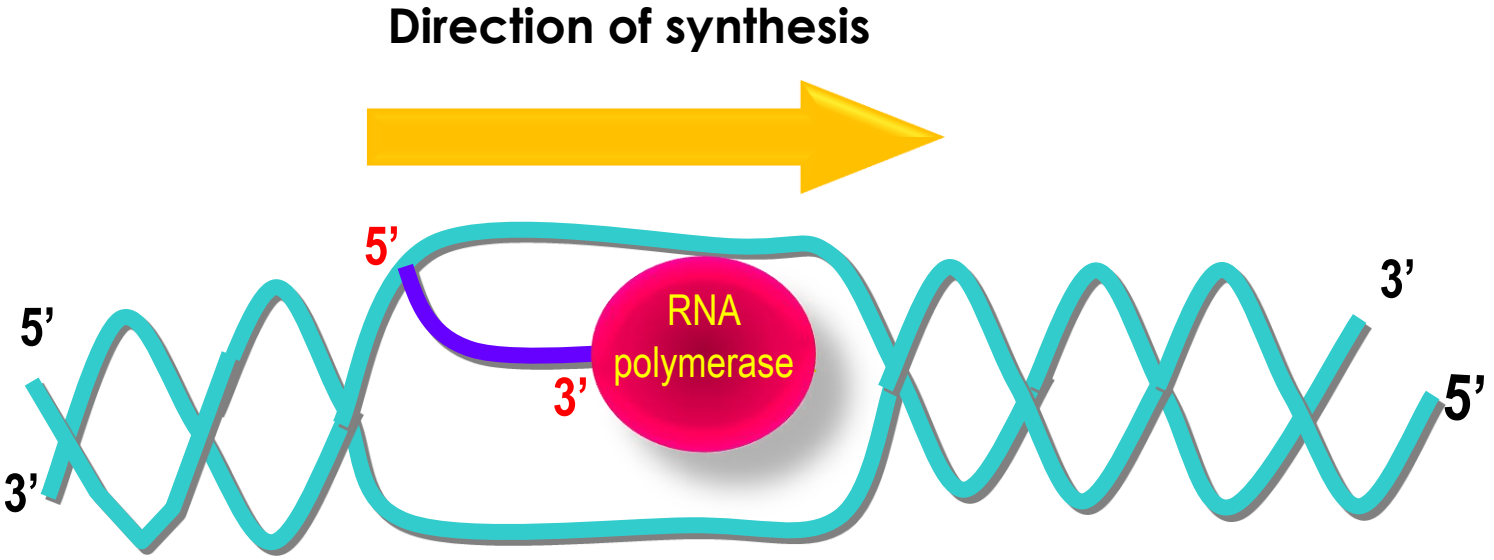
- Although a chromosome carries hundreds or thousands of genes, **only a fraction** of these are used at any given time.
- In a typical bacterial cell, about **30 percent** of the **genes are in use at any particular time**. In the cells of higher organisms having many more genes, the proportion in use at a time is much smaller.
- During cell growth, each gene or small group of related genes, is used to generate a separate RNA copy when, and if it is needed. Consequently there are many different messenger RNA molecules.
- Each of these mRNA molecules carries the information from a short segment of a chromosome.

Messenger RNA is made by RNA Polymerase

Function of RNA polymerase :

- **binds to the DNA** at the start of a gene and
- **opens the double helix**
- It then **manufactures an RNA message**

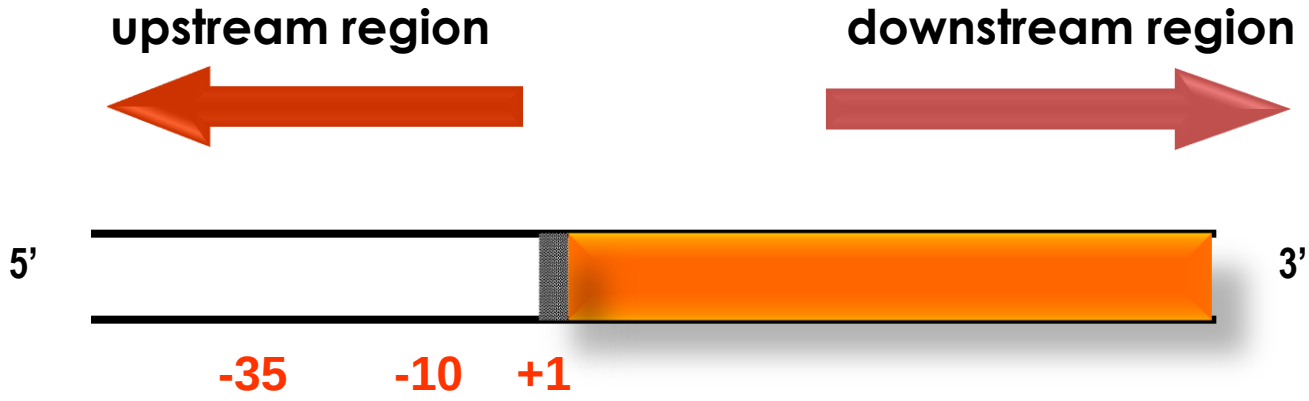



Direction of synthesis
5’
5’
3’
RNA
polymerase
3’
3’
5’

How is the Beginning of a Gene Recognized?

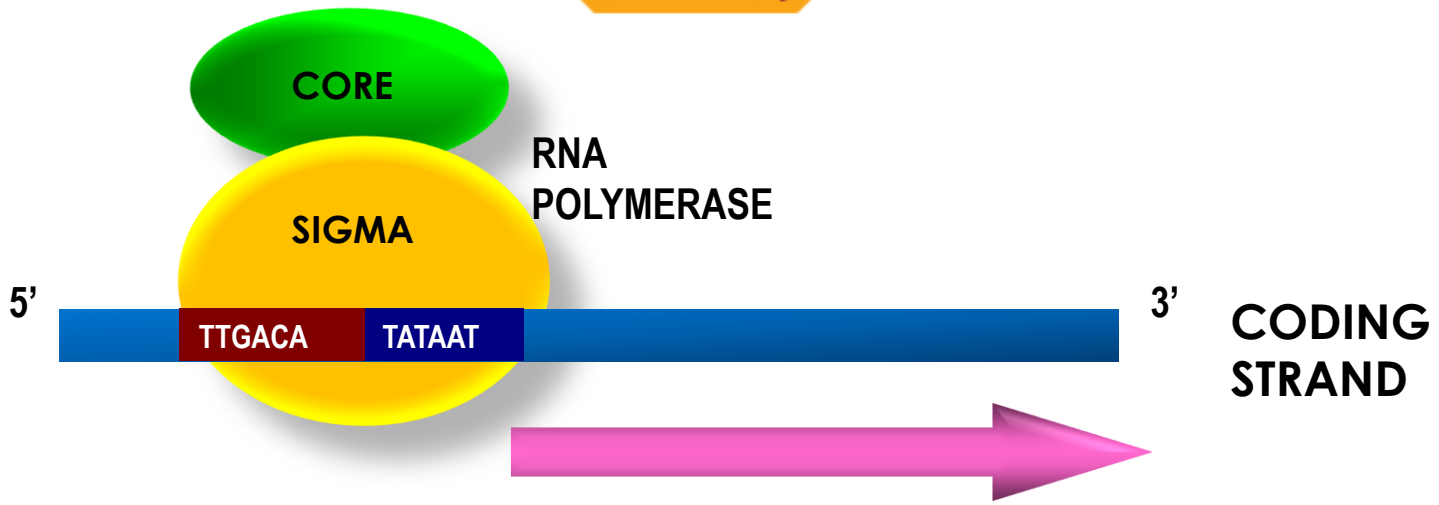
- We will only discuss about bacteria because they're much simpler. The principles of transcription are similar in higher organisms.
- RNA polymerase is made up of **several protein subunits** with different roles.
- A special subunit of bacterial RNA polymerase, the **sigma subunit** recognizes two special sequences of bases in the coding (non-template) sequences known as
  - the **-10 region** and
  - the **-35 region**;

  (because they are found by counting backwards 10 or 35 bases from the first base of the gene).



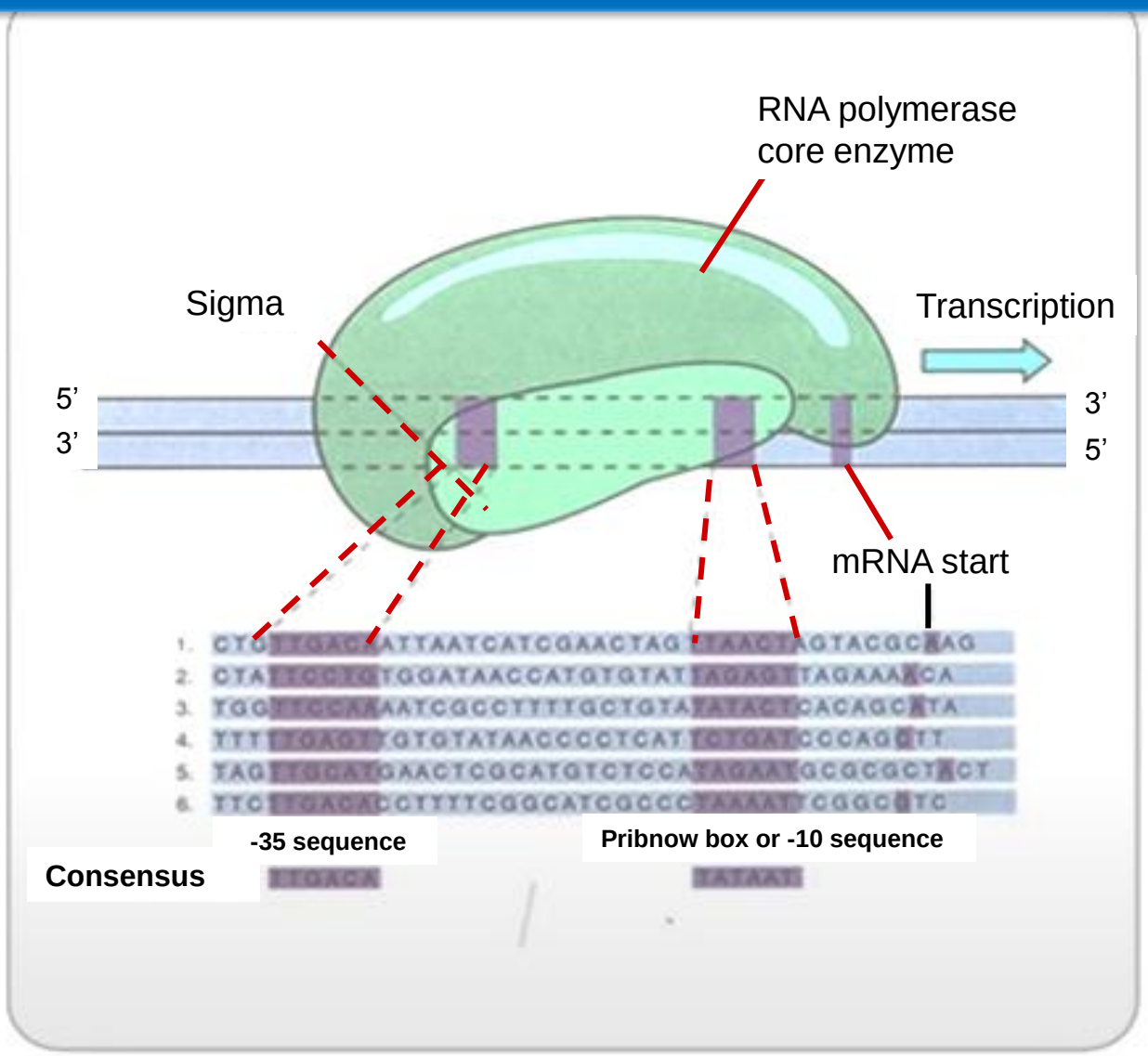
- The stretch of DNA region front of a gene is often referred to as the **upstream region** and the region where RNA polymerase binds is known as the **promoter**.

- To bind the sigma subunit properly, the base sequence needed at -10, **TATAAT** and the sequence at - 35, **TTGACA**. Such theoretically perfect sequences in this case (TATAAT and TTGACA) are known as **consensus sequences**.



- Consensus sequences are found by comparing many real life sequences and taking the average. In real life, **a few highly expressed genes** do have these **exact sequences** in their promoter.
- However, in practice, the **-l0** and **-35 region sequences are rarely perfect**, but as long as they are only "wrong" (not exact) by one or two bases, the sigma subunit will still recognize them. The **strength of a promoter** depends partly on **how closely matched** the ideal consensus sequence.

Promoter Sequence in Prokaryotes

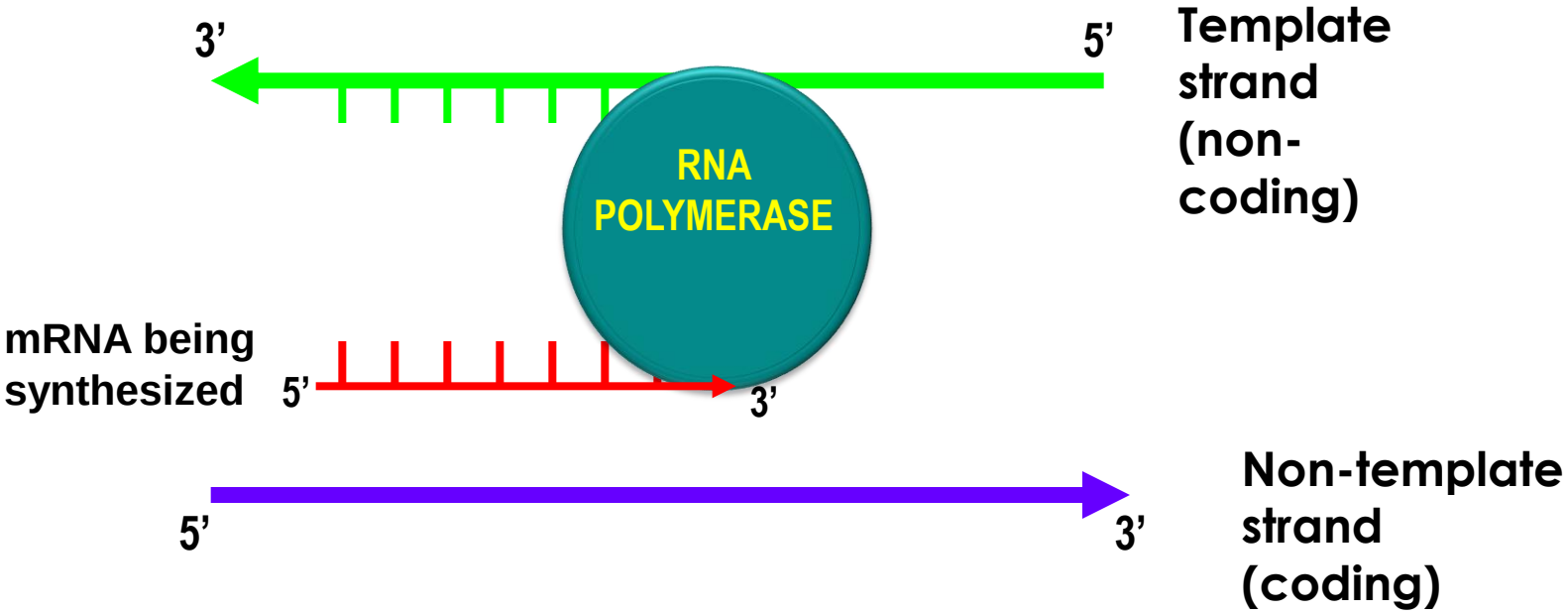


Manufacturing the Message

- Once sigma has found a promoter and the RNA polymerase has successfully bound to it, the sigma subunit drops off. The remaining part of bacterial RNA polymerase, known as the **core enzyme** then makes the mRNA.

- The DNA **double helix is opened up** and a **single strand of RNA** is generated **using one of the DNA strands** as a **template** for matching up the bases.

Manufacturing the Message



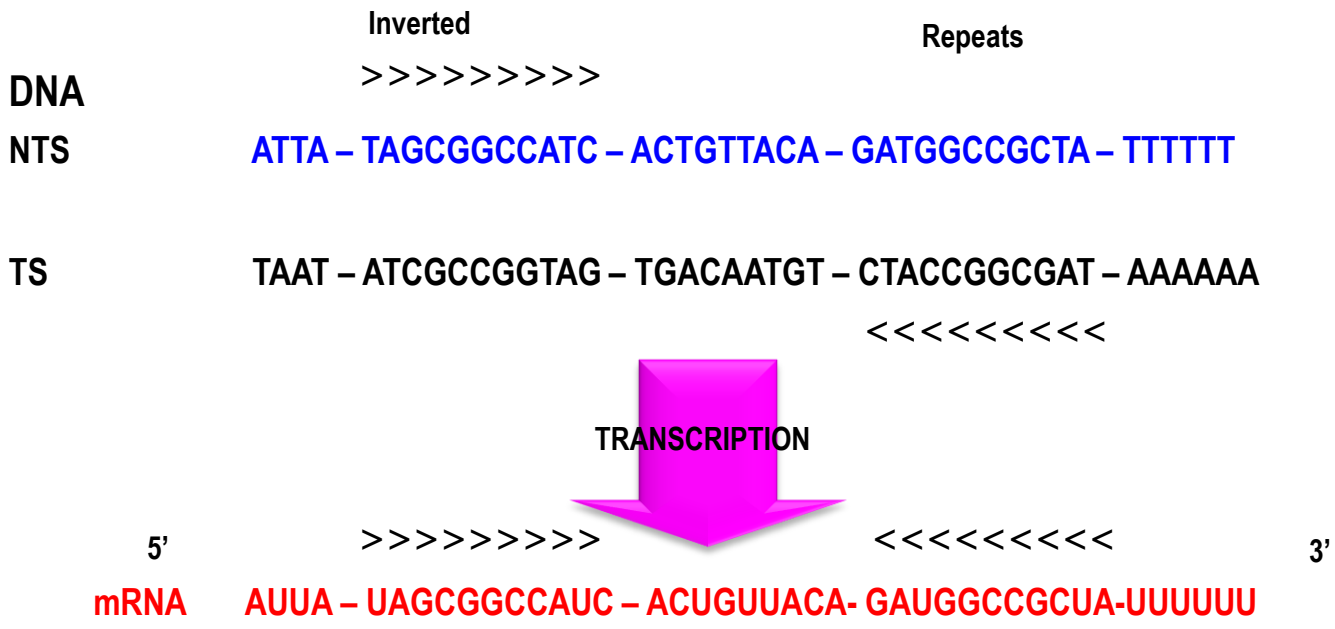
Template strand (non-coding)

Non-template strand (coding)

How Does RNA Polymerase Know Where to Stop

- Just as there is a special recognition site at the front of each gene, there is a special sequence at the end.
- A **terminator sequence** consists of **two inverted repeats separated** by **half a dozen bases** followed by **a string of As** (in the template strand of the DNA).
- The following figure will show you what is meant by this.

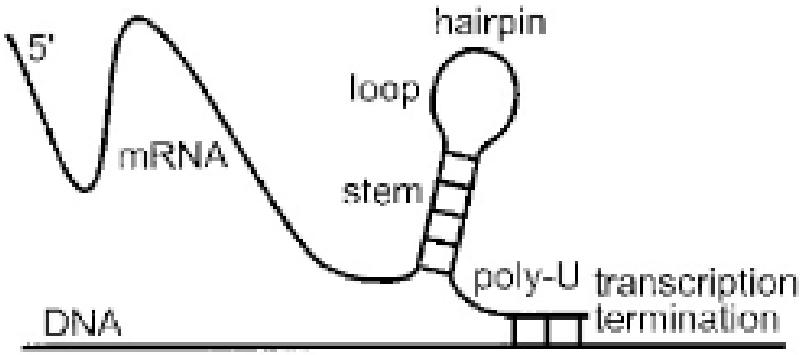
Termination Recognition



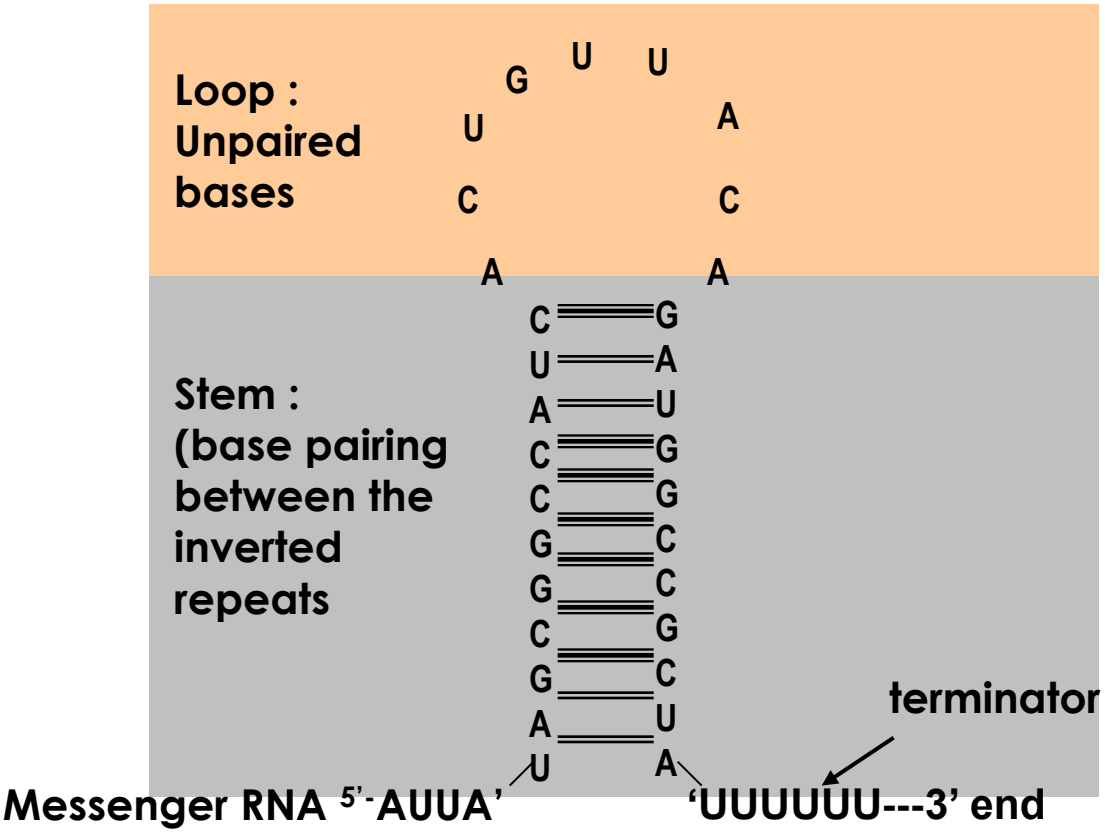
- Note that the **two inverted repeat sequences** are actually on **opposite strands** of the DNA. The **sequence of the mRNA** will be the **same as the non-template strand** of DNA except for the substitution of U for T.

- Although we often talk as if the corresponding single stranded mRNA as "inverted repeats," its second "repeat" is actually the complement of the inverse of the first.

- Because of this, such inverted repeat sequences on the same strand of an RNA molecule can pair up to generate a **"stem and loop"** or **"hairpin"** structure.

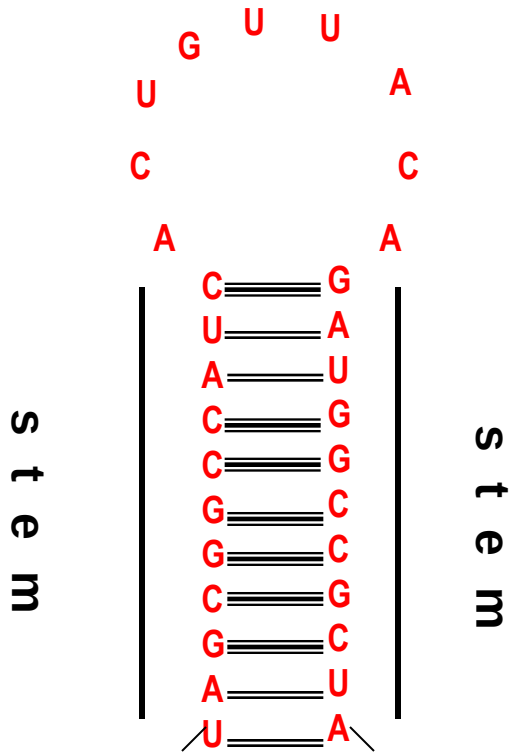


Inverted Repeats MakesA Stem and Loop



	stem	loop	stem
mRNA	AUUA – UAGCGGCCAUC –	ACUGUUACA-	GAUGGCCGCUA-UUUUUU

FOLDS



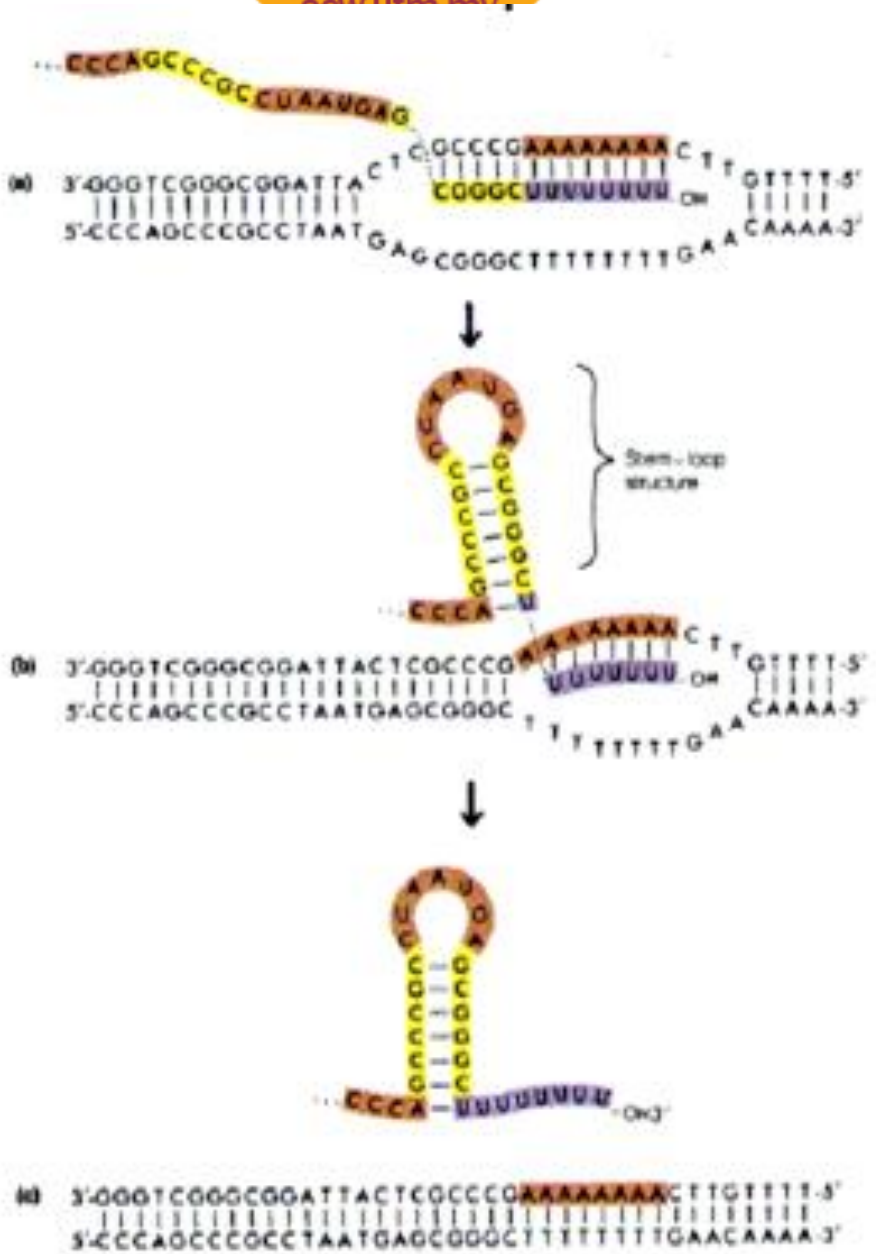
- The **string of As** in the DNA **gives rise to a run of Us** at the 3' end of the mRNA. Once the RNA polymerase reaches the stem and loop it **stops**.

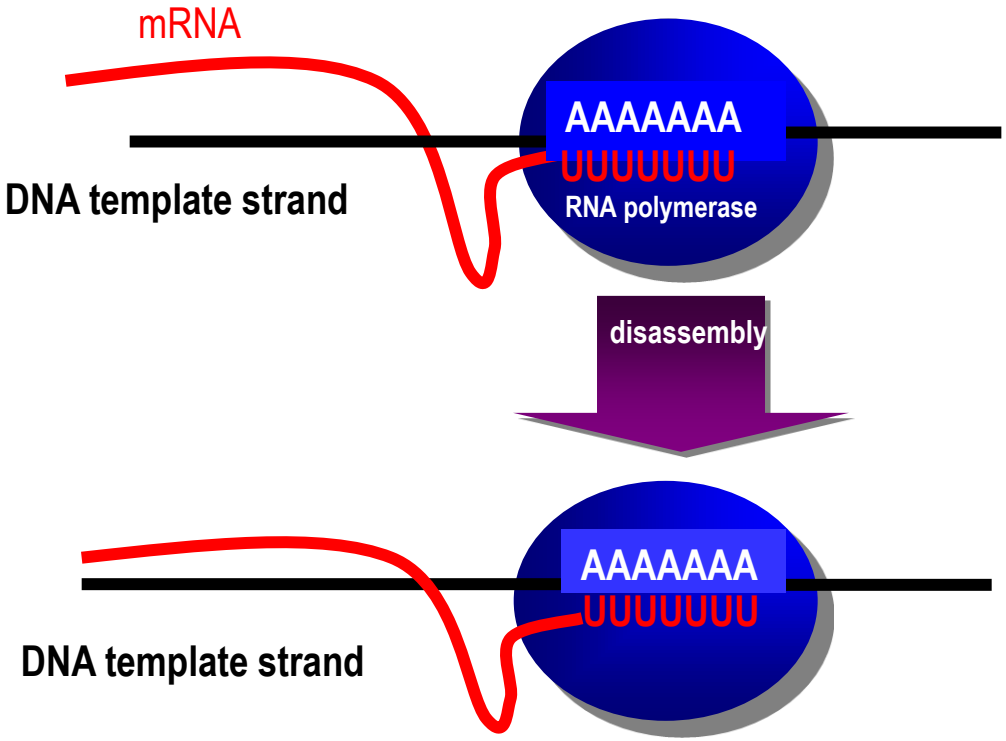
- Long RNA molecules contain **lots of possible hairpin structures** which cause RNA polymerase to slow down or stop briefly, depending on the size of the hairpin.

- This provides an opportunity for termination but, if there is no string of Us, the RNA polymerase will off again.

- However, a string of Us paired with a string of As in the template strand of DNA is a **very weak structure** and the RNA and DNA just fall apart while the RNA polymerase is idling.

- Once the DNA and RNA have separated at the terminator structure the RNA polymerase falls off and wanders away to find another gene.


Stem - loop structure


mRNA
AAAAAAA
UUUUUUU
RNA polymerase
DNA template strand
disassembly
AAAAAAA
UUUUUUU
DNA template strand

How Does the Cell Know Which Genes to Turn On?

- Although each gene has a promoter and a terminator for starting an finishing the synthesis of messenger RNA, this still does not tell us when to turn on a gene.

- Some genes are switched on all the time. They are sometimes known as **housekeeping genes** and they are said to be **expressed constitutively**. Most of these housekeeping genes have both their -10 and -35 region promoter sequences very **close or identical to consensus**.

- Consequently they are always recognized by the sigma subunit of RNA polymerase and are **switched on automatically under all condition**. Genes which are only needed under certain conditions usually have **poor recognition sequences** in the **-10 and -35 regions** of their promoters.

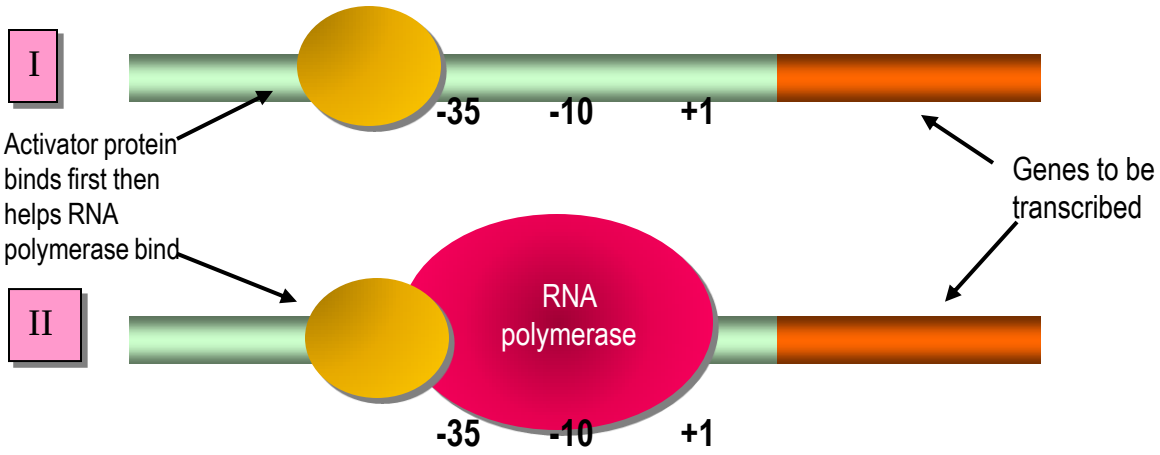
- In such cases the **promoter sequence is not recognized** by the sigma subunit unless another **accessory protein** is there to help.

- These accessory proteins are known as **gene activator protein** and are **different** for different genes.

- Each activator protein **may recognize one or more genes**.
- **A group of genes** which are all **recognized by the same activator protein** will be **expressed together under similar conditions**, even if the genes are at different places on the DNA.
- Higher organisms have many genes which are often expressed differently in different tissues.
- As a result, eukaryotic genes are often controlled by multiple activator proteins also known as **transcription factors**. So for now we'll stick to bacterial genes as examples.

What Activates the Activator?

- In living cells, especially in more complex higher organisms, there may indeed be **a series of regulators**, each regulating the next. Ultimately, however, the cell must respond to some outside influence.

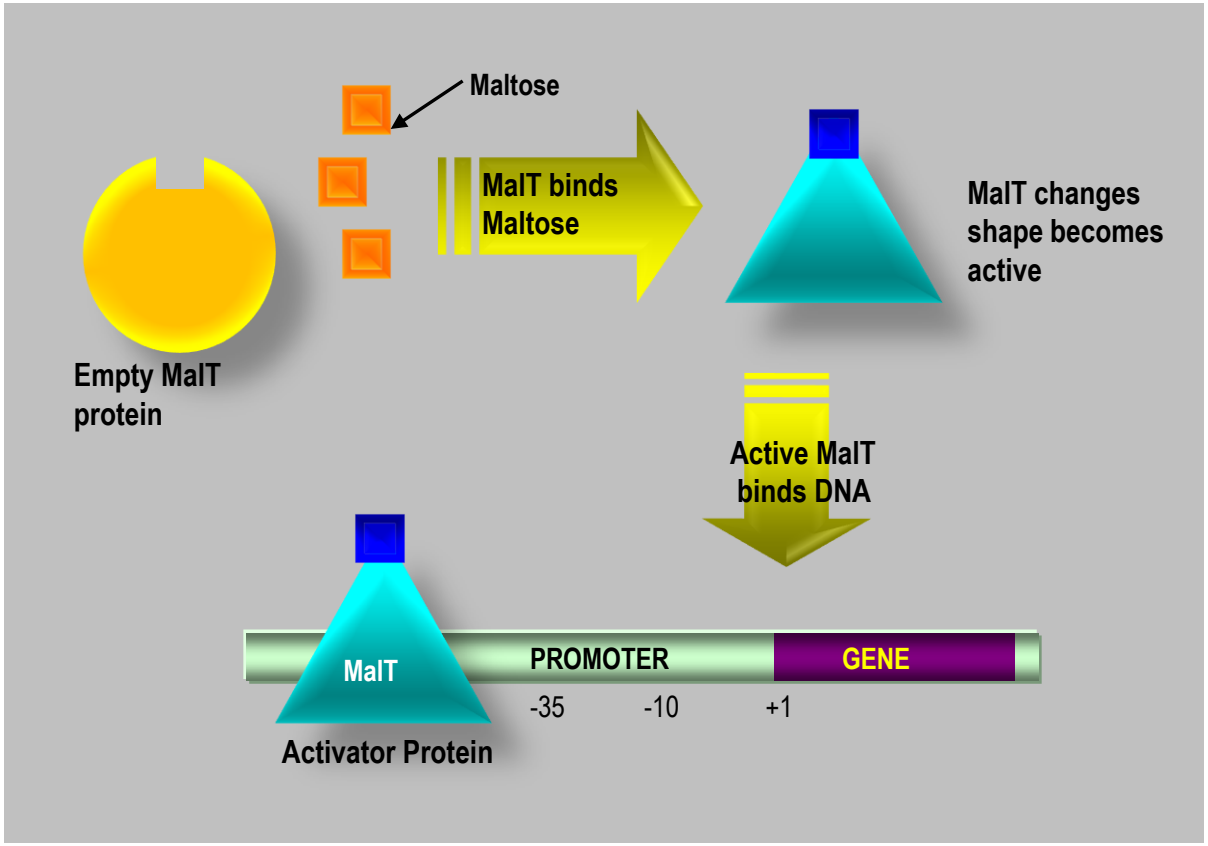


Activator

- As a simple example of an **activator**, let's consider the use in the bacterium ***Escherichia coli***.

- Maltose is a sugar made originally from the starch in malt. *E. coli* can grow using this sugar to satisfy all of its needs for energy.

MalT- Activator protein



- An activator protein, called **MalT**, detects maltose by binding to it. The Malt protein changes shape when it binds maltose. The original "empty" form of MalT cannot bind to DNA.
- The **active form** (**MalT + maltose**) can **bind to DNA** and it **finds the genes** needed for growth on maltose and **activates them**. The result of this is that **the genes intended for using maltose** are only expressed when this particular sugar is available.
- The same general principle applies to most nutrients although the details of the regulation often vary from case to case.

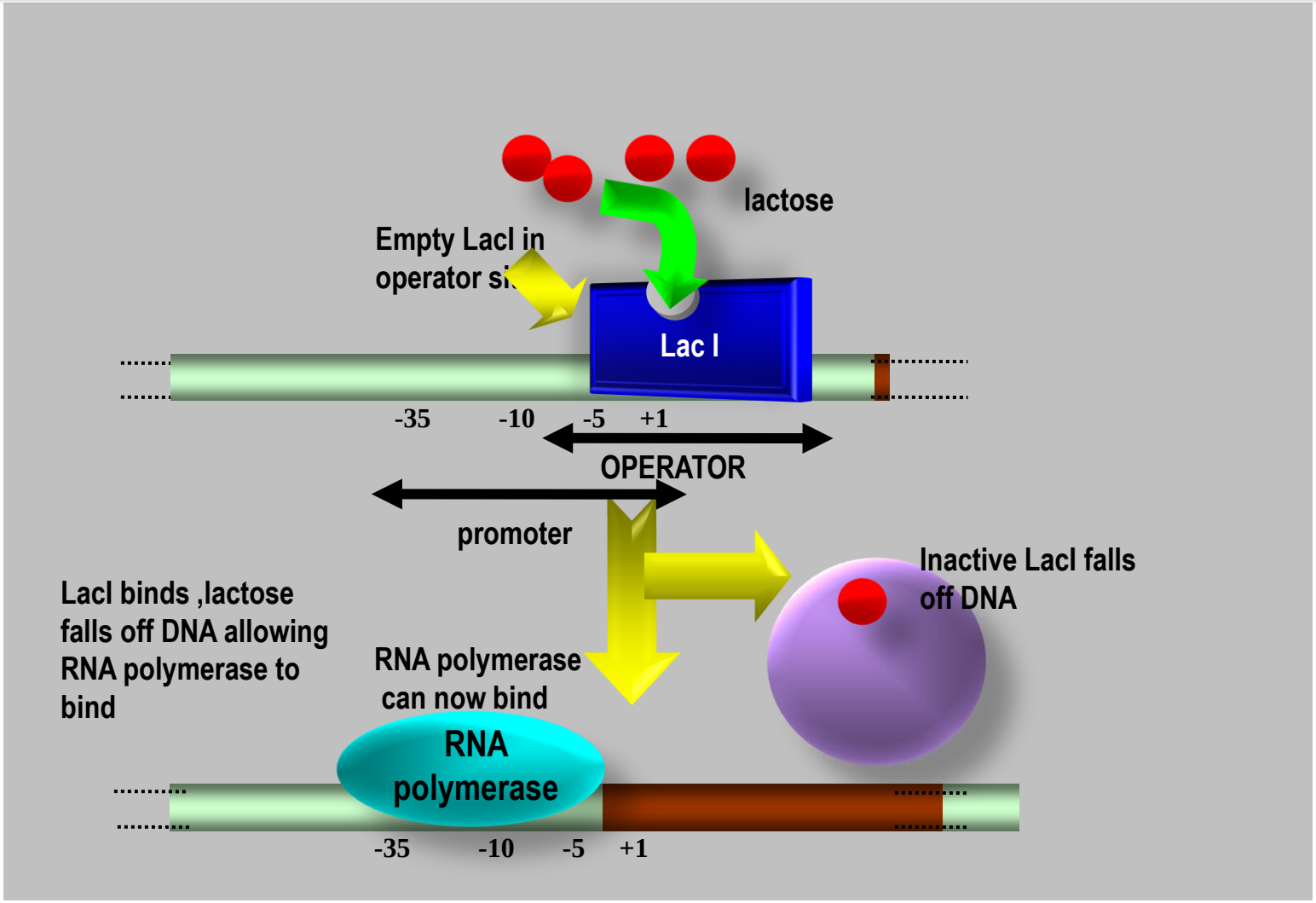
Negative Regulation

- Just as there are activator proteins which help turn genes on, there also proteins that can **turn genes off**.

- Historically, these negative regulators are actually discovered first. They are known as **repressors** and they work in a similar way to activators except they **have the opposite effect**.

- The best known example is the **lactose** and the **LacI protein**. Lactose is another sugar, found in milk, which bacteria like *E. coli* can grow on.

Negative Regulation: LacI Repressor Protein Detects Lactose



When there is no lactose….

- If no lactose is available the **LacI protein binds to the stretch of DNA** between the promoter and the genes for using lactose. The **site where a repressor binds** is called the **operator** sequence.
- The **repressor blocks the binding** of RNA polymerase, simply by getting in the way. When **lactose is present** it will **bind to the LacI protein**. The LacI protein then changes shape and falls off the DNA.
- Now the RNA polymerase can bind, and the genes for using lactose are switched on. The overall result is the same as for maltose: when **lactose is available**, the **genes** for using it **are switched on** and when there is **no lactose**, the **genes are turned off.**

Most Regulator Proteins Bind Small Molecules

- Whether our regulator protein is an activator or a repressor, we need to provide it with a signal of some sort. The most common way to do this is using some small molecule which fits into a binding site on the regulatory protein. This is called the **signal molecule**.

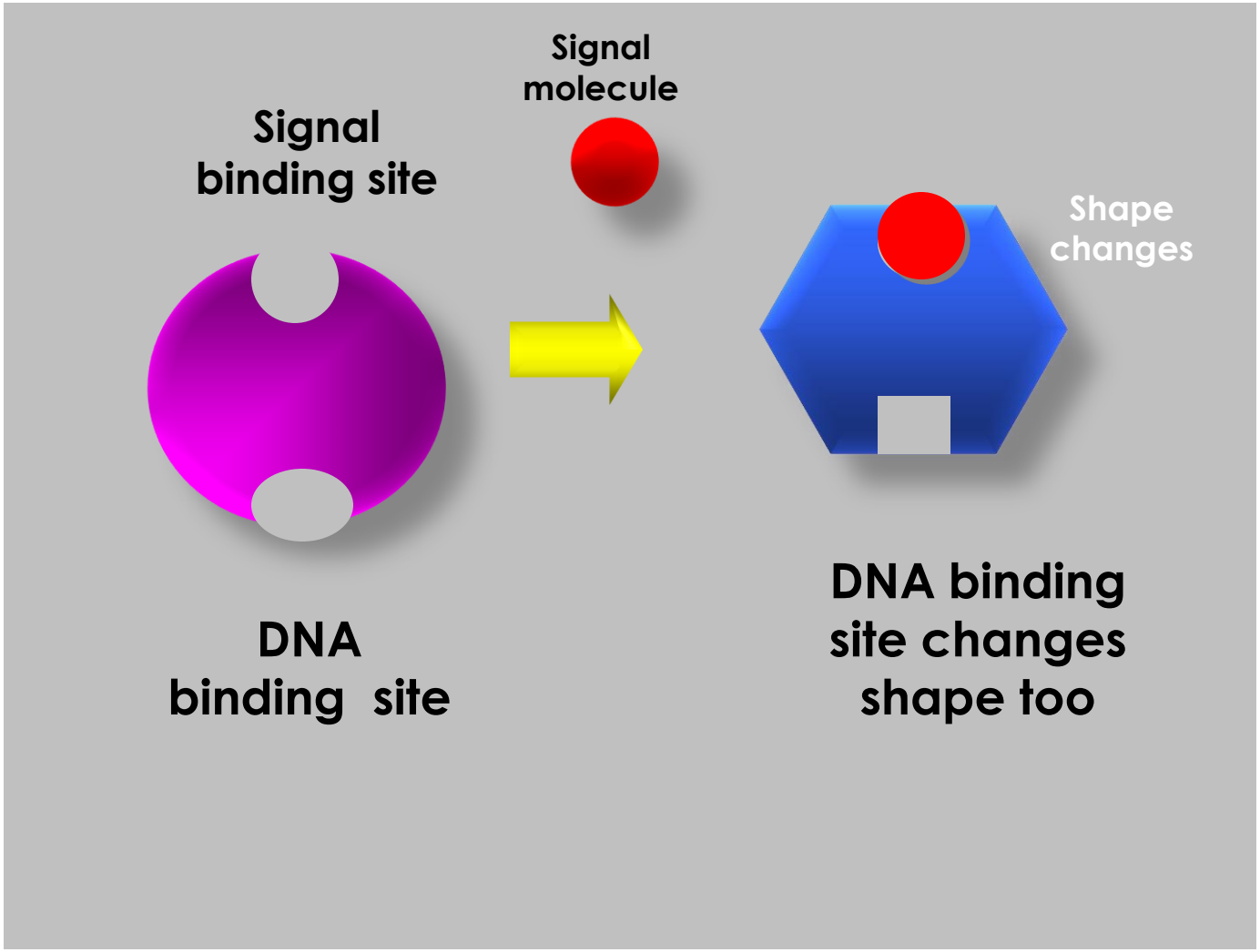
- In the case of using a nutrient for growth, the obvious choice is the **nutrient molecule** itself. As we have seen, this is true for the lactose repressor and the maltose activator.

Most Regulator Proteins Change Shape

- When a regulator protein binds to signal molecule it changes shape. Regulator proteins have two alternative forms, **the DNA binding form** and the **non-binding form**.

- Binding, or loss of the small signal molecule, causes the larger protein to change between two alternative shapes. Proteins that operate by **changing shape** in this manner are called **allosteric proteins**.

Regulatory protein with binding site

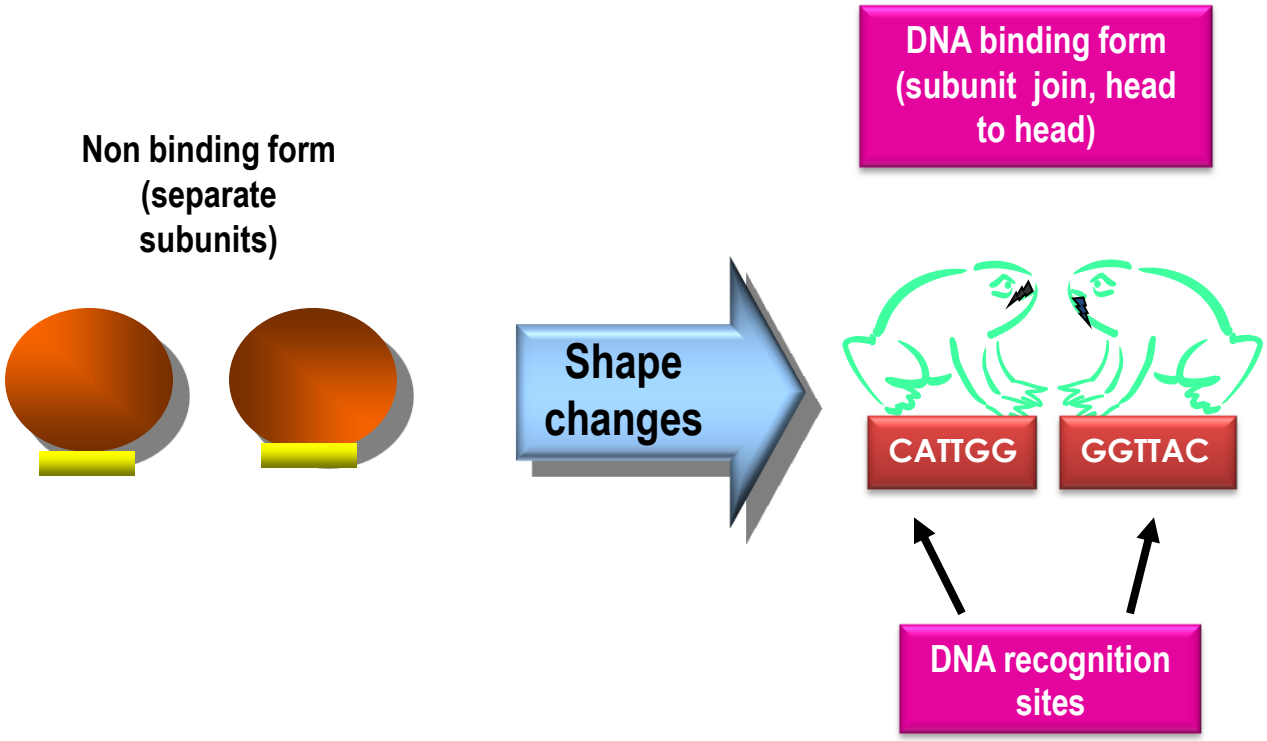


Most Regulator Proteins Have Two or Four Subunits

- Almost all real regulator proteins act as **pairs or in groups of four**. All of the subunits bind the signal molecule and then they **all change shape together**. Because there is an even number of protein subunits bound to the DNA, the recognition site on the DNA is also duplicated - well sort of.

- Actually the recognition site is not a direct repeat but an **inverted repeat**. This is because the subunits of the regulator protein **bind to each other head to head rather than head to tail**.

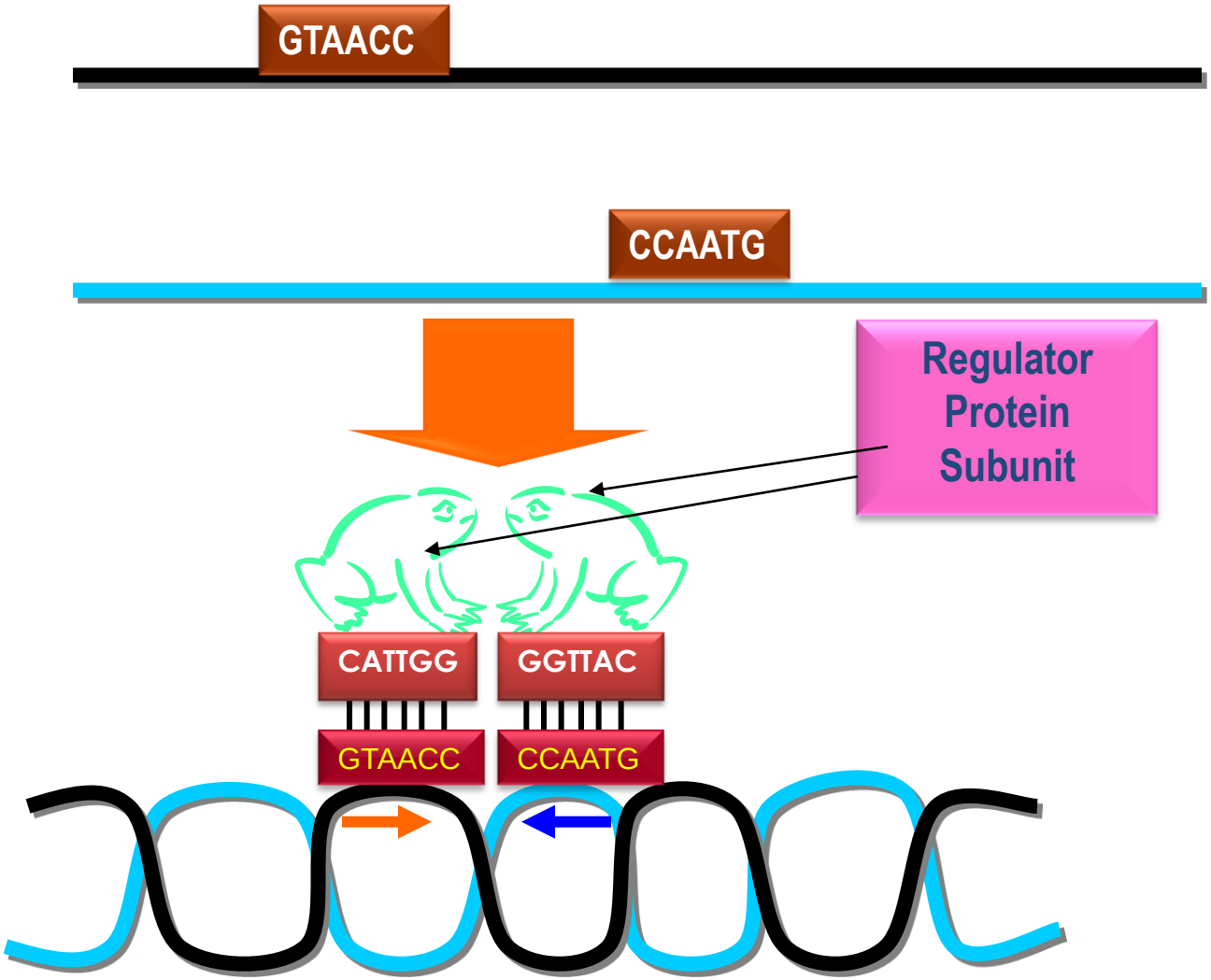
Subunit of a DNA binding protein



- Consequently, the **two protein molecules** are **pointing in opposite directions**. Because they have **identical binding sites for DNA**, they recognize the same sequence of bases, but in opposite direction.

- Therefore, **one protein subunit binds to the recognition sequence** on the **template strand** of the double helical DNA, and **its partner** binds to the **same sequence** but on the **non-template strand** of the DNA pointing in the opposite direction.

Inverted Repeats Binding with Protein



- This is actually simpler in practice than it sounds, precisely because the **DNA molecule is a double helix**, and **twists around to accommodate** the proteins most easily this way.

- Although the **two recognition sequences** are **on different strands they end up on the same side** of the DNA molecule due to its helical twisting.

Crp Protein - An Example of a Global Control Protein

- So far we have considered how to control genes for single functions such as using a particular sugar for bacterial growth. We must now consider the **coordinated control of large groups of genes**.
- This is known as **global regulation** and the proteins in charge of it are called **global regulators**. The **Crp** global control protein is in charge of **selecting from the menu** which nutrients to use for growth in bacteria like *E. coli*.
- Just as those of us who has our favourite food, bacteria also select their favourite foods when given a choice.

- Many bacteria can grow on a wide range of possible sugars such as fructose (fruit sugar), lactose (milk sugar), maltose (from starch breakdown) as well as glucose. When given a mixture of glucose, fructose and maltose, *E. coli* will use the **glucose** and ignore others.

- In molecular terms, this means switching off genes for using all of the other sugars when glucose is available. The Crp protein is a **global activator** that is **required for switching on the genes** for using **other sugars than glucose** such as maltose, for lactose and for all of the alternative nutrients to glucose.

Crp Protein – global activator

- The **Crp protein** is **allosteric**, like the MalT and LacI proteins. In order to bind DNA and activate genes, the Crp protein must first bind to a small **signal molecule** known as **cyclic AMP**. Maybe you have been wondering what Crp stands for.
  Crp : - **C**yclic AMP **R**eceptor **P**rotein.

- **Cyclic AMP is a global signal** that the **cell has run out of glucose**, its favorite energy-source. Only when this has occurred can the genes for using less favored nutrients be switched.

- Consequently, in order to switch on genes for using any individual sugar, say, lactose, we need both an **individual signal**, availability of lactose, and **a global signal**, cyclic AMP which signals the need for nutrition.

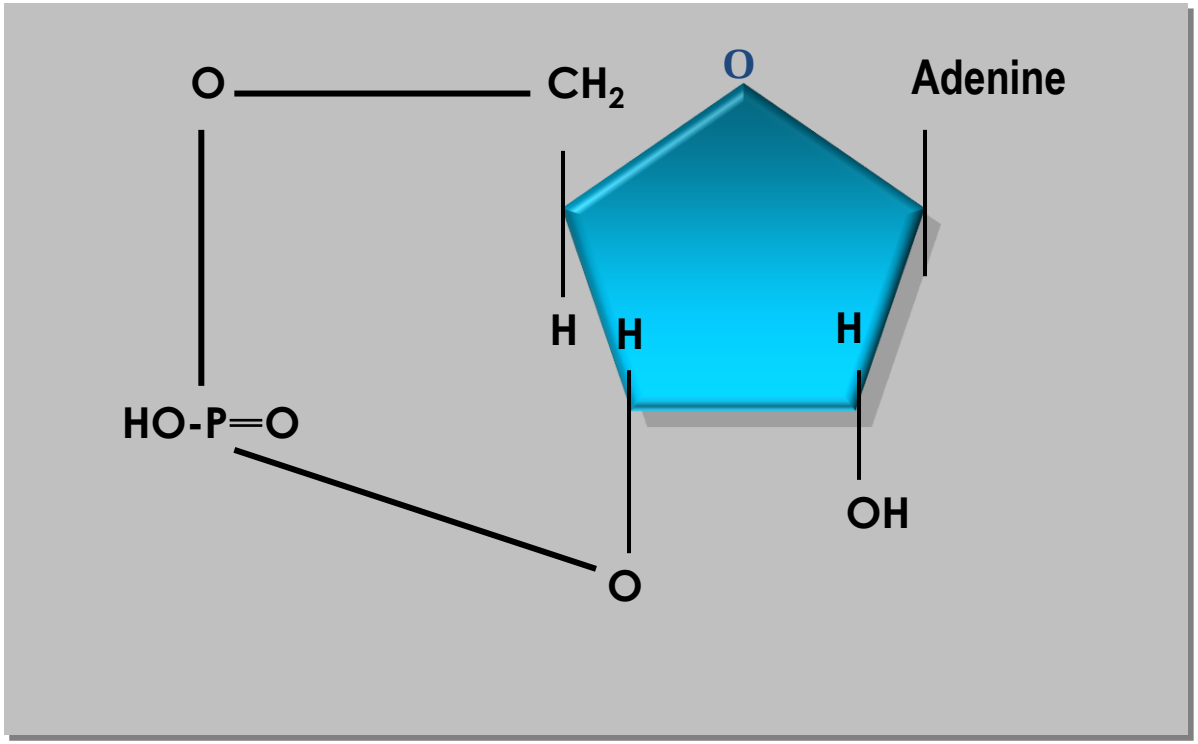
- In practice, most genes respond to two signals, sometimes more.

- Usually one is a **specific signal** and the other is a more **general signal** that applies to many genes.

Regulatory Nucleotides

- **Cyclic AMP** is a **cyclic version of adenosine monophosphate** in which the phosphate group is bent around and attached to both the 5' and 3' positions of the ribose sugar.
- Although it is not used as a building block when making nucleic acids, cyclic AMP is nonetheless a **nucleotide** of some sort.
- A variety of modified nucleotides are used by cells as signal molecules and consequently called **regulatory nucleotides**. Like cyclic AMP, they are mostly used as global signals.
- Another example is **isopentenyl adenosine**, found in plants where it acts to control cell division.

cAMP



The Operon Model for Gene Regulation

- The above scheme for regulating bacterial genes was first proposed by Francois Jacob ("Fronswa Zhakob") and Jacques Monod ("Zhak Mono") using the lactose genes as an example.
- Since then, a vast number of bacterial genes have been fitted to this model or slight variants of it. Jacob and Monod named the various components of this scheme, the repressor, the operator, etc .

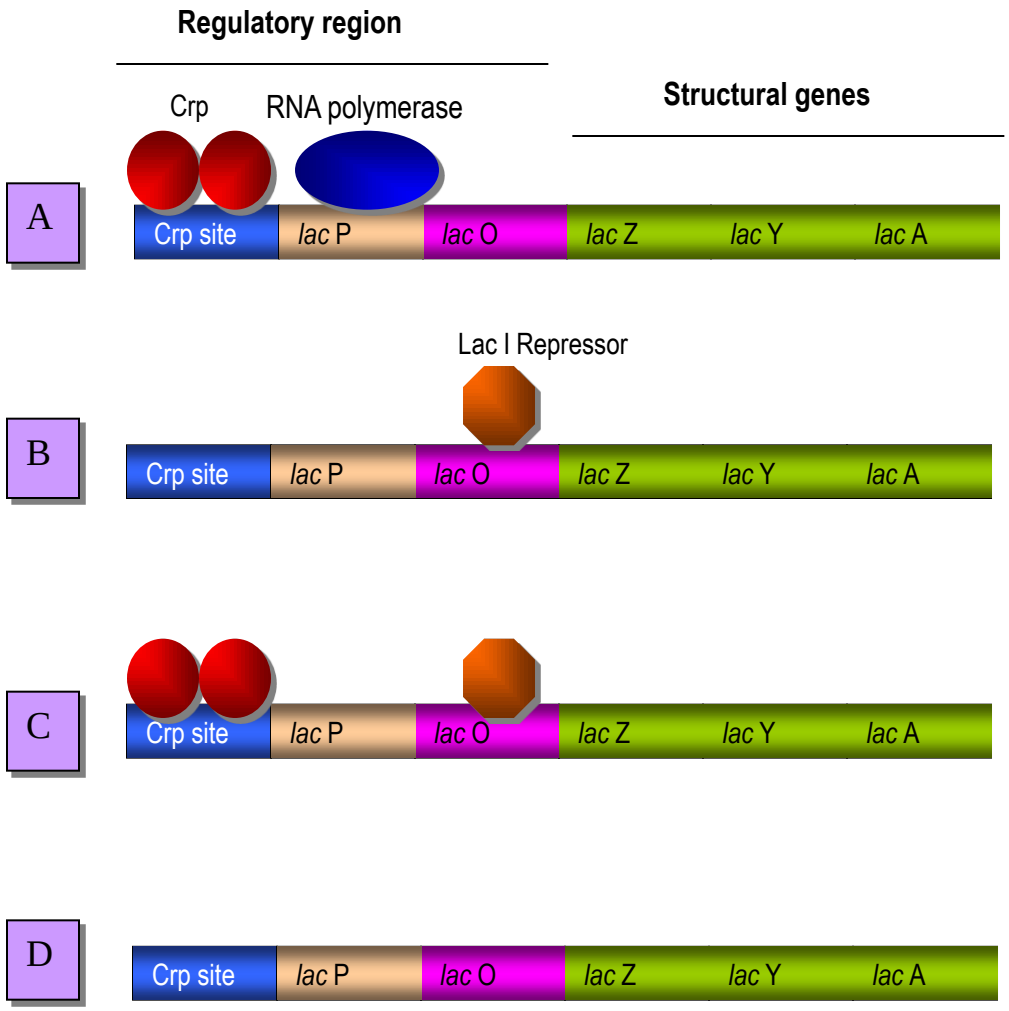
The Operon

- Up till now we have talked as if each gene had its own promoter and regulatory sites. In fact, many bacterial genes are found **in groups** that **are transcribed together from the same starting point** to give a single messenger RNA.

- A cluster of genes all switched on together by being transcribed from the same promoter is known as an **operon**. Despite having more genes than bacteria, **higher organisms do not have operons**; the genes are **regulated one at a time**.

Lac Operon

- Nonetheless, genes of higher organisms are regulated by the binding of control proteins, **both global** and **specific** in front of the gene.
- Some **operons have only a single gene**, most have **two** to **half a dozen** and a few have more. Geneticists have an obsession with abbreviations, if possible, of three letters.
- Another convention is to write gene names in italics. Thus, the **lactose operon** is generally known as the ***lac* operon**. The *lac* operon consists of three genes, ***lac* A**, ***lac* Y**, and ***lac* Z**.
  - ***lac* A - transacetylase**
  - ***lac* Y- Lac permease**
  - ***lac* Z- β-galactosidase**



	Crp protein	Repressor	RNA polymerase	On or Off
A	Present	Absent	Binds	On
B	Absent	Present	Cannot bind	Off
C	Present	Present	Cannot Bind	Off
D	Absent	Absent	Cannot bind	Off

- Whether or not the *lac* operon is switched on or off depends on two regulator proteins, **Lac I** and **Crp**. The various possibilities are illustrated in the last page.

- Only when the repressor, **LacI**, **is absent** and the **Crp protein is present** to give a helping hand, can the **RNA polymerase bind to the promoter** and make the messenger RNA.

Regulation by Antisense RNA

- As we have seen, messenger RNA is transcribed using **only one DNA strand** as the **template strand**. The other strand of DNA is not used. But suppose we did use the non-template strand and transcribed RNA from it?

- We would produce an **RNA molecule complementary** in sequence to the **mRNA**. This is known as **antisense RNA** and can base pair with its complementary mRNA, just as the two strands of DNA in the original gene base pair with each other.

Antisense in Bacteria and Higher Organisms

- Antisense RNA is occasionally used in **gene regulation by bacteria** and **higher organisms**.
- If antisense RNA is made, it will base pair with the mRNA and prevent it from binding to the ribosome.
- Consequently the **mRNA cannot be translated** to make protein and the gene is 'effectively switched off', even though mRNA has been made.

- In practice, antisense RNA is not made by transcribing the non-template strand of the same gene as the mRNA. Another, quite distinct **"anti-gene"** is used for making the antisense RNA.

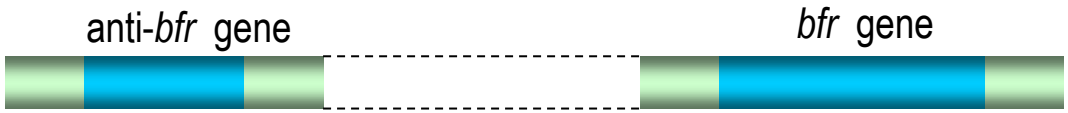
- **Bacterioferritin protein** used by bacteria to **store surplus iron atoms**. The ***bfr*** gene encodes bacterioferritin itself and the **anti-gene** encodes the **antisense RNA**.

- Since only relatively **short piece of antisense RNA** is needed to **block** the mRNA, the **anti-gene** is similar in sequence but **shorter than the original gene**.

What will happen if iron content is high or low?

- When the iron concentration in the culture medium is **low**, bacterioferritin is not needed, but it is made if the iron level goes up. Bacterioferritin is used to store surplus iron atoms.

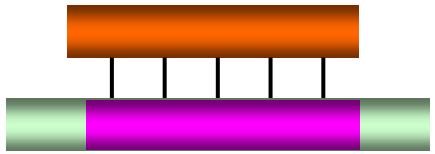
- The ***bfr* gene** itself is **transcribed to give mRNA in both conditions**. However, the **anti-bfr gene is only transcribed to give antisense RNA** in **low iron**. This prevents -synthesis of the bacterioferritin protein when iron is limited.



transcription

Anti-bfr messenger

If anti-*bfr* RNA is present it binds mRNA



messenger RNA with anti *bfr* cannot be translated into protein

bfr messenger RNA

In absence of anti-*bfr* RNA, the mRNA is translated

BACTERIOFERRITIN

So what turns the anti-bfr gene on and off?

- A global regulatory protein known as **Fur (Ferric uptake regulator)** detects and binds iron. When plenty of iron is present, **Fur** acts as a **repressor** and **turns off the transcription** of a dozen or more operons needed for adapting the cell to **low iron concentration**.

- In particular, when there is a change from low to high of iron concentration, Fur + iron (activated).

- **Fur + iron turns off** the **anti-bfr gene (repress anti-bfr gene)**. This **turns on** the production of bacterioferritin.

Fe content	"On" *bfr* gene	"On" anti *bfr*	Interaction of antisense Bfr with mRNA *bfr*	Bfr Protein synthesized	Repression by Fur against anti *bfr* gene
high	+	-	-	+	+
low	+	+	+	-	-

- So by using antisense RNA we can regulate one gene the opposite way to a group of others. Artificially synthesized antisense RNA will interfere with gene expression, or anything else involving RNA.

- Antisense RNA is being tested experimentally to suppress cancer by stopping chromosome division.

Bacterial Democracy - Quorum Sensing

- Its not just the cells of higher organisms that get together. Although bacteria live as single cells, under some circumstances they need to cooperate in communal ventures.

- Amazingly enough they regulate certain genes by a form of chemical voting known as **quorum sensing**. The basic idea is quite simple. Because bacteria are so tiny, if only a few are present they will be **unable** to make much impact.

- On the other hand, if billions are crowded together, their **joint effort** may be quite **significant**. So the bacteria involved all secrete a **signal molecule** called **autoinducer**, into the medium.

- If the level of **autoinducer** is high enough, this means that enough bacteria are present to have some effect and everybody switches on the genes for communal effort.

- The best known example is light emission by sea-faring bacteria. A single bacteria cannot make enough light to be seen, and only if billions cooperate it is worthwhile to get involved in light emission.

Vibrio fischeri and Luminescence

- *Vibrio fischeri* is a marine relative of our tiny friend *Escherichia coli*. Just as *E. coli* lives in the guts of animals, *Vibrio fischeri* wishes to live inside fishes.

- If a dense enough crowd of *Vibrio fischeri* gathers on organic matter at the bottom of the sea, they all turn on their lights together and the glow attracts a fish which swallows them.

- Some more daring luminous bacteria provide light for monsters like giant deep sea squid.

Lux Operon

- The enzyme **luciferase** is responsible for biological light emission and is widely used for detecting gene expression.

- Luciferase and some accessory proteins are coded for by the ***lux*** genes. The signal molecule, or **autoinducer**, is made by **LuxI protein** and binds to the **LuxR** protein.

- When **LuxR has bound autoinducer** it **switches ON the genes for luciferase**.

Autoinducer

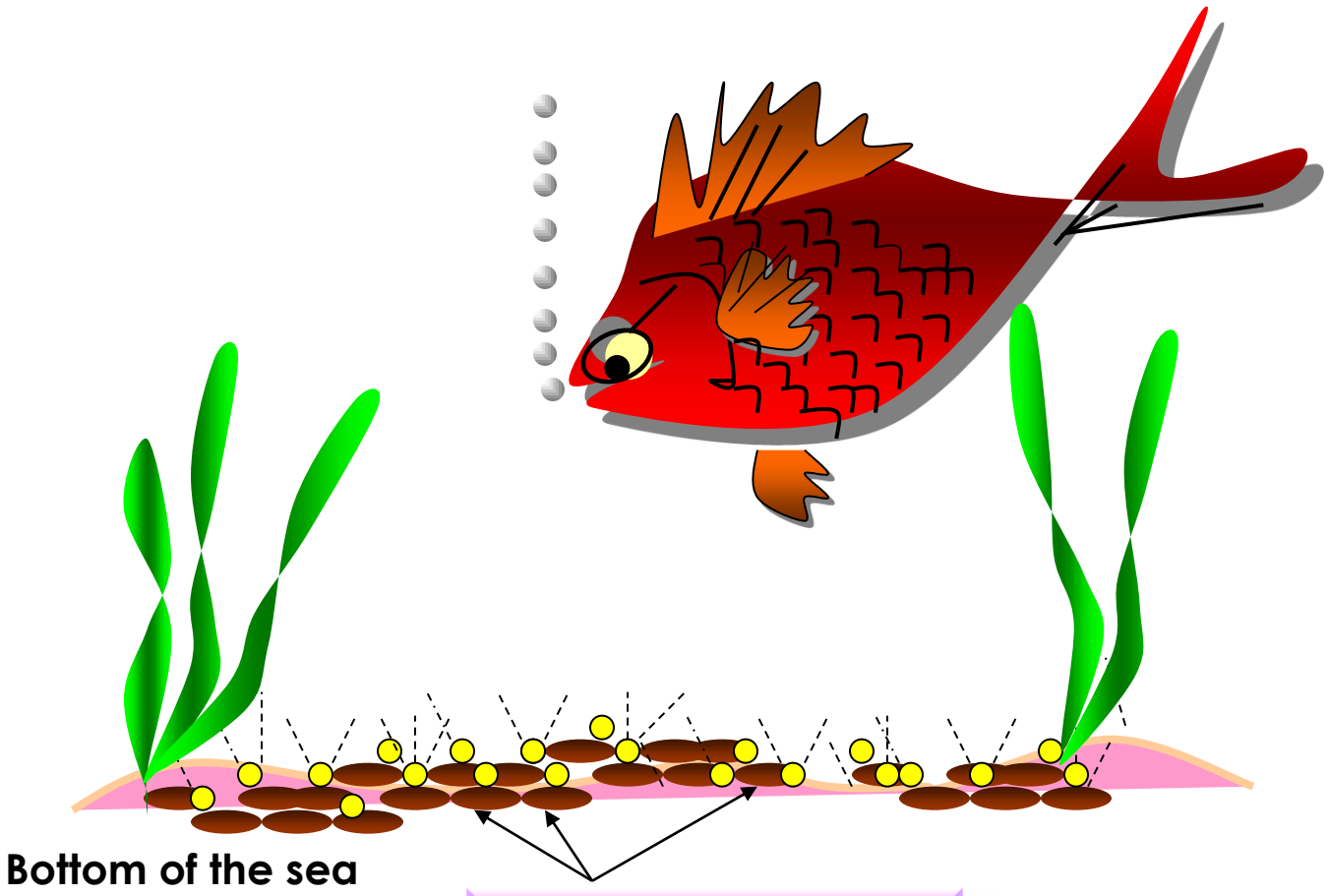
- The key property of autoinducer is that it drifts freely into and out of the deep blue sea.
- If lots of cells are huddled together (*berkumpul*), autoinducer from one cell will wander into others instead of being lost.
- Only if the population density rises above l0 million bacteria per milliliter does enough autoinducer build-up to **turn on the genes for luciferase**.

- Another, less enlightened, relative of *E. coli*, ***Erwinia carotovora***, lives by eating vegetables.

- Plant cells are many times the size of bacteria and have very thick walls. So breaking these down requires a **cooperative effort**.

- When enough bacteria are present on the plant, they all **secrete digestive enzymes** in unison under control of a quorum sensing system similar to **LuxI/LuxR**.

Light emission by seafaring bacteria



References:

- Madigan, M.T., Martinko, J.M., Dunlap, P.V. and Clark, D.P. (2009). Brock Biology of Microorganisms: Pearson Education, USA.

- Clark, D.P. and Russel, L.D. (2000). Molecular Biology Made Simple and Fun: Cache River Press, USA.