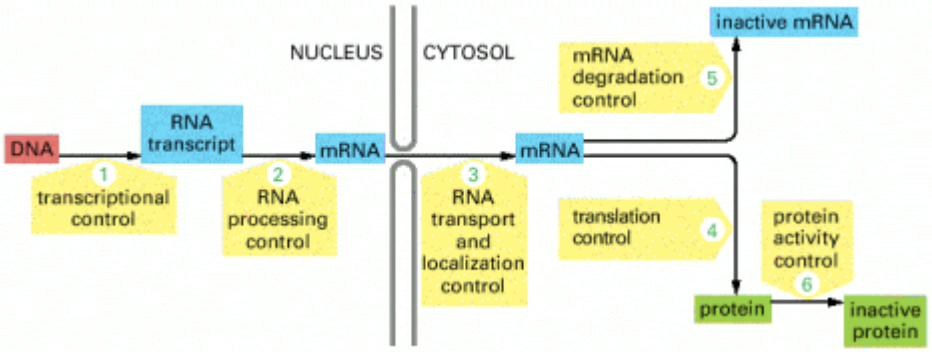
Transcriptional Regulation

This next figure basically presents all the steps involved in gene expression. There are six steps at which eukaryotic gene expression can be controlled. (Figure 7-5)



As we know already, in each cell there are over 3 billion base pairs per human cell, and if we consider this number logically, how can we find particular genes in which need to be transcribed? If we remember how the chromatin is organized, we will remember the presence of histones and the presence of sections of DNA in which we can find sequence-specific DNA-binding proteins. Histone code can demarcate regions for expression. This is controlled by the sequence-specific DNA-binding proteins, which are specific transcription factors that "read" the DNA sequence and binds to specific sequence motifs.

If we look closer at the structure of the double helix, we can see two different types of grooves, known as the minor groove and the major groove. The existence of these grooves is what allows proteins to "read" the DNA helix from the outside.

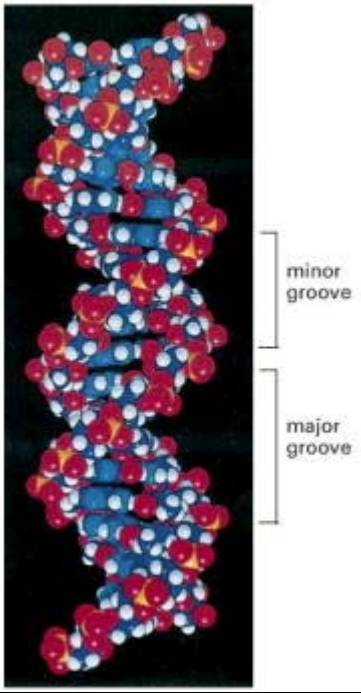
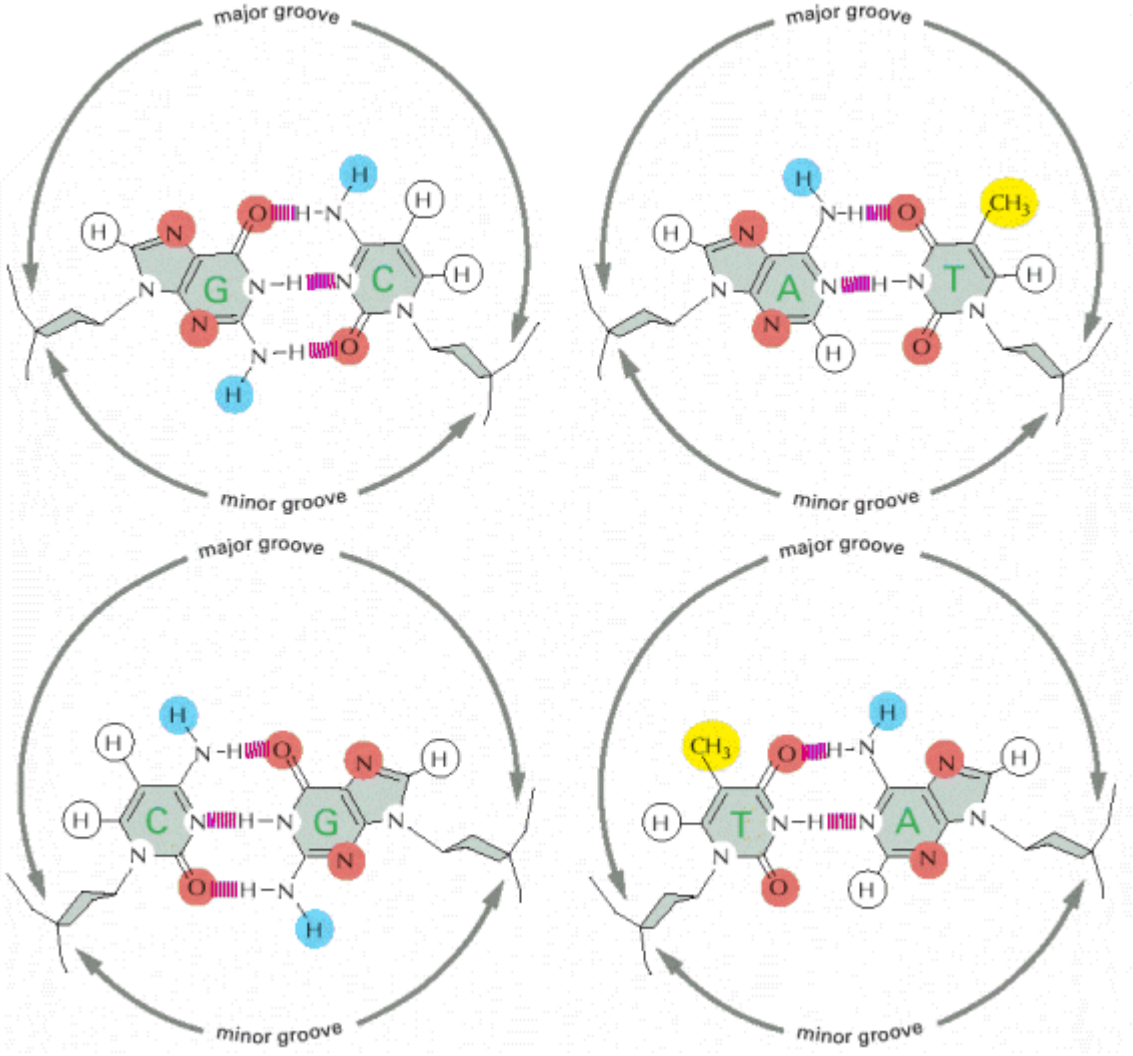


Figure 7-6. Double-helical structure of DNA. The major and minor grooves on the outside of the double helix are indicated. The atoms are colored as follows: carbon, *dark blue;* nitrogen, *light blue;* hydrogen, *white;* oxygen, *red;* phosphorus, *yellow.*

In the major groove, the bases can be discriminated by certain proteins. Now the figure below illustrates the minor and major grooves from a different perspective.



The "reading" and binding by DNA-binding proteins to target DNA sequences involves specific amino acids of proteins. Usually several amino acids help to locate and bind to the DNA. For an optimal fit, protein binding to DNA often induces contortions or bending of the DNA, which allows more surface area of DNA to interact with the protein.

So, in order to understand how these processes work, we must start from the first step, which is to identify important DNA sequences. To understand where these sequences are, we will go through an experimental example which analyzes the promoter (switch), identifying important motifs using reporter genes (lamp). The reporter gene will be able to tell us when the switch is on/off.

Reporter genes: - Proxy for the endogenous gene (genes of the organism)

Reports on timing and localization of expression driven by a specific promoter.

Preferably non-invasive & non-destructive

- Several common reporters

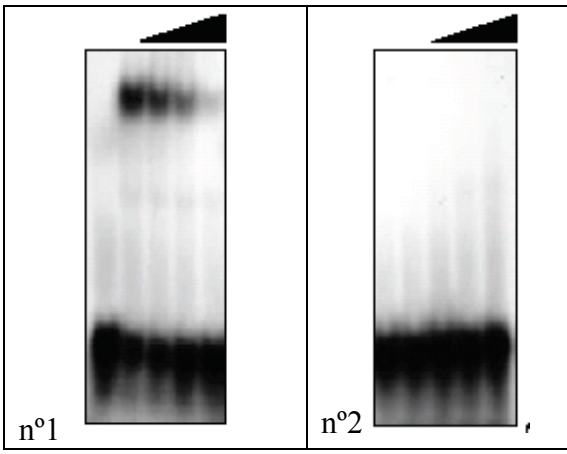
Green fluorescent protein (GFP)

Luciferase (Luc) – red light

B-Glucoronidase (GUS) – blue precipitate.

The method in which this is done is simple. On the sequence that you are investigating, you must separate different sections. In each sample, one of the sections is removed. With the help of the reporter we can recognize which sample has it's gene turned off. This is called the promoter deletion analysis. Once the sample is found, we know where the specific sequence that is necessary is found. The concept to know here is that this loss of function shows necessity. Now if we use this necessary sequence in combination with the minimal promoter and a reporter, and if this gives us a signal that it is turned on, we see that this gain of function shows sufficiency. So we conclude that this particular region is also sufficient for promoter activity. Together, the experiments show that this sequence is NECESSARY & SUFFICIENT for promoter activity.

Step two is to characterize protein interactions with these DNA sequences. The way this can be done is using EMSA (Electrophoretic Mobility Shift Assay). In order to perform this, we must first radioactively label the target sequence in which we are trying to investigate. The method in which we are going to investigate is a gel electrophoresis (PAGE), where the sequence will travel down from the cathode (negative) to the anode (positive). Two rows are set up, the first one containing only the sequence and the other one containing the target sequence plus the binding protein. What happens here is that the row where the sequence is free, the sample will travel further in the gel. The binding protein + sequence sample will be retarded, therefore traveling a smaller distance in the same amount of time.



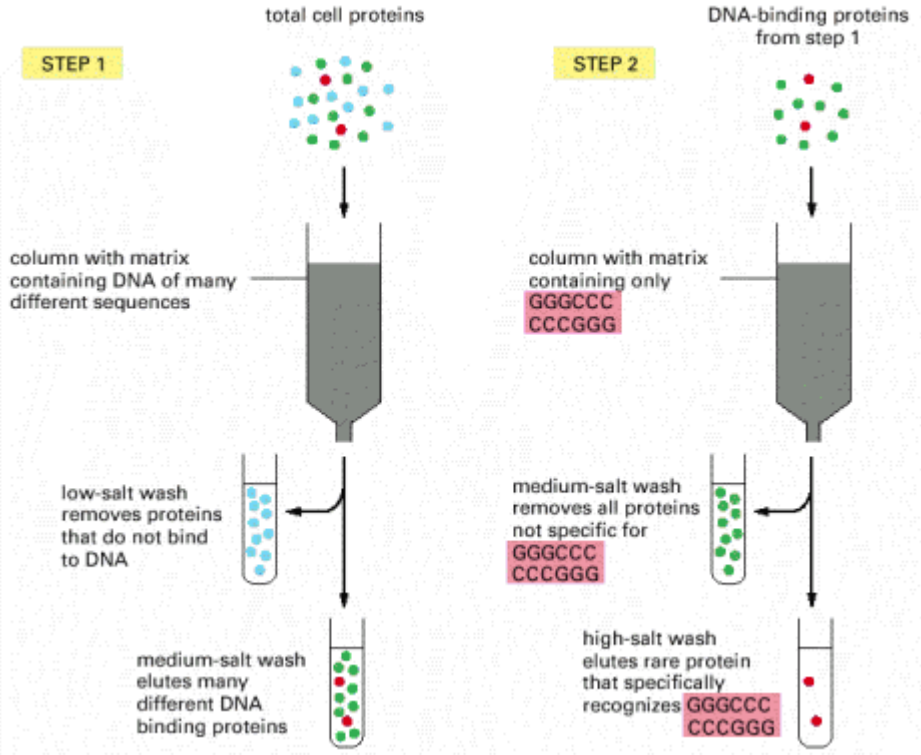
Now in order to check for specificity, we can use this same EMSA but now we also add in the different columns a competitor unlabeled sequence, illustrated in nº1. In this example we use a different amount of competitor sequences in each column, going from low in the left to high in the right. Notice that as we increase the amount of competitor sequence, we decrease the amount of labeled sequences bound to the protein. Now on nº2, a mutated unlabelled sequence was added, and it is visible that this didn't have a big effect on the sample.

Now using the EMSA we can also find different binding proteins for furthers examination. For example, if we add the labeled target sequence and in a different column we add the sequence plus binding proteins, we can separate the different proteins that bind to our sequence, since they will have different mobility in the gel. After running the gel and finding the different bands that correspond to the different binding proteins we need to purify them. In this case we will use the Affinity Chromatography technique.

In this figure below we can see that in the first step of Affinity Chromatography, all the proteins that can bind to the DNA are separated from the remainder of the proteins on a column containing a huge number of different DNA sequences. Most of the sequence-specific DNA-binding proteins have a weak affinity for bulk DNA, therefore they are retained on the column. This affinity is mostly due to a ionic attraction, so the proteins can be washed off using a solution that contains a moderate concentration of salt. Now that we have the DNA-binding proteins from the first step we can move to the second part of the

process. The mixture of DNA-binding proteins is passed through a column that contains only DNA of a particular sequence. Generally, all the DNA-binding proteins will stick to the column, mostly due to nonspecific interactions. These are again washed with salt solutions, which leave the column with only the proteins that bind specifically to the particular DNA sequence, due to stronger attraction of the two.

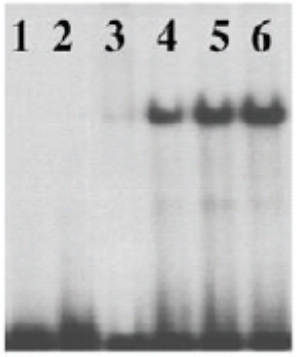
In the second step, the mixture of DNA-binding proteins is passed through a column that contains only DNA of a particular sequence. Typically, all the DNA-binding proteins will stick to the column, the great majority by nonspecific interactions. These are again eluted by solutions of moderate salt concentration, leaving on the column only those proteins (typically one or only a few) that bind specifically and therefore very tightly to the particular DNA sequence. These remaining proteins can be eluted from the column by solutions containing a very high concentration of salt. Now in order to remove these from the column we use a high-salt wash, to elute the desired proteins.



Now with these experiments we have identified and characterized these components of "genetic switches". For further characterization we could now do a detailed biochemical analyses (binding site selection assay, DNA foot printing) or do a detailed structural analyses (X-ray crystallography, NMR spectroscopy). For more information refer to chapter 8 of the textbook.

As mentioned, one of the ways we can further investigate the specificity of DNA binding of proteins, we can use Binding Site Selection Array. Where we do something similar to EMSA, we use a radioactive labeled pool of random oligonucleotides (which are simply short sequences of nucleotides). On one column of the gel we place the nucleotides, and on a different column we place the nucleotides + proteins. After running the gel we find the band where the protein was bound to the sequence, and we purify this shifted band. According to Dr. Campbell, this purification is done by simply cutting it out of the gel

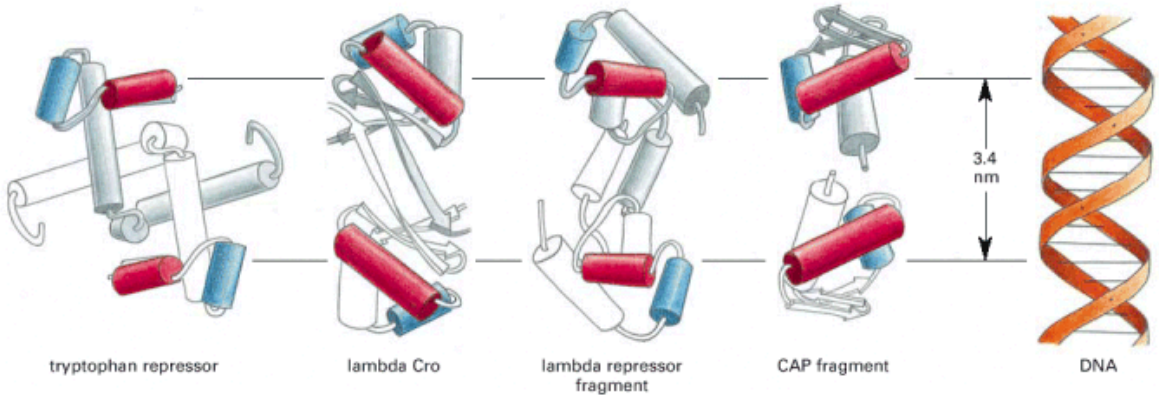
using either scissors or a scalpel (more common). The DNA is then eluted from the gel and used as template in the next PCR reaction. The EMSA step is done again with the purified protein, and this step is repeated multiple times, enrichment occurs in each round of selection. Then the final shifted band is cloned in plasmid vector. In the picture below it is possible to observe the increase in strength of the shifted band as you go through each round of selection. We also know that the sequences at the end of the selection process share a common consensus sequence that the protein binds.



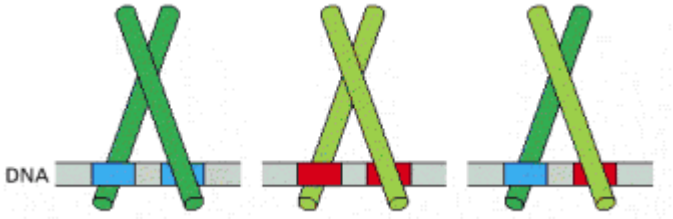
1. Free probe
2. First selection
3. Second selection
4. Third selection
5. Fourth selection
6. Fifth selection

Now, what are the components of these genetic switches? Short DNA sequences are fundamental components of genetic switches, which are recognized by gene regulatory proteins. These sequences typically are smaller than 20 nucleotides in length. Single nucleotide variation in the binding site is not uncommon, but generally at least 3 invariant nucleotides are present. This is what provides the specificity in recognition. Each sequence of length n, has a $1:4^n$ chance of occurring at random in a DNA sequence. An example of this is TATA, which occurs by chance every 4^4 (=256) nucleotides.

Gene regulatory proteins contain motifs that "read" DNA: The Helix-Turn-Helix Motif Is One of the Simplest and Most Common DNA-binding Motifs. It is found in both eukaryotes and prokaryotes. It is composed of two α helices connected by a short extended chain of amino acids, which constitutes the "turn". The helices are held at a fixed angle through interactions between the two. The more C-terminal helix is called recognition helix (shown in red) because it fits into the major groove of the DNA; its amino acids play an important part in recognizing the specific DNA sequence to which the protein binds. The polypeptide chain outside the HTH domain also makes important contact with the DNA, helping to fine-tune the interaction. The bigger picture below demonstrates a feature of many sequence-specific DNA-binding proteins. They tend to bind as dimmers to DNA sequences that are composed of two very similar "half-sites", which are also arranged symmetrically, as shown in the smaller picture. This arrangement allows each protein monomer to make a nearly identical set of contacts and enormously increase the binding affinity, which actually **squares** the affinity constant, doubling the number of contacts which doubles the free energy of the interaction.

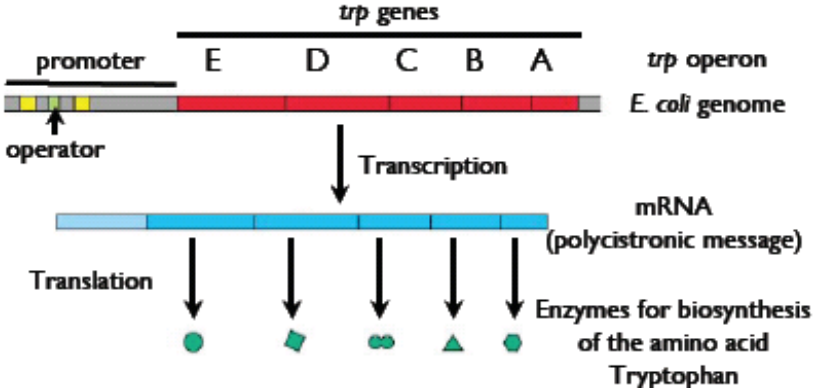


Homeodomain proteins are one special class of HTH proteins. They are composed of almost identical stretch of 60 amino acids. Another common motif are the Zinc finger motifs. Zinc fingers are co-ordinated by histidine and/or cysteine residues. The reading the and binding by DNA-binding proteins to target DNA sequences involves specific amino acids of the protein. Gene regulatory motifs tend to work together and the specificity in binding is dependent on motif interactions. For example the two leucine zippers that form a dimmer and bind to DNA. Now there are homodimers and heterodimers, which have different specificity as shown in the picture below. The leucine zippers from the left, *left and center*, are homodimers, which simply bind to symmetric DNA sequences. Now the one in the *right* is an illustration of the combination of the two monomers which forms a heterodimer, that can recognize a hybrid DNA sequence, composed from one *red* and one *blue* region. It is also possible for these DNA-binding domains, such as HTH and homeodomains, to be covalently joined by polypeptides, which can also increase affinity.

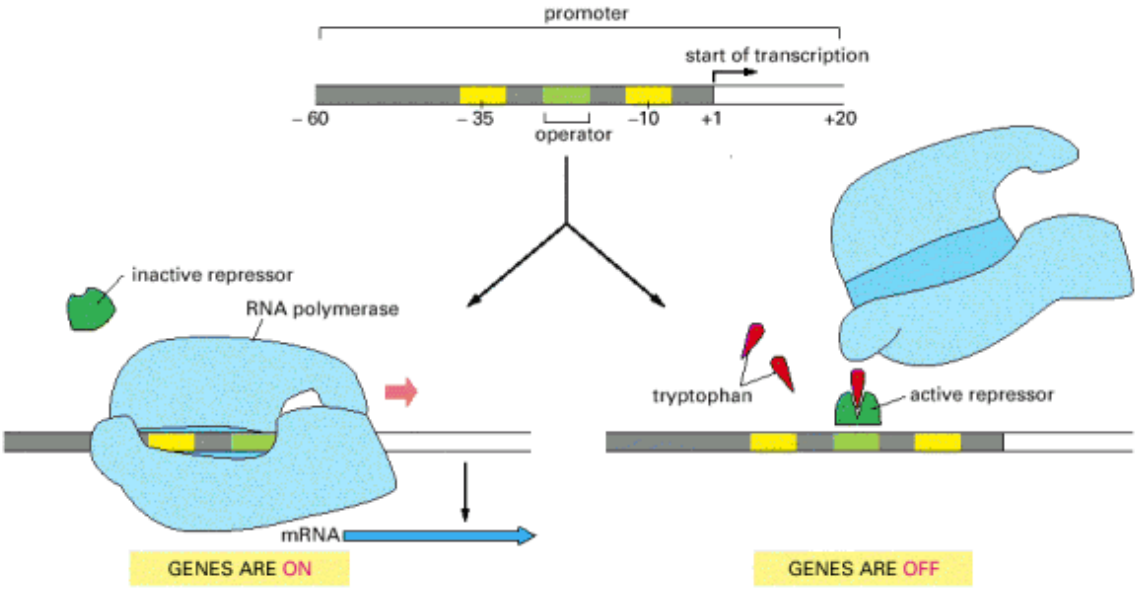


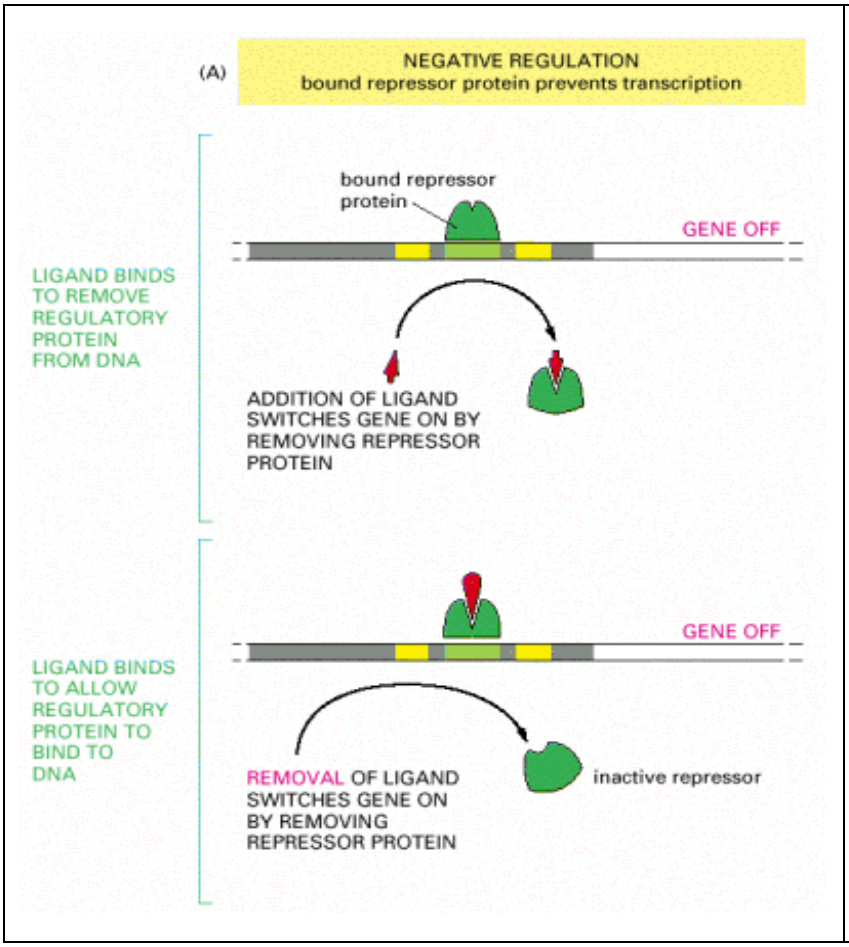
Now that we have a better understanding of the different components of a genetic switch we can actually look at an example, the *trp* repressor and the *trp* operon.

Operon: in a bacterial chromosome, it is a group of continuous genes transcribed into a single RNA molecule under the control of a single promoter.

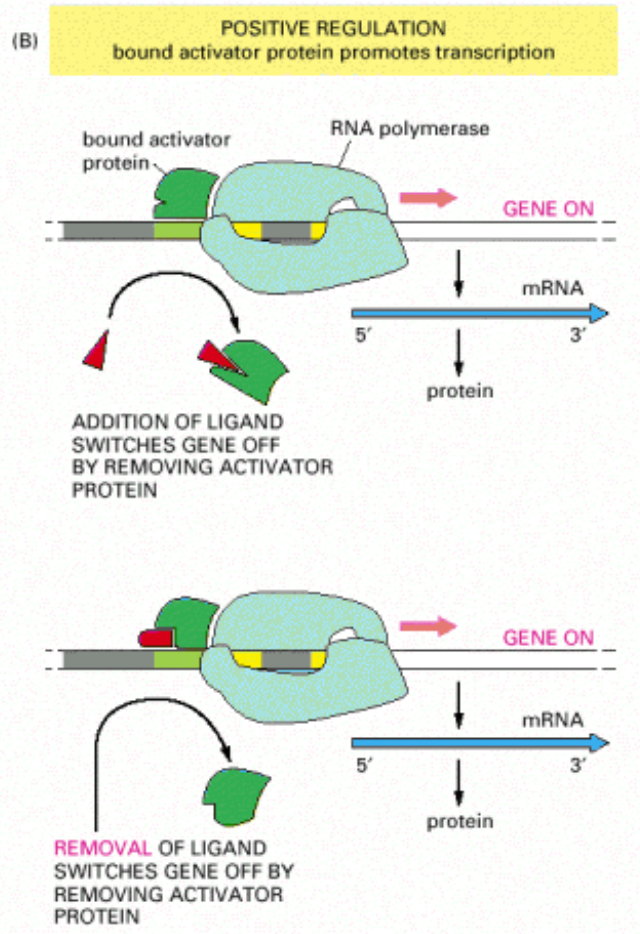


When the concentration of Tryptophan [Trp] are low the *trp* repressor is NOT bound to the operator site, therefore the transcription of the mRNA responsible for the production of the enzymes which produce Trp is allowed. However when [Trp] is high, the *trp* repressor is BOUND to the operator site, shutting down transcription. Ligand-mediated repression/de-repression is a common theme in bacterial genetic switches.



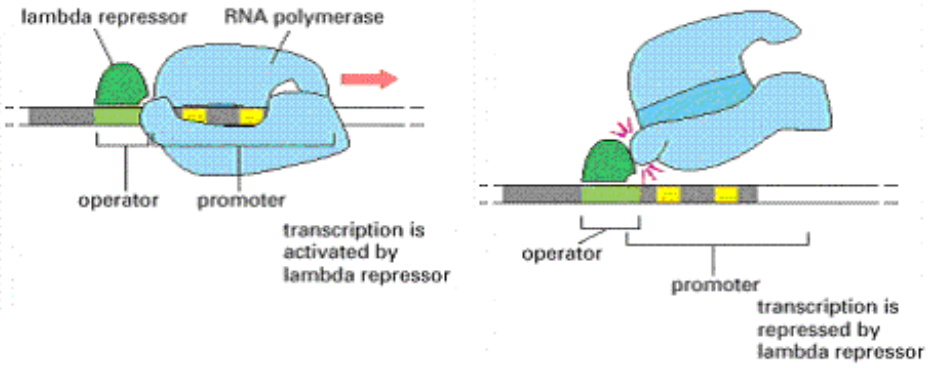


Negative regulation is when the bound repressor protein prevents transcription from happening. In this case the de-repressor (inducer) ligand binds to remove regulatory protein from DNA. Ligand switches gene ON by removing repressor – de-repression. The co-repressor is the ligand that binds to allow regulatory protein to bind to DNA. Ligand removal switches gene ON by removing repressor – co-repression.



During **Positive regulation**, the bound activator protein promotes transcription. The inactivator is a ligand that binds to remove regulatory protein from DNA. Ligand switches gene OFF by removing activator (de-activation). The inducer or co-activator, binds to allow regulatory protein to bind to DNA. Ligand removal switches gene OFF by removing activator (co-activation)

There are also examples of proteins that function as activators and repressors depending on the context. The example used in lecture was the Lambda repressor, which is a gene regulatory protein encoded by *lambda* phage genome. In some genes the lambda repressor protein acts as an activator, depending on the distance between the operator and the promoter, which in this case allows the lambda repressor to provide favorable contact with the polymerase. In other genes, the presence of the lambda repressor prevents RNA polymerase binding, since the distances are not favorable.



The *lac operon* is another more complex genetic switch. One of the components of the *lac* operon is the β-galactosidase which hydrolyses lactose. Another component is the lac permease, which transports lactose into bacterial cells.

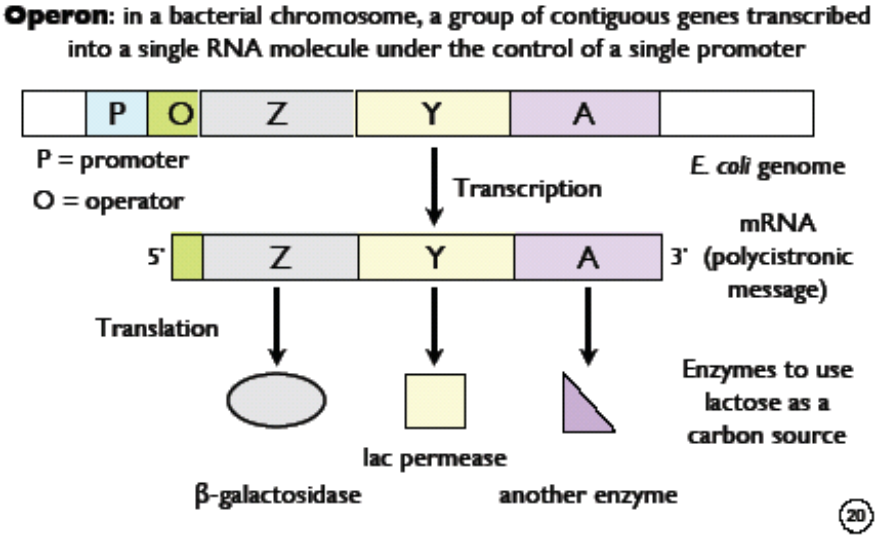
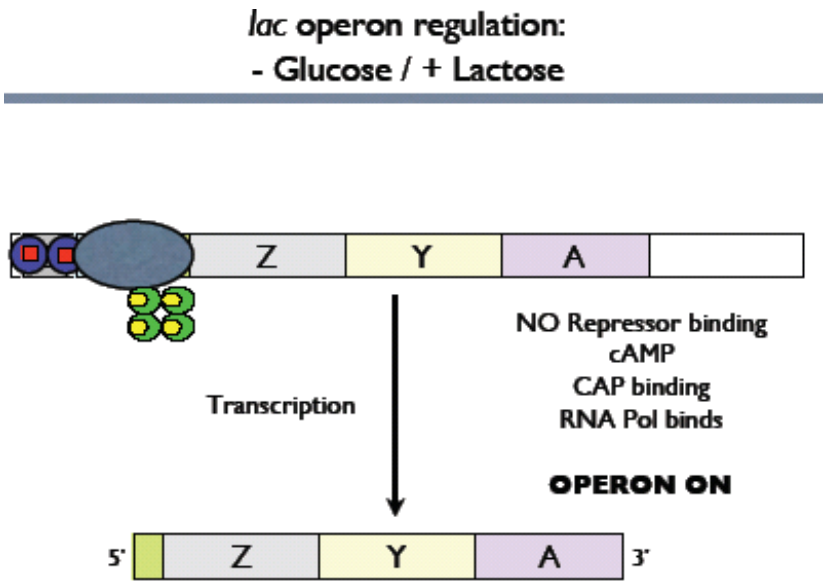


Fig1

From the picture below and the caption what I think is most important to know is that for optimal transcription you need lots of Lactose and little Glucose. Also have a general idea of why. Lac repressor binds to Operator. But if in presence of lactose the lac repressor is bound to allolactose, and not to the gene. Now understand what CAP does (which is a Helix-turn-helix domain). So when you have a lot of glucose, the levels of cAMP in the cell are low, since cAMP is synthesized in absence of glucose, in catabolism. This matters because without cAMP, CAP CANNOT bind to the CAP-binding site. Therefore if there is little glucose, CAP can bind and this facilitates the binding of the RNA polymerase to the promoter. Now if you have the case where there is Glucose and Lactose, the repressor will not be bound to the DNA, which is good, but the CAP won't be bound either so the operon will be turned OFF.



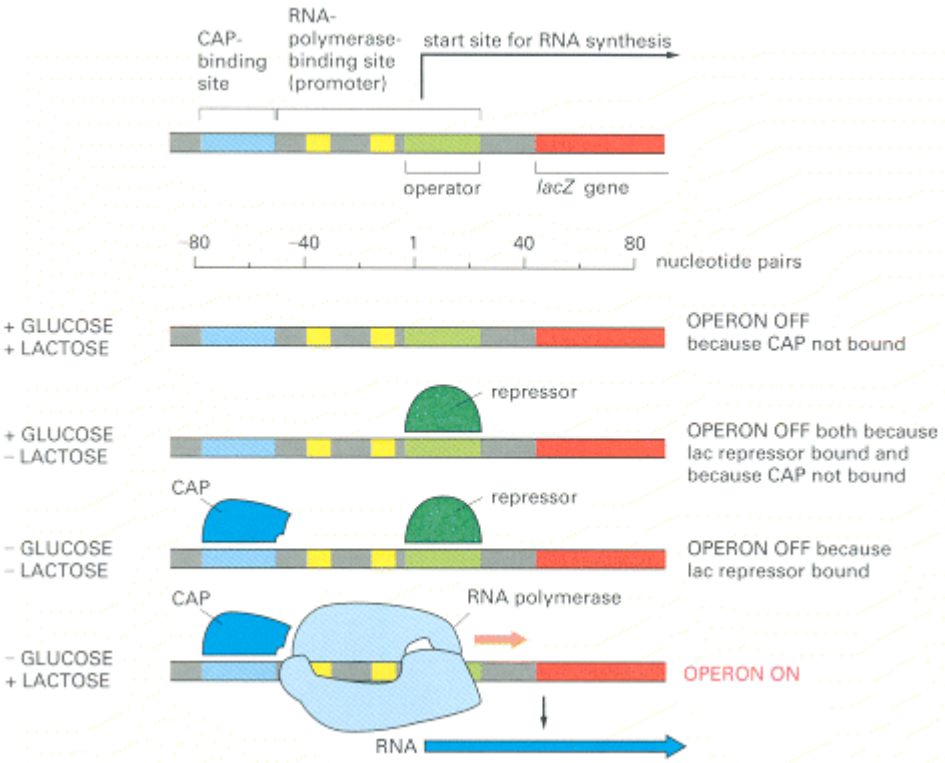


Fig7-39

Figure 7-38. Dual control of the *lac* operon. Glucose and lactose levels control the initiation of transcription of the *lac* operon through their effects on the lac repressor protein and CAP (catabolite activator protein). Lactose addition increases the concentration of allolactose, which binds to the repressor protein and removes it from the DNA. Glucose addition decreases the concentration of cAMP; because cAMP no longer binds to CAP, this gene activator protein dissociates from the DNA, turning off the operon. As shown in Figure 7-11, CAP is known to induce a bend in the DNA when it binds; for simplicity, the bend is not shown here. *LacZ*, the first gene of the *lac* operon, encodes the enzyme β-galactosidase, which breaks down the disaccharide lactose to galactose and glucose. The essential features of the *lac* operon are summarized in the figure, but in reality the situation is more complex. For one thing, there are several *lac* repressor binding sites located at different positions along the DNA. Although the one illustrated exerts the greatest effect, the others are required for full repression. In addition, expression of the *lac* operon is never completely shut down. A small amount of the enzyme β-galactosidase is required to convert lactose to allolactose thereby permitting the *lac* repressor to be inactivated when lactose is added to the growth medium.

Promoter – Nucleotide sequence in DNA to which RNA polymerase binds to begin transcription.

Operator – Short region of DNA in a bacterial chromosome that controls the transcrioption of an adjacent gene.

Eukariotic Transcription regulation

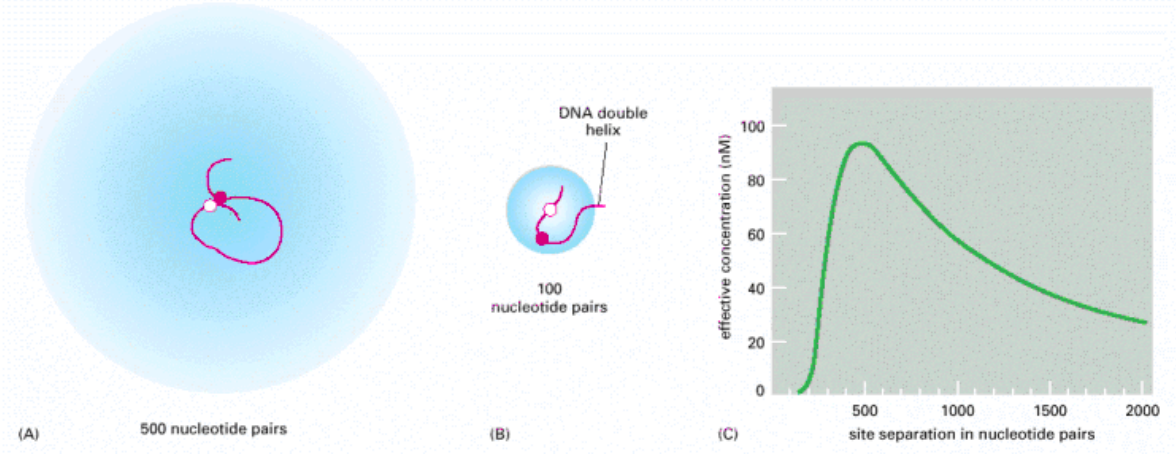
We can expect that the regulation of transcription in prokaryotes is much simpler than in eukaryotes. That's right. Prokaryotes have simple development but they must respond to fluctuation in the environment they live in. Eukaryotes have more complex development and they also must be able to respond to fluctuation in the environment. So we can expect a more complex regulation of transcription in Eukariotes.

There are 3 important differences between eukaryotic and bacterial transcriptional regulation:

1. Eukaryotic DNA transcription regulations are more complex, providing opportunities for regulation not available to bacteria
2. Eukaryotic RNA Polymerase II (which makes mRNA) cannot initiate transcription on it's own, it requires a set of general transcription factors, that provides multiple steps for changing rate of transcription.
3. Eukaryotic gene regulatory proteinsare packaged into chromatin, thousands of nucleotides away from the promoter, therefore many possible regulatory sequences for a given gene.

Histone code demarcates regions for expression; sequence-specific DNA-binding proteins control this expression. Specific transcription factors "read" the DNA sequence and bind to specific sequence motifs. If we recall how RNA Polymerase II requires an assembly of factors we can visualize this better. There are a number of rate limiting steps in eukaryotic transcription, for example: Enzymatic reactions catalyzed by components of the basal transcriptional machinery (eg. Phosphorylation of Pol II by TFIIH); Availability of general transcription factors: Recruitment of general transcription factors.

An important concept to know is that eukaryotic gene regulatory proteins can control gene expression from a distance. The binding of two proteins to different sites on DNA increases the probability of interaction between the proteins. On the picture below we can see this, the intensity of blue cloud indicates likelihood of collision. Frequency of collision increased in tethered by DNA- down to 100bp after which collision is restricted to bending limitations on DNA. Note that the optimal distance is 500 nucleotides, to increase the probability. Multiples of 10bp important, why? Because this is the distance in which you come back to the same side of the helix. The gene activation from a distance is very common in eukaryotes, and it can occur in bacteria, but it is rare.



Minimal Promoter - The region necessary for binding at the TATA box???. Important for transcription to take place, but it is not sufficient, since you also need an enhancer.

In transcription you need the binding of the general transcription factors, RNA polymerase, mediator, chromatin remodeling complexes and histone acetylases. Also know that a typical eukaryotic gene has many activator proteins, which together determine the rate and pattern of transcription. Sometimes this happens from a distance as we just saw, these gene regulatory proteins help the assembly or the parts. In addition, activators attract the ATP-dependent chromatin remodeling complexes and histone acetylases.

This gene regulatory sequences can reside upstream and downstream of coding sequence, it can even be "intragenic" (introns). These sequences are binding sites for gene regulatory proteins.

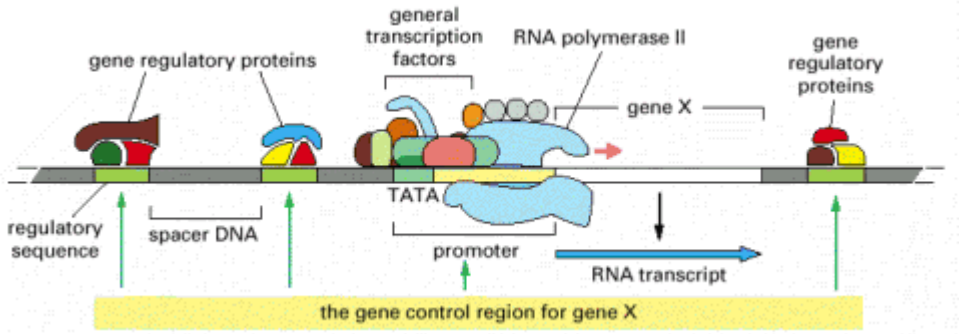
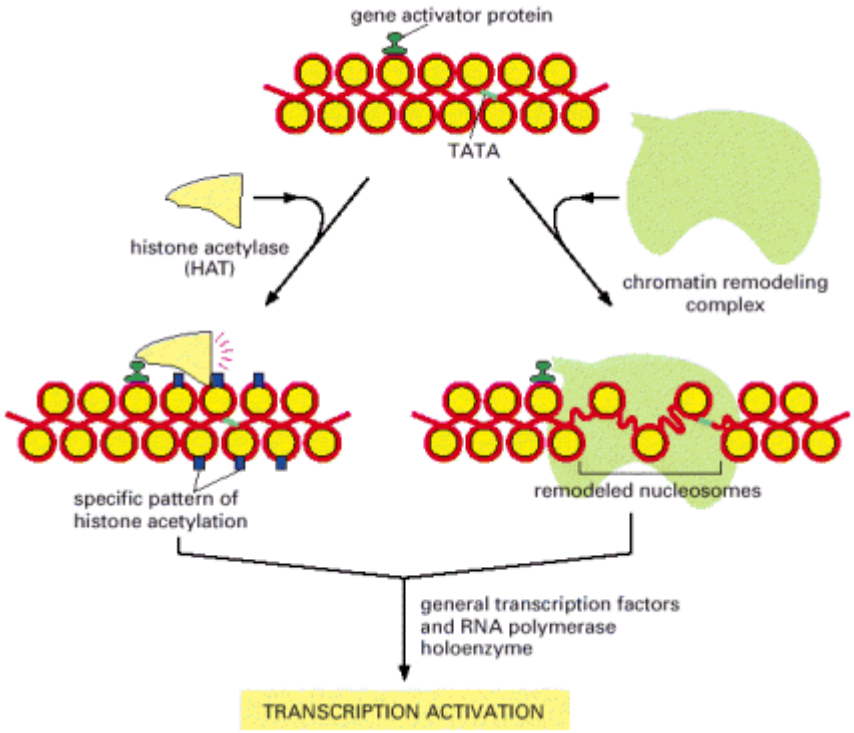


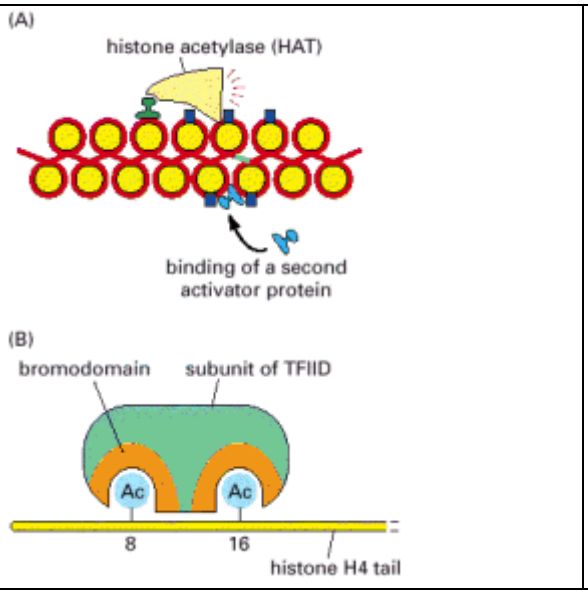
Figure 7-41. The gene control region of a typical eucaryotic gene. The *promoter* is the DNA sequence where the general transcription factors and the polymerase assemble (see Figure 6-16). The *regulatory sequences* serve as binding sites for gene regulatory proteins, whose presence on the DNA affects the rate of transcription initiation. These sequences can be located adjacent to the promoter, far upstream of it, or even within introns or downstream of the gene. DNA looping is thought to allow gene regulatory proteins bound at any of these positions to interact with the proteins that assemble at the promoter. Whereas the general transcription factors that assemble at the promoter are similar for all polymerase II transcribed genes, the gene regulatory proteins and the locations of their binding sites relative to the promoter are different for each gene.

Eukaryotic gene activator proteins tend to be modular (they can sweep around?) in structure: DNA binding (DBD) and transcriptional activation (TAD) domains. Whatever this means. Increas rate of transcription up to 1000 fold is possible, weak and non-specific interactions with RNA polymerase holoenzyme (enzyme produce from many sub-units). What this means is that the activator interacts with specific proteins in such way that it helps to assemble some of the transcription factors, increasing the speed of transcription. This recruitments of the transcriptional machinery is via interaction between TAD and mediator protein complex (ca. 20 proteins). Translational fusion experiment, where TAD is replaced by mediator proves this idea, apparently.

Now recall: Eukaryotic sequence-specific gene activator proteins can increase chromatin accessibility. This is done by the recruit of histone acetylase, decondense chromatin, facilitate recruitment of the transcriptional machinery. Remember that when we studied Chromatin we learned that sometimes the chromatin has to be remodeled (page 4/8 of the notes on DNA and chromosomes). Figure below show how it's done.



If we can also recall: Histones can be acetylated on their accessible N-terminal "tails". This acetylation occurs on Lysine (K) residues. Acetylation makes the histones more neutrally charged, which tends to repel DNA, making it more accessible. I think this is pretty important.



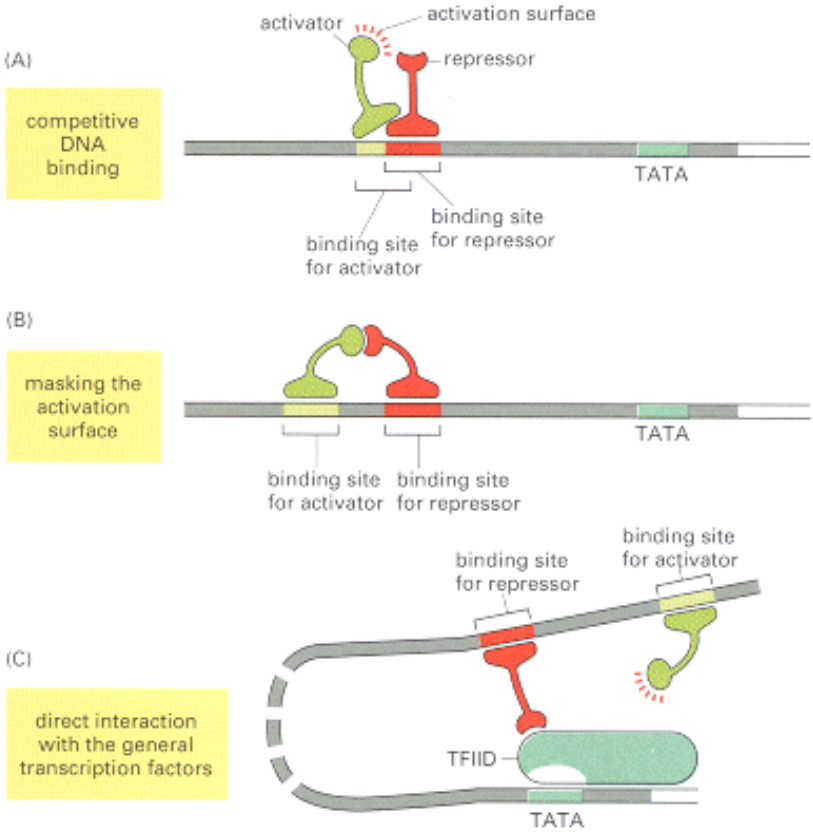
Now recall:

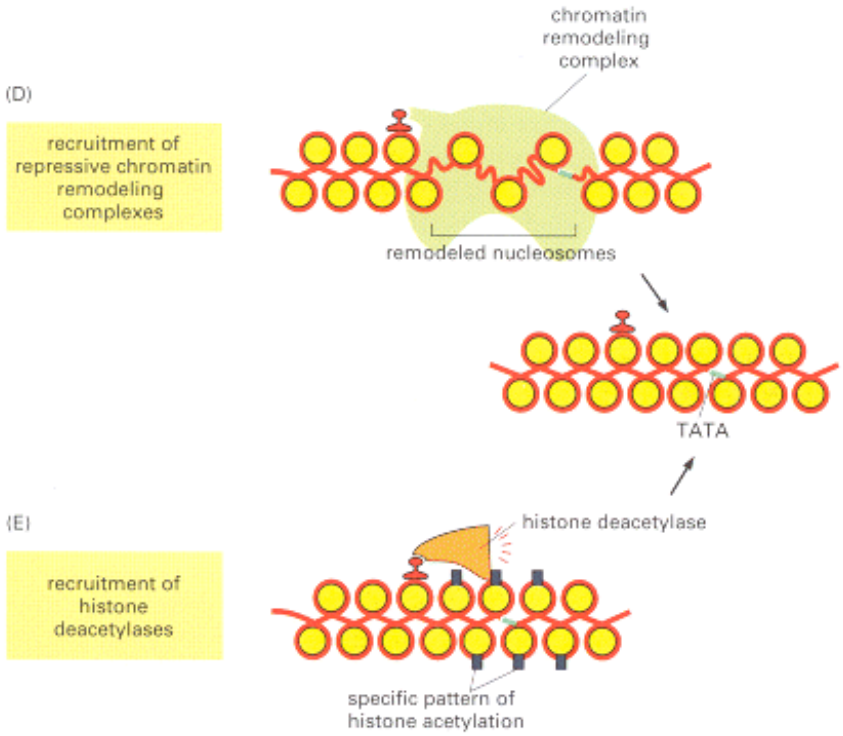
- Histone acetylation is done by gene activator proteins facilitates recruitment of other transcription factors.
- Histone acetylation by gene activator proteins facilitates recruitment of general transcriptional machinery (e.g. TFIID)

We also have to recall that: Gene activator proteins facilitate chromatin remodeling. Gene activator proteins facilitate histone modification. Gene activator proteins facilitate chromatin remodeling assembly of RNA Polymerase II Pre-initiation complex at Promoter.

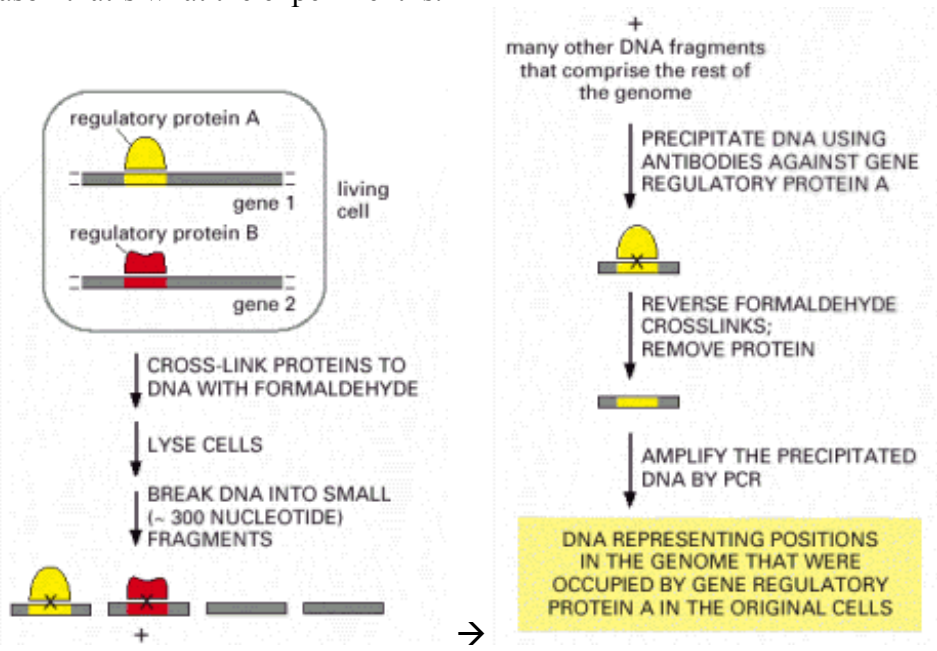
Now that we remember all this, we can understand it a bit better hopefully. So eukaryotic gene activator proteins tend to function synergistically (synergy = The interaction of two or more agents or forces so that their combined effect is greater than the sum of their individual effects). Experimental evidence shows that recruitment of more transcription factors results in a greater than additive effect on transcription, which shows transcriptional synergy. There is plenty of evidence for this *in vivo,* multiple binding sites in one gene regulatory region.

There are also Eukaryotic Gene Repressor Proteins, which can mediate their effect through multiple mechanisms. Refer to picture below. Also know that depending on the components you can get proteins functioning as activators or repressors, even a single protein can do both.



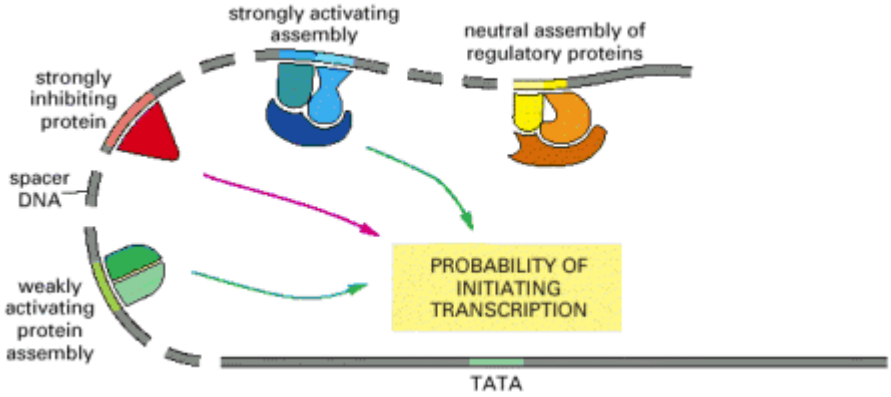


Interactions between transcription factors and chromatin can be examined; the example used in class was the Chromatin Immunoprecipitation (ChiP). Basically you do all this in the picture just to find which DNA sequences the regulatory proteins bind, and for some reason that's what the experiment is.



Also know that Eukaryotic Gene Regulatory Proteins often work in complexes (combinatorial action). Different complexes are formed by different proteins, and one protein can be used in one complex that activates transcription or it might also be used to repress transcription. And then to make it more fun, you may have several complexes

working together in different parts of the DNA, some may be inhibitory and some may be activating. They don't work like switches anymore they are more like dimmers.



These differences in transcription factor use define cells, tissues and organisms. Also know that a large number of genes are dedicated to regulation of transcription. About 5% of the total number of genes is used in regulation of transcription. In mammals it is between 5-10%. These gene regulatory proteins can also be regulated (regulator is regulated)

LECTURE 6

*Establishing transcription networks – Genetic network in *Arabidapsis* light perception.

In general eukaryotes have complex developtment and must respond to fluctuations in environment. A good example is the phytochrome, which is a photoreceptor that plants use to detect light:

- Soluble pigmented proteins of ca. 125kD.
- Homodimer
- Each subunit contains covalently linked linear tetrapyrole chromophore.
- Photointerconversion between two stable isomers.
- One absorbs red light (Pr), one absorbs far red (Pfr)
- Pfr thought to be the active form
- Encoded by multiple genes (e.g. 5 in arabidopsis, PhyA-PhyE)

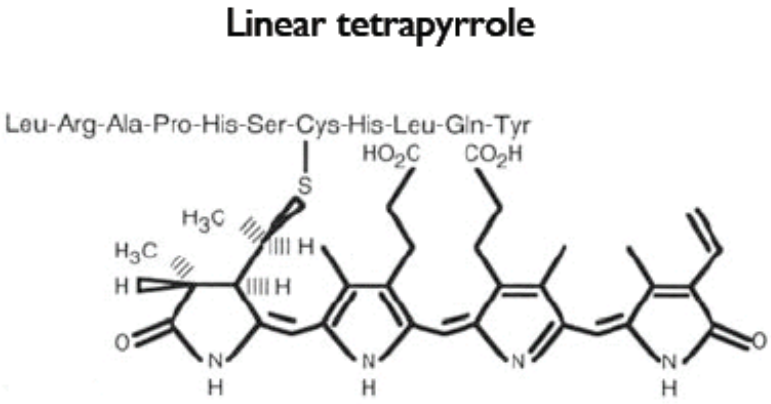
-Photomorphogenesis is light-mediated development. The photomorphogenesis of plants is often studied by using tightly-frequency-controlled light sources to grow the plants. Then **Skotomorphogenesis** is the opposite.

-Chromophore is a term that describes the moiety of a molecule responsible for its color. These consist of regions in a molecule where the spacing between two different molecular orbitals falls within the range of the visible spectrum. That is to say, light hits the chromophore and can be absorbed by exciting an electron from its ground state into an excited state. When a molecule absorbs certain visible wavelengths in this way, but transmits or reflects others, the molecule has a color.

There is a Thioether linkage between chrmophore and cysteine in apoprotein. A picture of the Phytochrome Chromophore is shown below. So as we have already seen this phytochrome is found in plants in order to respond to light, that's why when you leave a plant in the dark they develop a certain way, and if they are in the light they develop differently. So basically this phytochrome changes conformation when light shines in it, it goes from Pr (inactive) form to Pfr (active) form. This phytochrome photointerconversion affects plant development.

Pr → red light → Pfr → far-red light → Pr

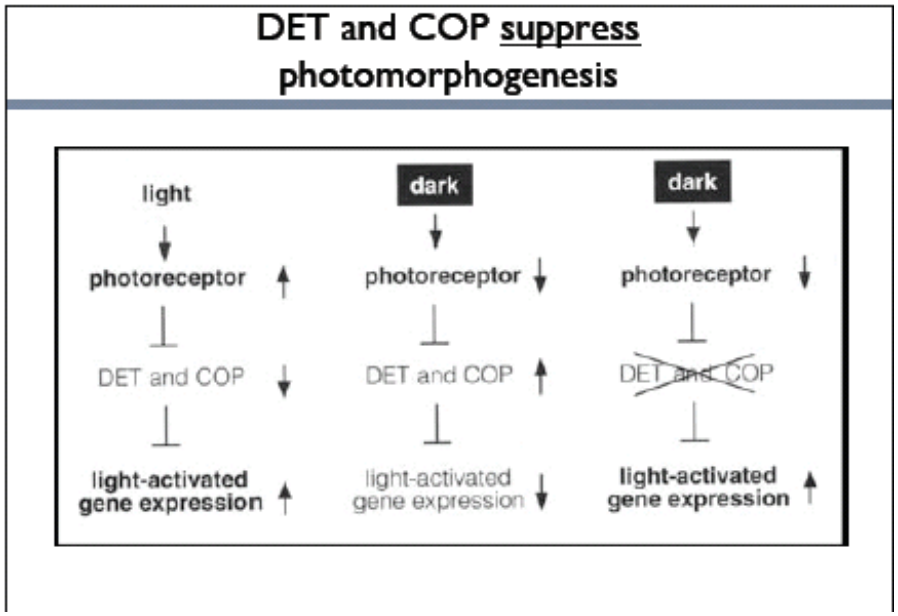
Now the question is, how is phytochrome perception of light transformed into a response? As you can probably already predict, as red light shines in the Pr, we get the Pfr form, which is active. This creates a signal transduction which affects the transcriptional network, altering gene expression. So then you get photo-morphogenesis. Most of what is known about the developmental responses affected by phytochrome is derived from the study of *Arabidopsis thaliana* mutants.



This example of the phytochrome pathway has hallmarks of genetic networks that are found in many eukaryotes: Importance of nuclear import, "master control genes" – encode transcriptional regulators.

Phytochromes appear to directly activate gene expression. Just on these slides there were diagrams and I think what he wants us to know is that in PhyA the phytochrome is activating the gene directly from the nucleous, and in PhyB the regulator is moving from the cytoplasm. PhyA and PhyB act on some of the same genes, but they each act on a set of specific genes. For examlple PhyA acts on FAR1 and PhyB doesn't.

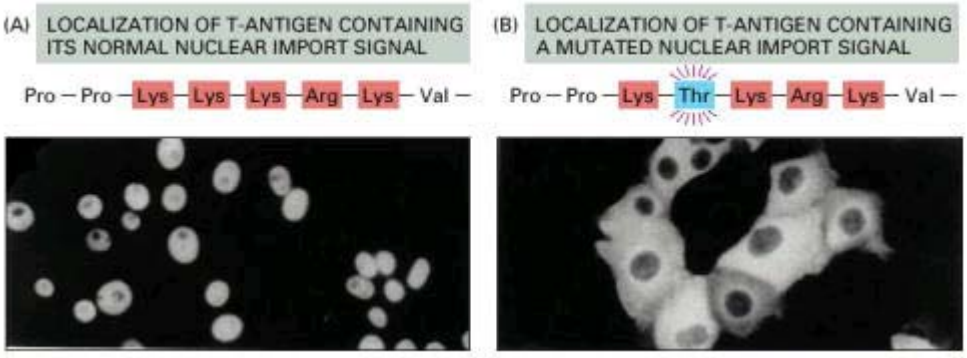
I don't actually know what this means in slide 14, but in darkness GFP:PHY fusion in cytoplasm and GFP:PHY fusion moves to nucleus with red light. Also I'm not sure if we have to know this next slide, which says DET and COP suppress photomorphogenesis. Which I guess is not that bad to remember, because photomorphogenesis happens in presence of light, so if there is light there will be less DET and COP. After the picture we go more in detail on this COP.



COP1 is a zinc finger protein that functions as a gene regulatory protein. It has a nuclear localization signal; but, this signal appears to function only in the dark. In the dark the protein is in the nucleus where it functions as a repressor of genes that are light-activated genes involved in photomorphogenesis.

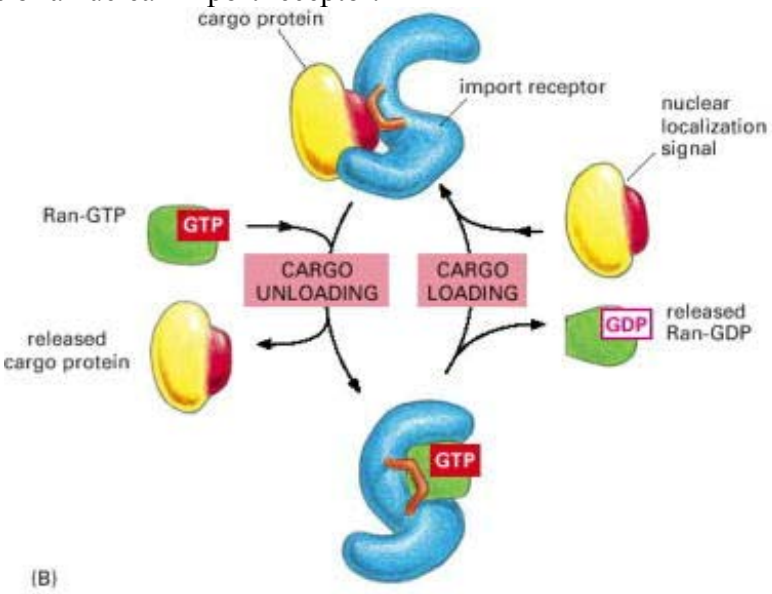
*Establishing transcription networks – Nuclear import of transcription factors

Now he wants us to know that the nucleus is a double membrane-bound organelle. Also the nucleus has a size exclusion limit, above which active transport is required. Then this next picture says on the title: Nuclear Localisation Signals Direct Nuclear Proteins to the Nucleus. What the picture is showing is that this signaling molecule is usually found in the nucleus of the cell, you can see that on A. Then on B when you mutate the protein, the signal can't get into the nucleus so you can see it on the surroundings but not so much inside the nucleus.



This next picture is showing us nuclear import receptor proteins and Ran GTPase. The picture should illustrate how the binding of the Ran-GTP might cause nuclear import

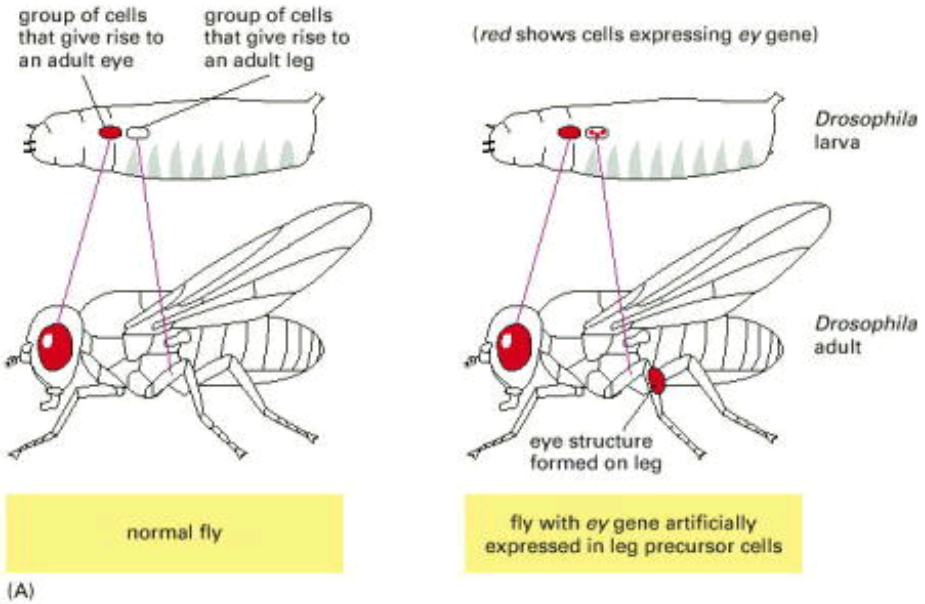
receptors to release their cargo. The cycle of loading in the cytosol and unloading in the nucleus of a nuclear import receptor.

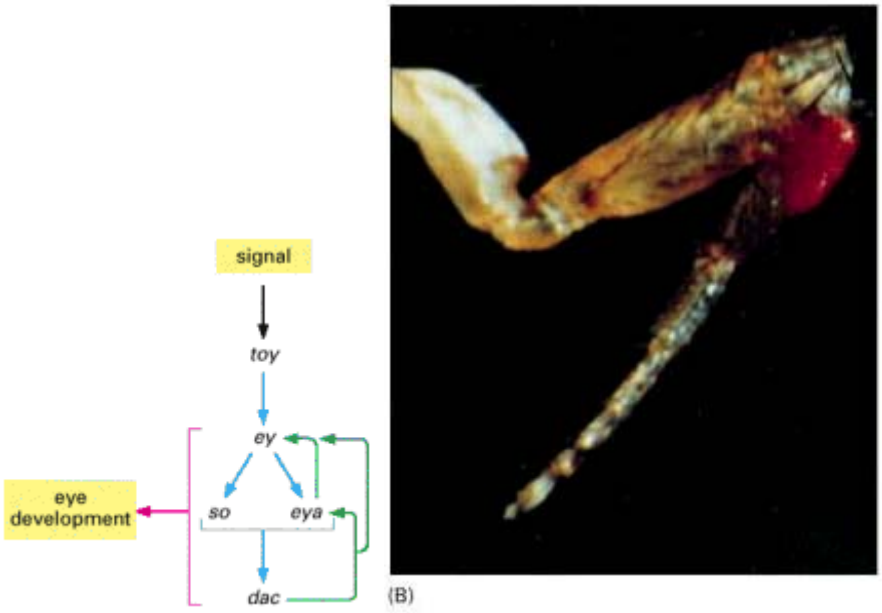


Apparently control of nuclear import is a common theme in transcriptional regulation. Another example of this is NF-AT (nuclear factor of activated T-Cells). Don't need to know the details but if you want to see the picture of an example here it is. http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=mboc4.figgrp.2171

*Establishing transcription networks – Master control genes

Master control genes frequently encode transcriptional regulators. Now he showed us the example with the eyeless (ey) & eye differentiation.





Master control genes, frequently encode for transcriptional regulators. “Canalisation” of cell differentiation by master control genes encoding gene regulatory proteins. Below there is a picture and its caption for this slide.

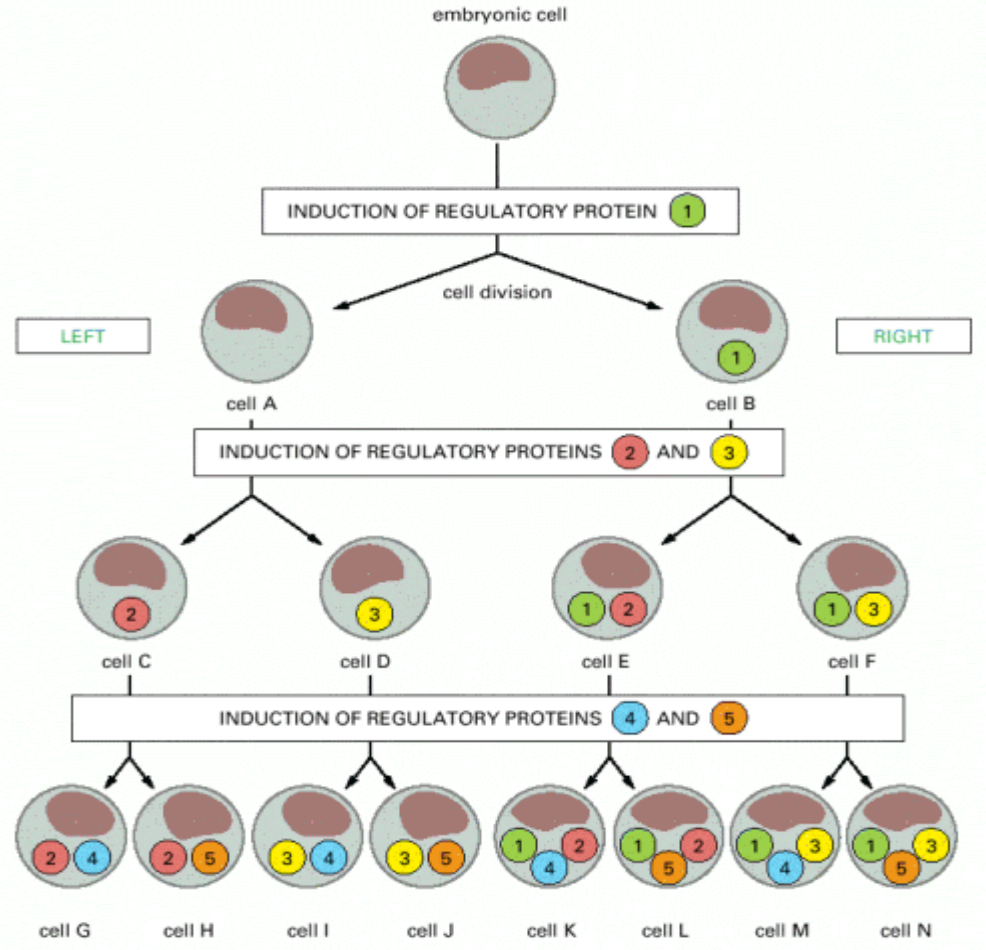
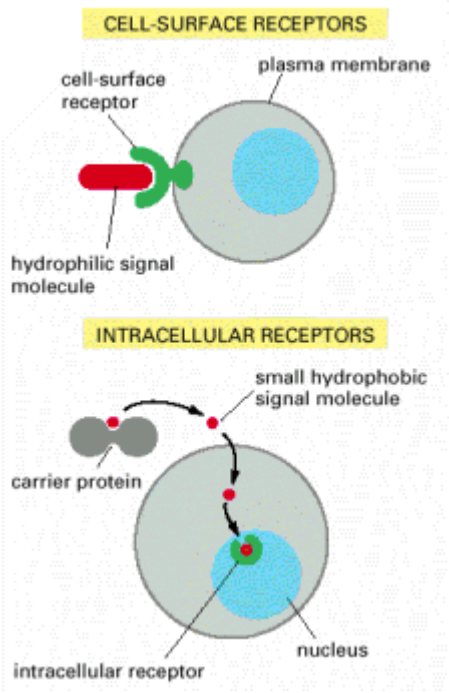


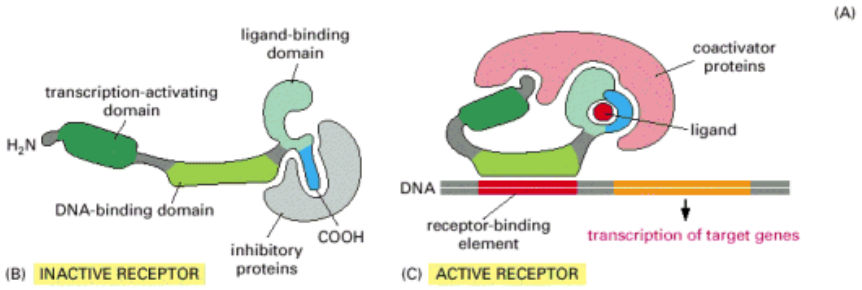
Figure 7-73. The importance of combinatorial gene control for development. Combinations of a few gene regulatory proteins can generate many cell types during development. In this simple, idealized scheme a "decision" to make one of a pair of different gene regulatory proteins (shown as numbered *circles)* is made after each cell division. Sensing its relative position in the embryo, the daughter cell toward the *left side* of the embryo is always induced to synthesize the even-numbered protein of each pair, while the daughter cell toward the *right side* of the embryo is induced to synthesize the odd-numbered protein. The production of each gene regulatory protein is assumed to be self-perpetuating once it has become initiated (see Figure 7-68). In this way, through cell memory, the final combinatorial specification is built up step by step. In this purely hypothetical example, eight final cell types (G-N) have been created using five different gene regulatory proteins.

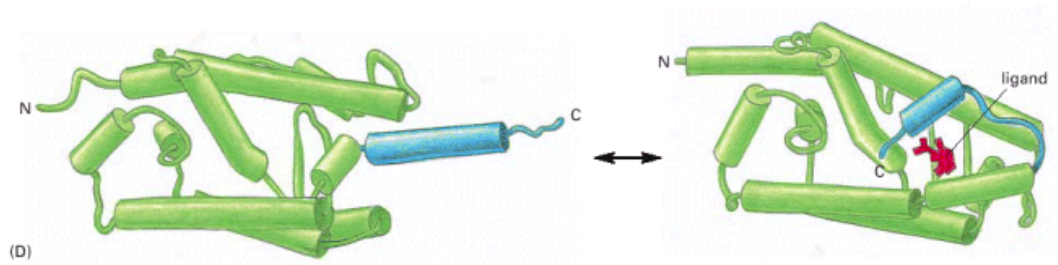
Now the next concept to know is: Terminal differentiation is frequently initiated by extra cellular signals. So basically if you have a stem cell it will only be able to specialize with the influence of it's surrounding, usually chemical signals, much of the time it is hormonal. There are also hydrophobic or hydrophilic extra cellular signals. Cell with a receptor for the signal = target cell.



*Transcriptional control and signal transduction – Nuclear hormone receptor pathways

Hormones are one type of extra cellular signaling molecule, and its target cells may be in many places in the body. There are hydrophobic hormones (cortisol from the adrenal gland, estradiol from the ovary, testosterone from the testis). On the next picture the blanks were the transcriptional activation domain, DNA binding domain and Ligand (hormone) binding domain.

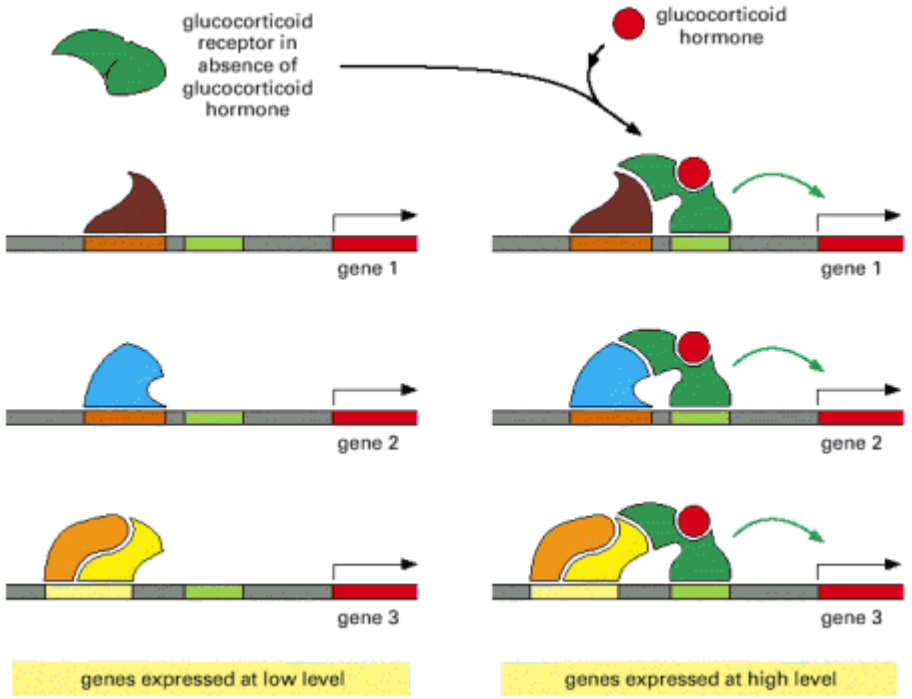


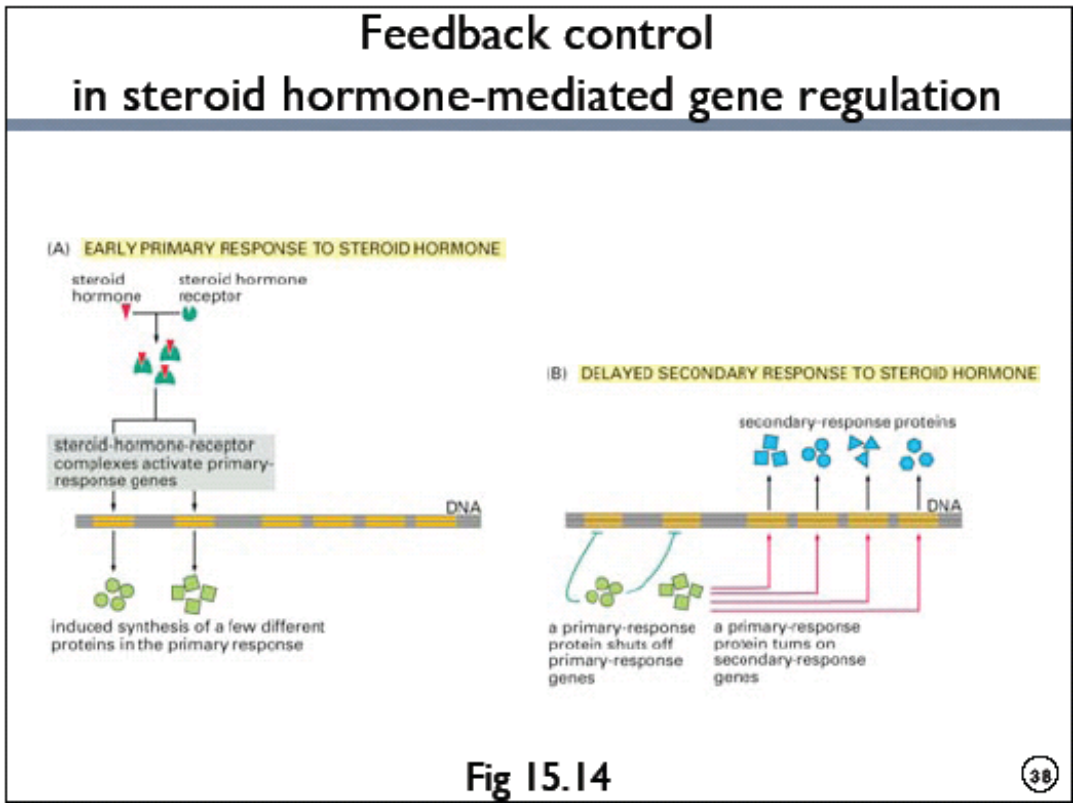


When bound to the ligand the receptor acts as a gene regulatory protein to alter the transcription of a subset of genes to give rise to a biological effect. Ligand binding induces a conformational change of the hormone receptor.

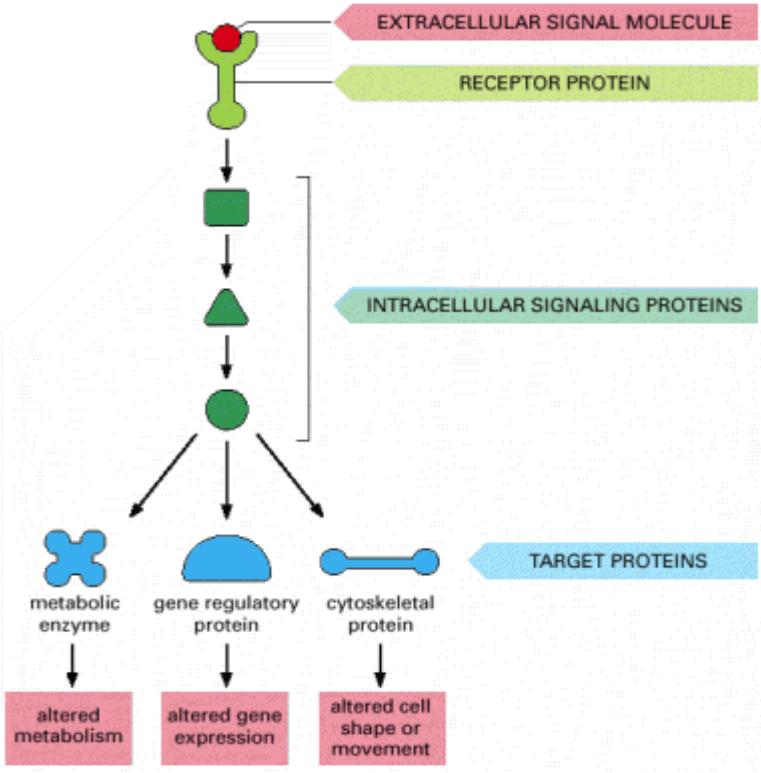
From the picture above just know that Co-activators can be histone acetylases, chromatin-remodelling complexes, proteins that bind to the RNA polymerase holoenzyme complex, which increases transcription.

Glucocorticoid receptor's DNA-binding domain has a zinc finger motif. Also know that several genes may be activated, response depends on the gene regulatory proteins present, regulatory region of the gene, and type of glucocorticoid receptor.





But cell signaling is generally more complex than this. Transcriptional regulation is a common theme in signal transduction pathways. More examples on the slides but I think its pointless.



LECTURE 7

On this lecture we looked at more steps involved in gene expression:
Transcriptional regulation
- Insulators
- DNA methylation

mRNA processing
- Intron Splicing
- Transcript termination

mRNA transport
mRNA localization
Translational regulation
mRNA degradation

So first we will deal with Transcriptional regulation via insulators.

Insulators prevent spread of Heterochromatin. Insulators directionally block action of enhancers. In the picture below enhancer ensures only the expression of gene B, I think that that's what he said in lecture, but the caption below explains it better. But basically insulators function to avoid formation of heterochromatin. So if for example you insert a new gene in a fly, if it is inserted near Heterochromatin it probably won't be expressed too much, but if you insert this gene and an insulator then you get expression. The other function is that it can block the action of enhancers, if insulator is between gene and enhancer.

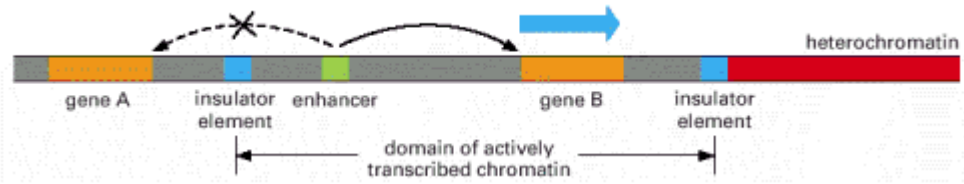


Figure 7-61. Schematic diagram summarizing the properties of insulators. Insulators both prevent the spread of heterochromatin *(right-hand side* of diagram) and directionally block the action of enhancers *(left-hand side).* Thus gene B is properly regulated and gene B's enhancer is prevented from influencing the transcription of gene A.

Heterochromatin – Region of a chromosome that remains unusually condensed chromatin; transcriptionally inactive during interphase.

Insulators thought to divide genome into 2 domains: one for Gene regulation, and one solely for chromatin structure. So we also have to know that Transcriptional regulation state of a gene can be passed onto daughter cells, this is a form of cell memory. So his example was a gene that encodes for a protein A that is a gene regulatory protein that activates its own transcription. So you can see the positive feedback here. This positive feedback is passed on to the daughter cells, so this gene continues to be expressed. In other cases, this memory can be perpetuated by cycles of co-operative binding by negative regulatory proteins. Here is an example of that:

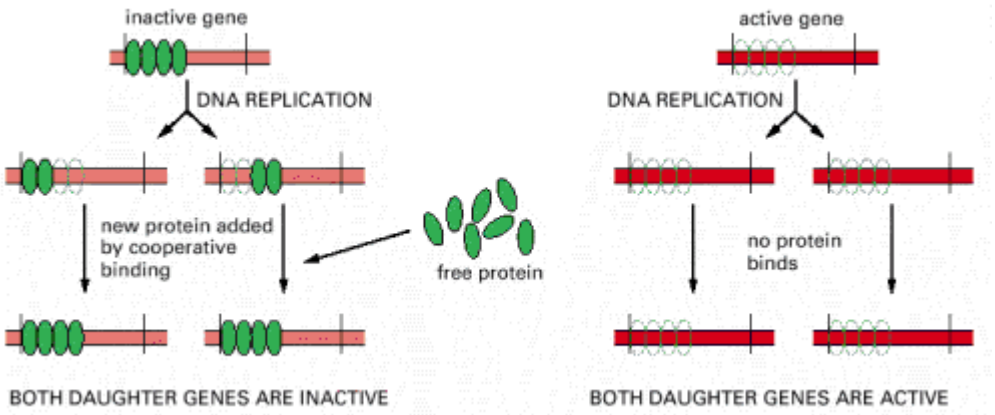


Figure 7-76. A general scheme that permits the direct inheritance of states of gene expression during DNA replication. In this hypothetical model, portions of a cooperatively bound cluster of chromosomal proteins are transferred directly from the parental DNA helix *(top left)* to both daughter helices. The inherited cluster then causes each of the daughter DNA helices to bind additional copies of the same proteins. Because the binding is cooperative, DNA synthesized from an identical parental DNA helix that lacks the bound proteins *(top right)* will remain free of them. If the bound proteins turn off gene transcription, then the inactive gene state will be directly inherited, as illustrated. If the bound proteins activate transcription, then the active gene state will be directly inherited (not shown). When the cooperative protein binding requires specific DNA sequences, these events will be limited to specific gene control regions; if the binding can be propagated all along the chromosome, however, it could account for the spreading effect associated with the heritable chromatin states discussed in Chapter 4. Although the proteins are depicted as being identical, the same principle can explain how cooperatively assembling combinations of different proteins can be propagated stably.

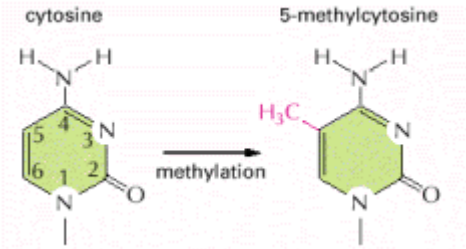
Transcriptional regulation state of whole chromosomes can be passed onto daughter cells, for example in the X chromosome inactivation. This is a dosage compensation to equalize the dosage of X chromosome gene products between males and females. Formation of Barr body (Condensed X chromosomes), this happens by random choice of X (maternal vs paternal copy).

This X chromosome inactivation involves:

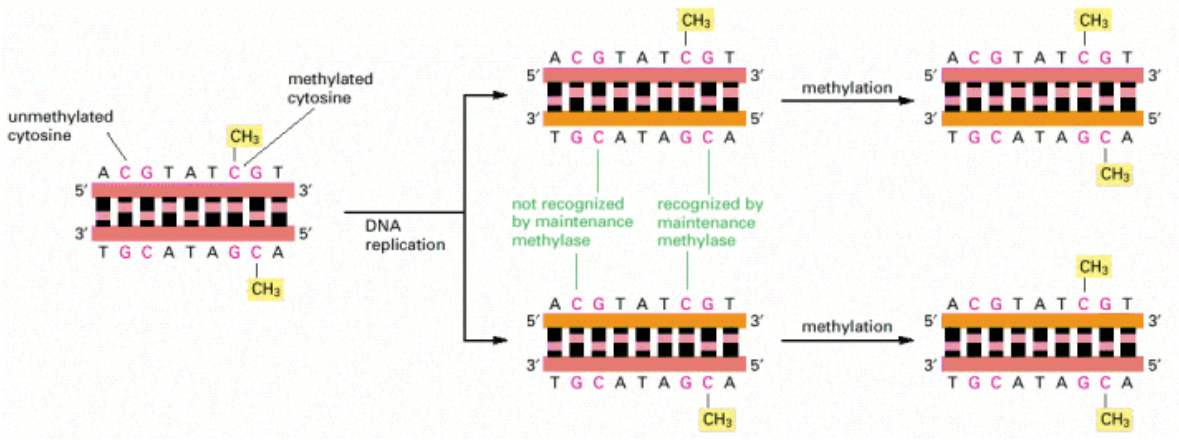
X-inactivation center (XIC) seeds and facilitates the spread of Heterochromatin

XIC encodes the non-translated RNA, XIST (X-inactivation specific transcript) RNA

XIST RNA spreads over chromosome, binding to the DNA, and induces Heterochromatin formation, which involves histone hypoacetylation & DNA methylation.

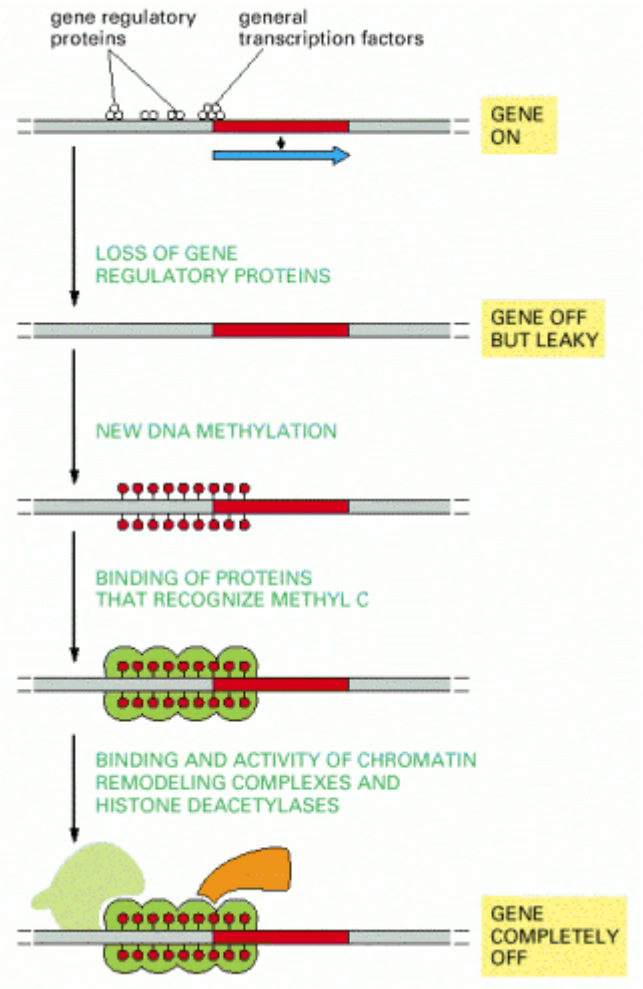


So lets understand this DNA methylation. In vertebrates **CYTOSINE** in CG dinucleotides can be methylated to give 5-methylcytosine. Just to be clear CG dinucleotides means that they are like C next door to G. Anyway, this is catalysed by DNA methyltransferase. Stably inherited in daughter cells (through activity of maintenance methyltransferase.



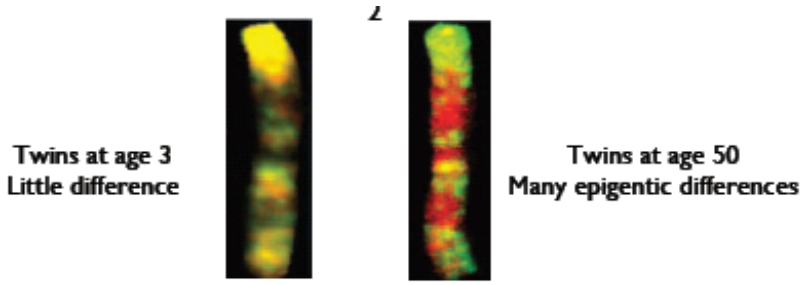
DNA methylation is thought to direct chromatin remodeling and gene silencing. DNA methylation Locks Genes in a silent state. Such changes in gene expression are stable throughout development – said to be epigenetic. Reset at meiosis and re-established post-implantation of embryo (by de novo methyltransferases). Prevents genes from being expressed at later stages of development. Suppress the movement of mobile genetic elements (transposons). It also prevents movements creating mutations.

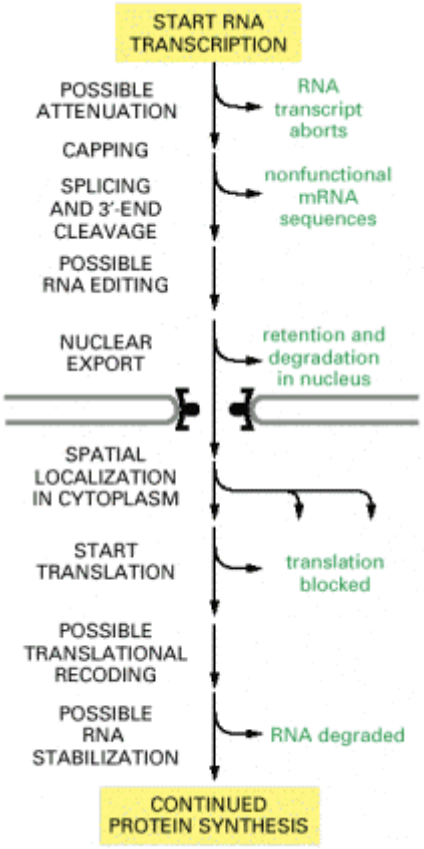
Figure 7-82. How DNA methylation may help turn off genes. The binding of gene regulatory proteins and the general transcription machinery near an active promoter may prevent DNA methylation by excluding *de novo* methylases. If most of these proteins dissociate from the DNA, however, as generally occurs when a cell no longer produces the required activator proteins, the DNA becomes methylated, which enables other proteins to bind, and these shut down the gene completely by further altering chromatin structure (see Figure 7-49).



The pattern of DNA methylation changes over the lifetime of individuals and varies between individuals. Comparison of methylation of chromosome I between zygotic twins shows this differentiation. Yellos indicates that methylation was equivalent between the pair of twins, Green indicates greater

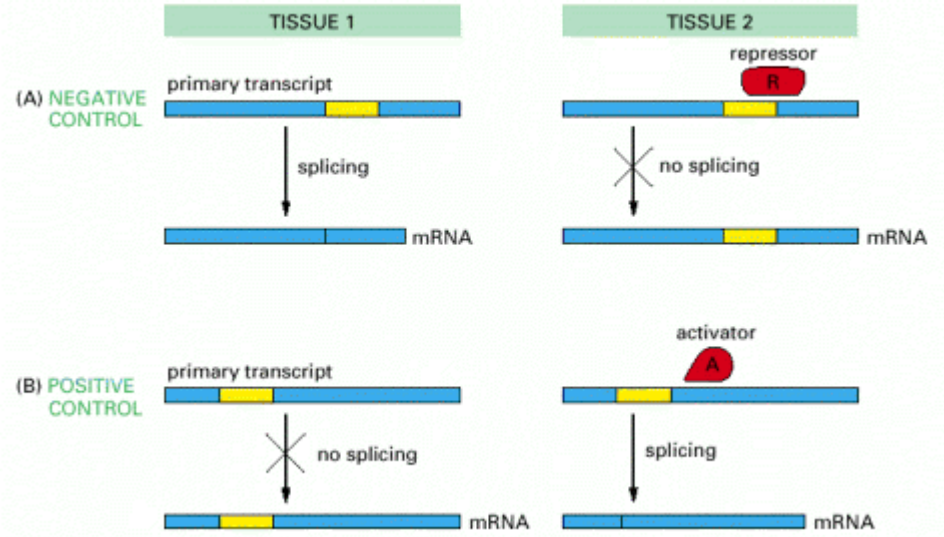
Basically remember that DNA methylation turns off the GENE. There is also such thing as Post-translational regulation, which are shown in the picture below, however only a few of those might be used for a determined gene. Green indicates greater methylation in twin 1 relative to twin 2. Red indicates lesser methylation (hypomethylation) in twin 1 relative to twin 2.





As we can see from the picture, alternative mRNA splicing can produce different forms of a protein from the same gene.

On the picture below we have negative regulation and positive regulation.



Alternate mRNA splicing can produce different forms of a protein from the same gene. Sometime many possibilities can give origin to many different proteins. Alternative splicing possibilities for Drosophila DSCAM gene are one example. Another example of a drastic developmental effect is during sex determination of Drosophila, which depends on a regulated series of splicing events.

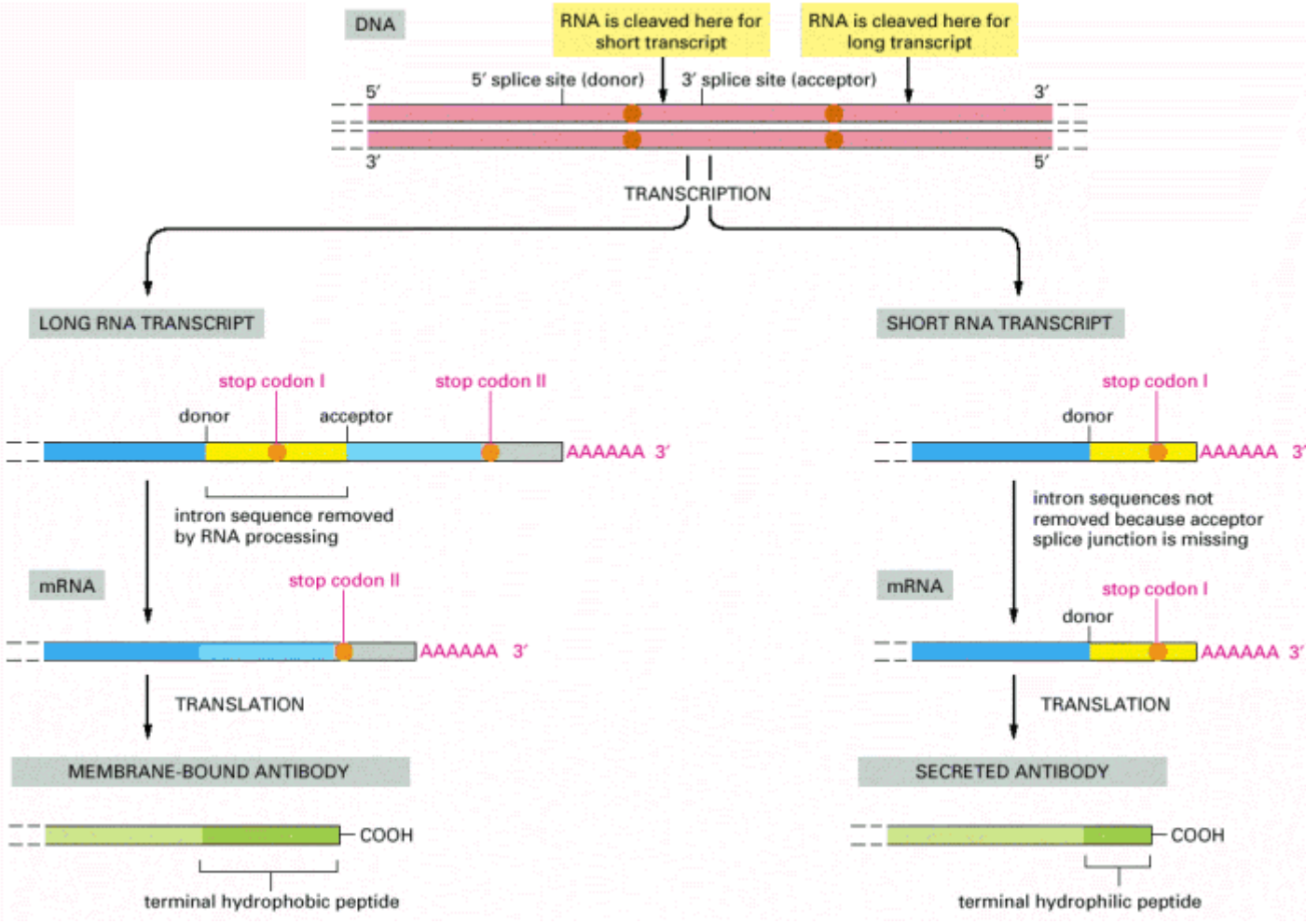


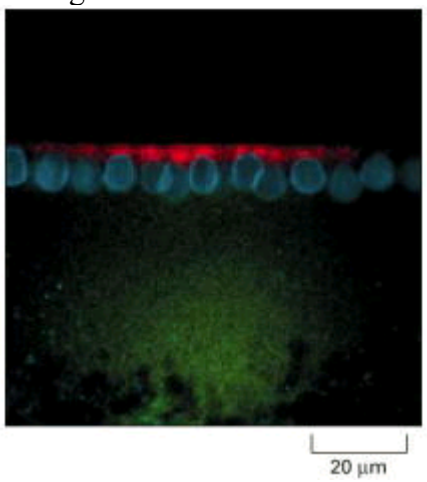
Figure 7-93. Regulation of the site of RNA cleavage and poly-A addition determines whether an antibody molecule is secreted or remains membrane-bound. In unstimulated B lymphocytes *(left),* a long RNA transcript is produced, and the intron sequence near its 3′ end is removed by RNA splicing to give rise to an mRNA molecule that codes for a membrane-bound antibody molecule. In contrast, after antigen stimulation *(right)* the primary RNA transcript is cleaved upstream from the splice site in front of the last exon sequence. As a result, some of the intron sequence that is removed from the long transcript remains as coding sequence in the short transcript. These are the nucleotide sequences that encode the hydrophilic C-terminal portion of the secreted antibody molecule.

There can also be regulation of mRNA transport from the nucleus to the cytoplasm. The nucleus is a double membrane-bound organelle, with controlled gating into and out of the organelle. After processing “left over” RNA fragments (introns, 3’cleaved sequences, incomplete/damaged RNA) are degraded in the nucleus in the RNA exonuclease-containing

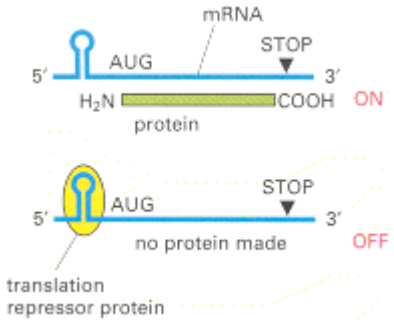
nucleus in the RNA exonuclease-containg degradates. Mature mRNA transcripts must be transported from nucleus to cytoplasm by regulated nuclear transport. With no stress the mouse CTN-RNA is retained in the nucleus by binding of nuclear retention factors to a retention motif in the 3'UTR of the mRNA and only small amounts of RNA are translated. At times of stress, the nuclear retention motif is cleaved and the coding sequences of the mRNA enter the cytoplasm to be translated. This is an example of how you can keep a protein in the nucleus.

Now also know that some mRNAs are localized to specific regions of the cytosol. The importance of the 3' untranslated region (UTR) in mRNA localization is:

- Two different version of the same RNA injected into Drosophila embryo.
- One mRNA had 3' UTR (labeled with red fluorescent dye) - localized
- One mRNA had 3' UTR missing (labeled with green fluorescent dye) –not localized
- Nuclei stained blue
- Only the mRNA with 3' UTR localizes to the apical side of the nuclei, the mRNA lacking the 3' UTR fail to localize. (picture below A)



A→



B→

Proteins that bind to the 5' and 3' untranslated regions of mRNAs negatively regulate translation. (Picture above B). Also the phosphorylation of initiation factor eIF-2 globally regulates translation.

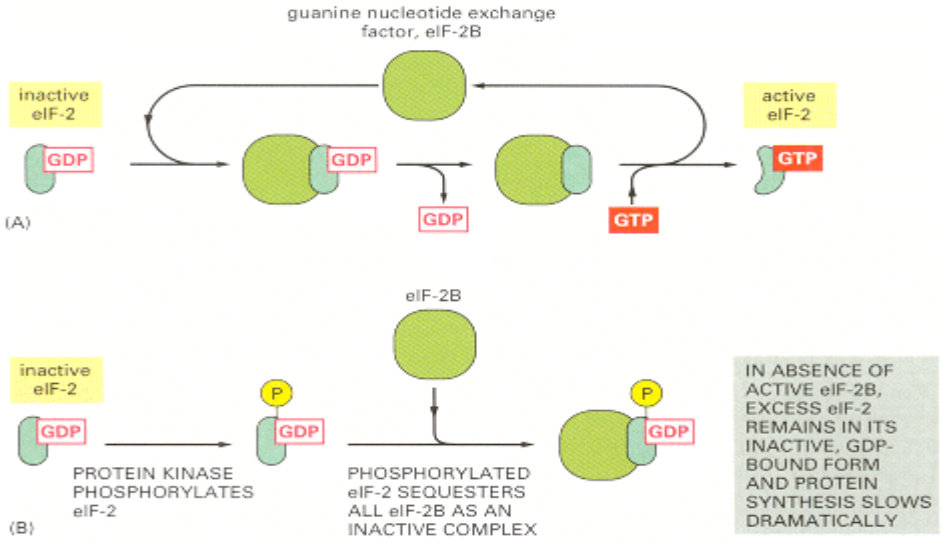


Figure 7-101. The elF-2 cycle. (A) The recycling of used elF-2 by a guanine nucleotide exchange factor (elF-2B). (B) elF-2 phosphorylation controls protein synthesis rates by tying up elF-2B.

Initiation at different AUG codons upstream of the translational start can regulate translation. Normally translation is initiated at AUG downstream from 5' end of mRNA. Keep in mind that AUG occurs by chance with a frequency of $1:4^n$, where n=3, so in every 64 nt – sometime 5' UTR is hundreds of nt long - how to initiate translation from the correct AUG? The context of the AUG is important (i.e., the bases surrounding it are important – some AUG codons are surrounded by nt which support good translation initiation, other are poor for translation initiation. So if the context is good you get more frequent initiation. The small ribosomal subunit "scans" the mRNA and initiates translation either infrequently or frequently depending on whether the codon is in a poor or good context respectively. "Kozak Scanning Hypothesis" can create proteins with different N-temini at different frequencies.

Internal ribosome entry site (IRES) can provide opportunities for translational control. IRES allows 5' cap-independent tranlation initiation. No absolute requirement for eIF-4E to initiate transcription. IRES typically several hundred nt in length and fold into special structure. Allows transcription of specific genes when eIF-4E is limiting, or under instances when eIR-4G is unable to bind eIF-4E (e.g., during programmed cell death – apoptosis)

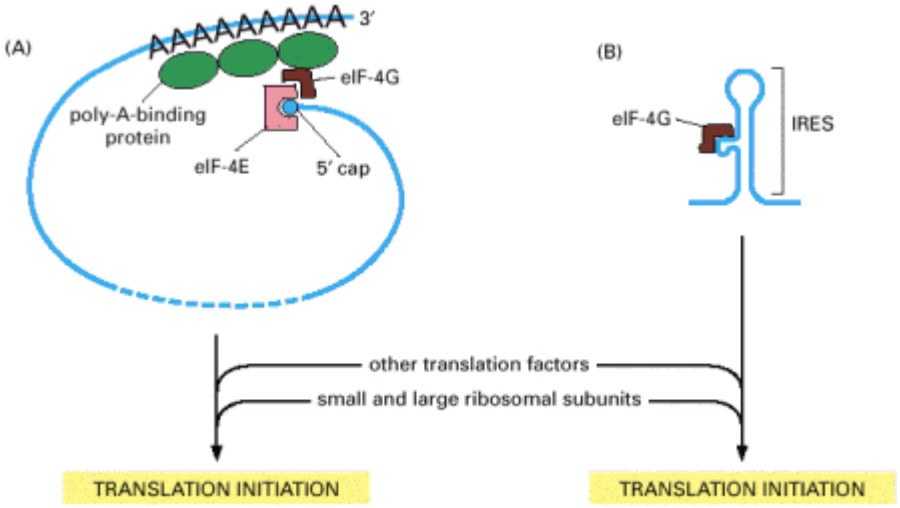


Figure 7-102. Two mechanisms of translation initiation. (A) The cap-dependent mechanism requires a set of initiation factors whose assembly on the mRNA is stimulated by the presence of a 5′ cap and a poly-A tail (see also Figure 6-71). (B) The IRES-dependent mechanism requires only a subset of the normal translation initiating factors, and these assemble directly on the folded IRES.

Gene expression can be controlled by a change in mRNA stability. Nonsense-mediated mRNA decay functions as a mRNA quality control system in eukaryotes. So basically there are nuclear proteins tha mark exon-exon boundaries on a spliced mRNA molecule. To be more clear, the DNA was transcribed and it contains mRNA for several exons and introns, and these proteins are binded to the sections where the introns were cut off, they bind during the the splicing of the mRNA. These proteins are also thought to help

with the movement of this mRNA to the cytosol from the nucleus. Now after it is in the cytosol it is "tested" by the ribosome, which locates all these proteins. However if a in-frame stop codon is found before the protein that shows the exon-exon boundary, the mRNA is subjected to nonsense-mediated decay. If not then the mRNA goes on to be translated. This is believed to occur just outside the nucleus.

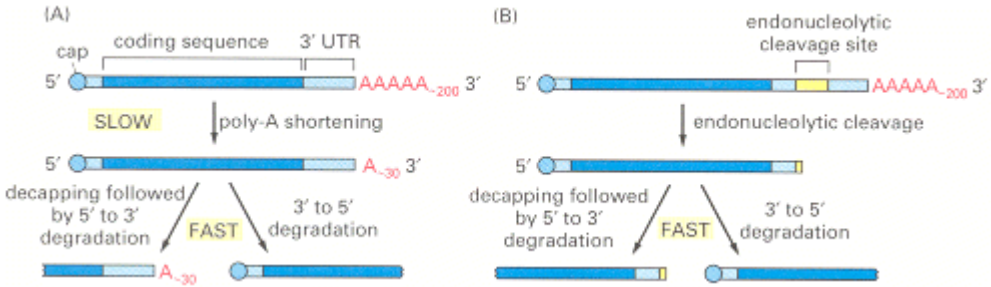


Figure 7-103. Two mechanisms of eucaryotic mRNA decay. (A) Deadenylation-dependent decay. Most eucaryotic mRNAs are degraded by this pathway. The critical threshold of poly-A tail length that induces decay may correspond to the loss of the poly-A binding proteins (see Figure 6-40). As shown in Figure 7-104, the deadenylation enzyme associates with both the 3′ poly-A tail and the 5′ cap, and this arrangement may coordinate decapping with poly-A shortening. Although 5′ to 3′ and 3′ to 5′ degradation are shown on separate RNA molecules, these two processes can occur together on the same molecule. (B) Deadenylation-independent decay. It is not yet known with certainty whether decapping follows endonucleolytic cleavage of the mRNA.

The 5' Cap and the 3' poly-A tail are used in translation initiation and mRNA decay. The translation initiation machinery associates with 5' cap and 3' poly-A tail. Deadenylation-dependent mRNA decau initiated by an enzyme (called DAN, deadenylating nuclease) that shortens poly-A tail in 3' to 5' direction.