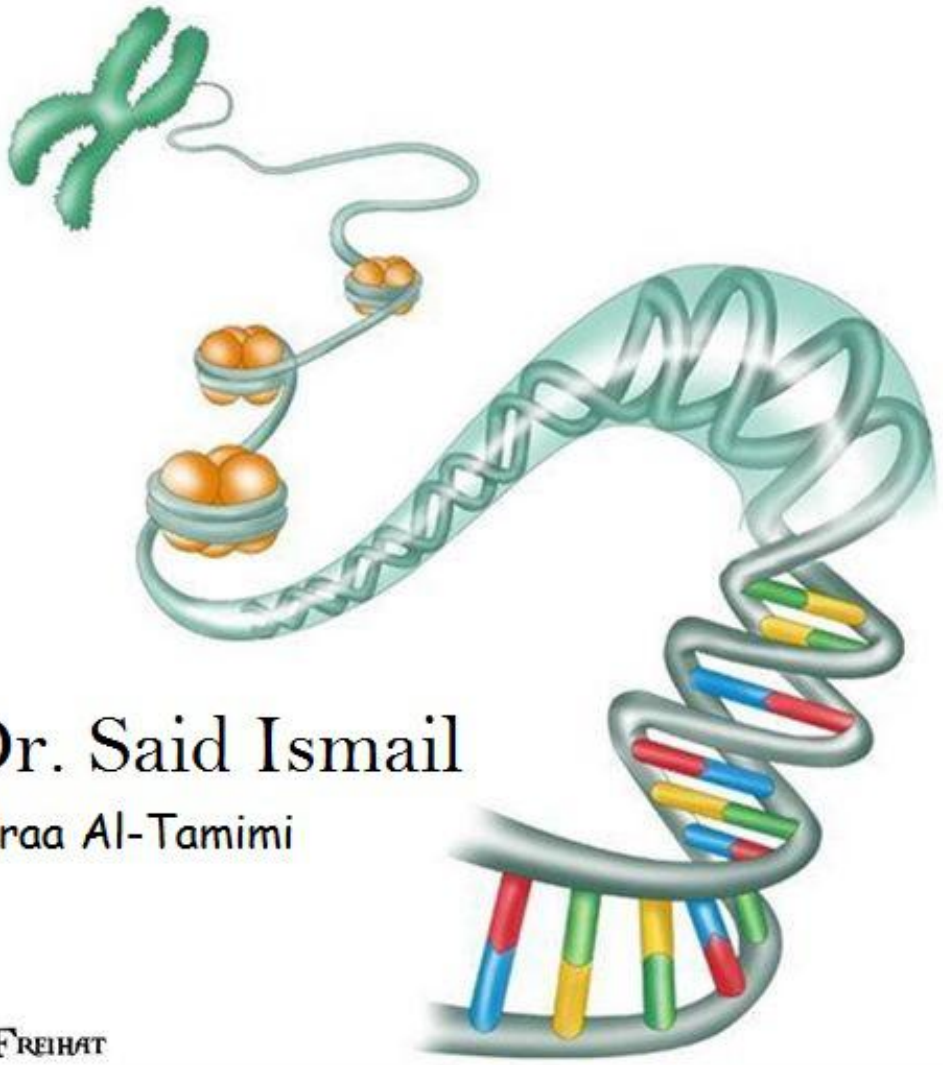
University of Jordan
Faculty of Medicine
Batch 2013-2019

Medical Committee
The University of Jordan

Genetics & Molecular Biology

○ Slides ● Sheet ○ Handout ○ other.....



Sheet#: 21
Dr. Name: Dr. Said Ismail
Done By: Zahraa Al-Tamimi

Designed by Nadeen Al-Freihat

Gene Regulation

The following issue is always brought up with regards to promoter sequences: when the doctor talks about them and when they are brought up in books, they are always mentioned in the coding sequence. Sequences are always defined based on the coding strand, even if the work being done by an enzyme (ex. RNA Pol, transcription factors) is on the template strand.

Beginning of this lecture's material

Last time, we talked about regulation of gene expression in eukaryotes. We divided it into 5 levels: regulation at the DNA level, transcriptional level, post-transcriptional level, translational level, and post-translational level.

We mentioned that the first two examples in the **DNA level** are the most important ones. Namely, **chromatin remodeling** (conversion between heterochromatin and euchromatin , condensed chromatin and loose or diffused chromatin), and **DNA methylation** (cells methylate their promoter regions to hide it from transcriptional machinery). **Gene rearrangement** is an example specific to B cells (Antibody producing cells). **Gene amplification** is abnormal and occurs in cancer cells (HER-2 gene and breast cancer). **Gene deletion** is another extreme measure by which RBCs delete genes they don't need any more, not just switch them off.

Now we move on to the Transcriptional level (most important level):

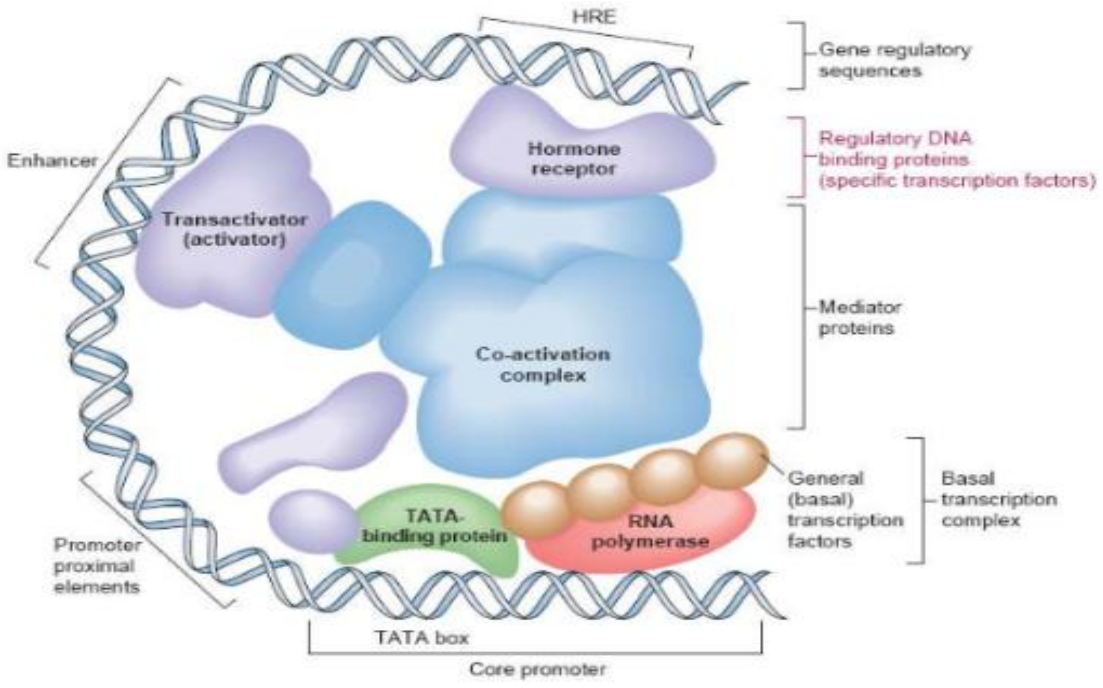
We should be familiar with the two sides of this story. Namely, the **Cis-acting elements** (DNA sequences) and **trans-acting elements** (transcription factors which bind and identify these sequences).

The two most important **Cis-acting elements** are the promoters and enhancers. **Trans-acting elements** are the transcription factors that bind to the Cis-acting elements.

To summarize, there are sequences to the 5' of the gene (just before the gene), and there are transcription factors that bind those 5' sequences.

Promoter and enhancer regions each have different elements. An element is a short sequence identified by a specific transcription factor. So, when we talk about a promoter, it's not just one sequence that is the binding site for just one transcription factor. It has the TATA box, the GC rich boxes, the CAAT boxes. There are many sequence. It has proximal promoter elements, which are close to the gene, and distal promoter elements that are away from the gene. Each one of these elements is a docking site for a specific transcription factor. The same goes for enhancers. An enhancer isn't just one element. Rather, it's made up of several sequences/elements that each act as binding sites for a transcription factor.

Let's look at the following figure:



We agreed that the core promoter is just upstream of the transcription start site (the beginning of the coding region of the gene). One of the main consensus elements (conserved elements) of the promoter is the TATA box. However, there are other elements such as the promoter proximal elements and promoter distal elements (although they are not pointed out in the figure above). Then there is the enhancer, which can be very far away from the promoter. Notice that there are transcription factors binding at the

promoter and transcription factors binding at the enhancer. Those binding specifically at the core promoter are called **basal or general transcription factors**. While those binding the enhancer are called **specific transcription factors**. They are called general for promoters because a promoter is a general sequence that can be found before all genes (example :TATA boxes, the TATA binding protein and its transcription factors will be attached to every gene that has the TATA box and that why it's called general). Enhancers, however, are very specific. An enhancer can be shared by a few genes (not only a single gene) , but not more than that.

Why is the DNA bending this way (in the above figure)? It's to show you what most likely happens when transcription occurs. We know that there are transcription factors that bind promoters and others that bind enhancer. We also know that enhancers can be very far away from promoters. How would they communicate if the enhancers are so far away?

The DNA will bend ; so that the enhancer and the promoter can communicate via **mediator proteins**. Even if they are close to each other, they still use mediator proteins to communicate (they don't communicate directly). Mediator proteins are the proteins that pass on information from the enhancer transcription factors to the promoter transcription factors (from **specific to general**). Sometimes, they're called **co-activators**. Recall that we mentioned that a promoter on its own, with its transcription factors, can only induce very weak transcription. For efficient transcription to happen, a message from the enhancer is needed (and it's passed through the co-activators).

The final goal of the proteins in the picture is to recruit RNA Polymerase. Without them, RNA Polymerase can't identify the promoter and start transcription. As long as they are assembled here, it will keep transcribing until they break away.

Some transcription factors can directly bind to DNA, while others can't. Specific transcription factors can directly bind the enhancer. The general transcription factors, however, can't bind directly to the promoter. They

depend on the TATA binding protein to read the promoter, and then they bind to the TATA binding protein. DNA binding proteins have specific structural motifs that enable them to bind and read DNA. When you study these motifs (Zinc fingers, leucine zippers, helix-turn-helix, and helix-loop-helix) just know that they are structural motifs that enable a protein to bind DNA and read it; don't go into specifics when you read the book.

What about trans-activators? Are they activators or specific transcription factors? They are both. These proteins bind to enhancer regions and can be called **specific transcription factors** or **trans-activators**. "Trans" is a word that implies distance; so, trans-activators are proteins that activate transcription from a distance. They can also be called hormone receptors (hormone receptor specific transcription factor), meaning they can be activated by hormones. They are found in the cell inactive and when bind to a hormone they are activated . Once they are activated by the hormone, they work as transcription factors. In this case, the enhancer can be called a hormone response element ;it's a sequence that response to protein that is a receptor to a hormone. Once again, hormone receptors are a special kind of transcription factor that are activated by hormones.

So, we have agreed that there are three main types of proteins working at the transcription level: basal/general, specific (trans-activators, hormone receptors) or the co-activators that are in the middle.

Examples of gene regulation at the transcription level:

1) An example of a specific transcription factor is the **Glucocorticoid Receptor (GR)**. GR is a specific trans-activator that is receptive to hormones such as cortisol. This hormone receptor is found in the cytoplasm. However, transcription factors need to be in the nucleus. Thus, it enters the nucleus once the hormone is bound to it. HSB (heat shock protein) is holding GR in the cytoplasm and prevents it from entering the nucleus. How do nuclear proteins know that they should be in the nucleus? How do they leave the cytoplasm and enter nucleus? They do so through the **NLS**

(**N**uclear **L**ocalizing **S**ignal). NLS is a bit of a polypeptide that functions as a signal for transport (shuttle)proteins. Transport proteins hold on to the NLS and transport the protein to the nucleus. The HSB hides this signal (the NLS) because it's so large. Once cortisol arrives and binds to GR, HSB is displaced. Cortisol's affinity for GR is higher than that of HSB to GR. Cortisol is not as big as HSB so it won't hide the NLS once it's bound to GR.

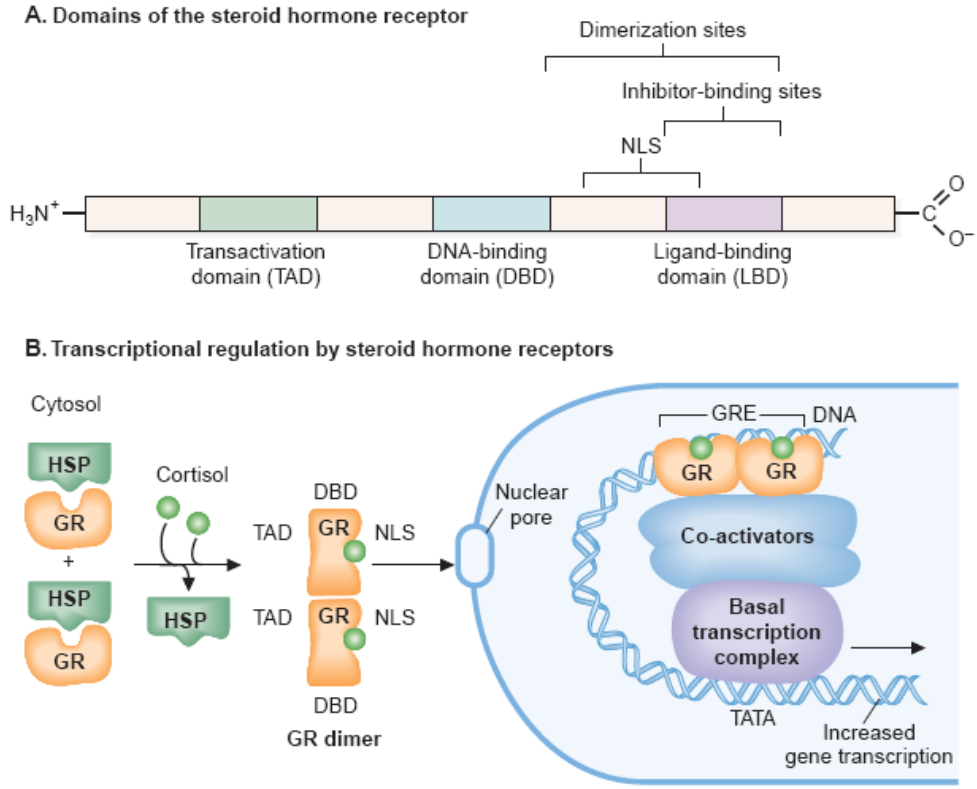


FIG. 16.12. Steroid hormone receptors. **A.** Domains of the steroid hormone receptor. The transactivation domain (TAD) binds coactivat DNA-binding domain (DBD) binds to the hormone response element in DNA; ligand-binding domain (LBD) binds hormone; NLS is the nucl localization signal; the dimerization sites are the portions of the protein involved in forming a dimer. The inhibitor-binding site binds heat-sh proteins and masks the nuclear localization signal. **B.** Transcriptional regulation by steroid hormone receptors. *HSP*, heat-shock proteins; *G* glucocorticoid response element; *GR*, glucocorticoid receptor.

Once GR is inside the nucleus, it knows exactly where to go (to the enhancer). It functions as transcription factor for its gene, through co activators(mediators)→basal transcription complex →RNA polymerase will start transcription

*Note: GR works in pairs (it functions in the form of a homodimer, two GR molecules together). Also, there are some transcription factors that function in the form of heterodimers.

2) Another example involves the **Thyroid Receptor (TR)**:

It's a T3 receptor(T3,هرمونات الغدة الدرقية), meaning T3 is its ligand. The cell controls its function differently from GR. TR is already found in the nucleus. It works as a heterodimer with **RXR**. Without binding to its ligand (T3), TR's conformation attracts co-repressor mediators which pass a negative signal to the promoter. When T3 arrives, it binds TR in the nucleus (at the enhancer), induces a conformational change, and starts to recruit co-activators.

**Note: RXR is a partner to many proteins and form with them heterodimer transcription factors. As a result, it's a target in cancer therapy. If you take it away, you can control the transcription of many genes ; because it is involved in many transcription factors.

Note: TR binds the enhancer and induces repression through repressive mediators.That's why **enhancers can be silencers some times and activators at other times.

- Enhancer sequence +TR = repressive mediator (co-repressor) = the enhancer sequence acts as a silencer .
- Enhancer +[TR+T3]= activating mediator (co-activator) = the enhancer sequence acts as an activator .

How do repressors and co-activators work? How do they exert their effects on the general transcription factor?

Co-repressors have HDAC activity (histone deacetylase activity). They remove acetyl groups from histones > exposure of positive charge> DNA binds to histones= no transcription

T3 binds to TR > conformational change > repressor goes away and activator binds > activator has histone acetylase activity > represses or acetylates lysine residues > this hides positive charge > DNA doesn't bind (loose DNA)= transcription .

It's not easy for any protein to bind and read DNA. It has to have structural motifs (zinc fingers..etc). Most transcriptional proteins have them.

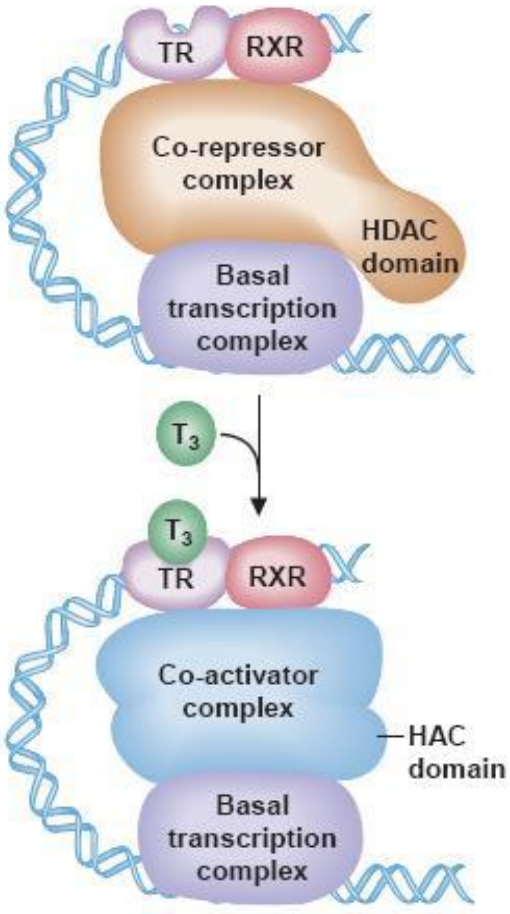


FIG. 16.13. Activity of the thyroid hormone receptor–retinoid receptor dimer (TR-RXR) in the presence and absence of thyroid hormone (T_3). *HAC*, histone acetylase; *HDAC*, histone deacetylase.

Transcription factors control our genes, but they need to be controlled as well. They are proteins. If we want to control them, we can do so in several ways :

1) **Switch off/on its promoter**. In other words, there is another transcription factor that controls it. Some transcription factors are more potent than others(in two ways).First one, More potent, in this case, means that it binds to several enhancers of several genes, and thus, controls many genes. Second, A transcription factor could also be more potent if it activates 4/5 genes, which are also transcription factors. Thus, this potent transcription factor results in waves of more gene activation (Look at Figure 16.16).

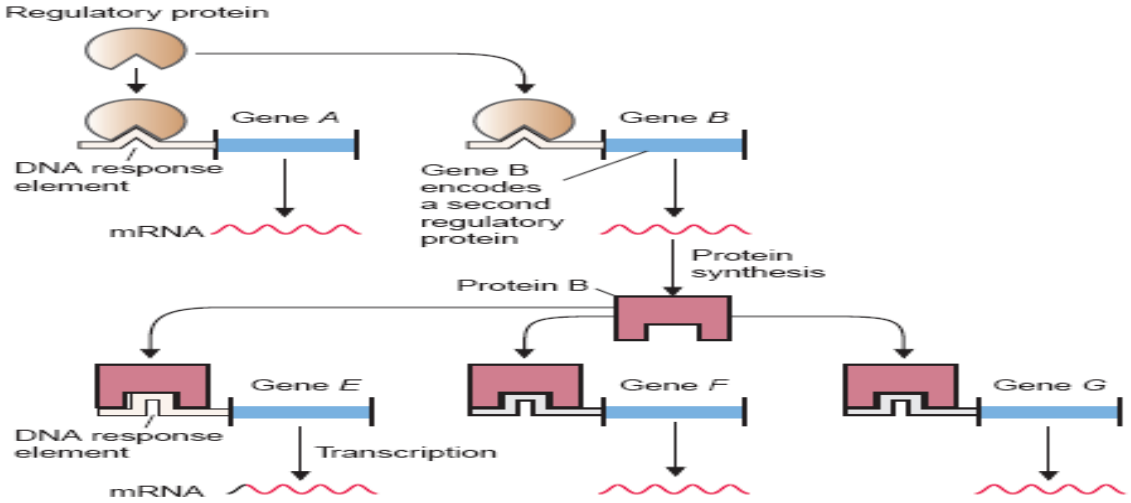


FIG. 16.16. Activation of sets of genes by a single inducer. Each gene in a set has a common DNA regulatory element, so one regulatory protein can activate all the genes in the set. In the example shown, the first regulatory protein stimulates the transcription of genes *A* and *B*, which have a common DNA regulatory sequence in their control regions. The protein product of gene *B* is itself a transcriptional activator, which in turn stimulates the transcription of genes *E*, *F*, and *G*, which likewise contain common response elements.

Negative feedback can stop this cascade of activation. Somewhere down, at the 7th or 8th wave of activation, a protein may be produced that has an inhibitory effect on that original, potent, transcription factor.

2) **Phosphorylation/ Dephosphorylation:** At the protein level, most of them are phosphoproteins. Phosphorylation = active, dephosophorylation= inactive. We can control them through kinases and phosphatases.

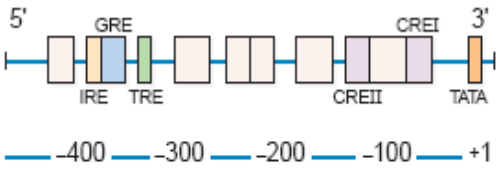


FIG. 16.17. A simplified view of the regulatory region of the PEPCK gene. *Boxes* represent various response elements in the 5′-flanking region of the gene. Not all elements are labeled. Regulatory proteins bind to these DNA elements and stimulate or inhibit the transcription of the gene. This gene encodes the enzyme PEPCK, which catalyzes a reaction of gluconeogenesis (the pathway for production of glucose) in the liver. Synthesis of the enzyme is stimulated by glucagon (by a cAMP-mediated process), by glucocorticoids, and by thyroid hormone. Synthesis of PEPCK is inhibited by insulin. *CRE*, cAMP response element; *TRE*, thyroid hormone response element; *GRE*, glucocorticoid response element; *IRE*, insulin response element.

3) **Control its partner:** an example of this is controlling TR through RXR. TR can't function on its own. It needs to form a heterodimer with RXR. Thus, controlling RXR levels in the cell controls TRs Activity.

The promoter region/enhancer region has several promoter elements (as we mentioned before), and each element binds a different set of transcription factors. The net result of the binding of several proteins at promoter region, and several at the enhancer regions, decides how much of the protein is eventually transcribed. These transcription factors "discuss" the situation, "decide" how much mRNA is to

be transcribed, and mediate the message to RNA Polymerase. They do this with high accuracy and efficiency(All of this can be done due to structural fitting between these proteins ,but the doctor thinks that these proteins have small brains and they have their own language to communicate between each other).

Sometimes, the cell encounters new changes that make the first decision no longer appropriate. The cell can modify the numbers of proteins produced in **the post-translational level.**

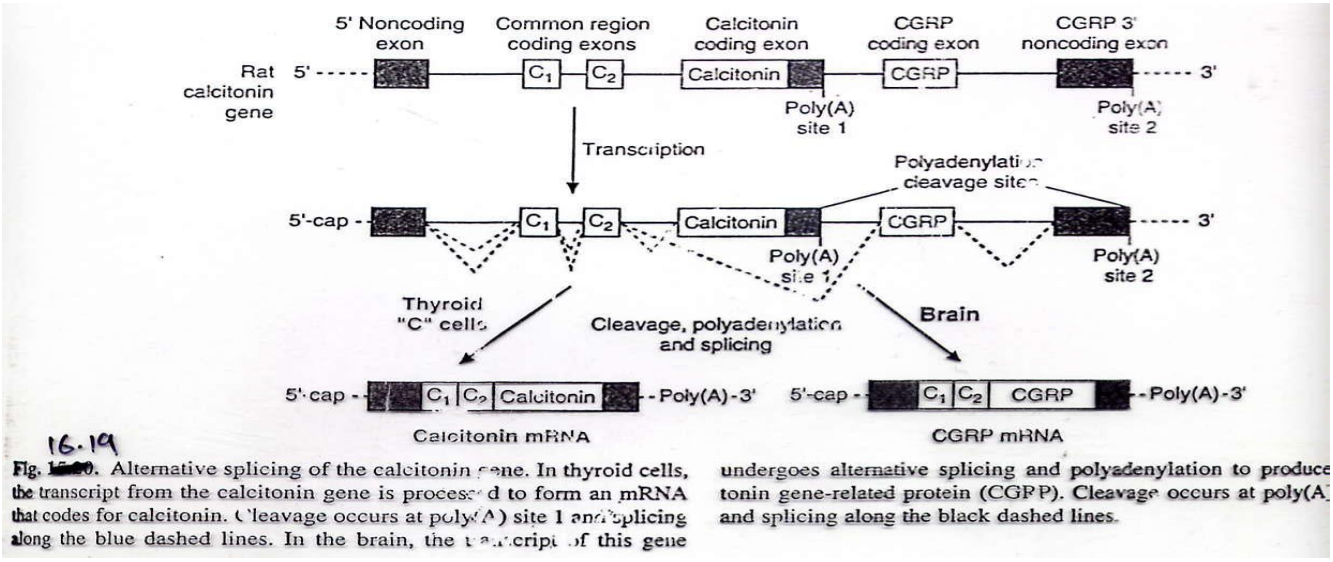
Regulation at the post-transcriptional level:

There is qualitative regulation and quantitative regulation at this levels.

Quantitative: the cell holds back the mRNA in the nucleus. It sends out less mRNA copies to the cytoplasm, and degrades the rest in the nucleus.

Qualitative: the cell can decide what proteins to produce from the same primary mRNA. This is done through **alternative splicing** and **alternative polyadenylation.**

An example of this is the gene in Fig. 16.19 in thyroid vs. brain cells. Different cells have the same gene, same primary mRNA, but cut in different ways in the two cells.



16.19

Fig. 16.19. Alternative splicing of the calcitonin gene. In thyroid cells, the transcript from the calcitonin gene is processed to form an mRNA that codes for calcitonin. Cleavage occurs at poly(A) site 1 and splicing along the blue dashed lines. In the brain, the transcript of this gene undergoes alternative splicing and polyadenylation to produce tonin gene-related protein (CGRP). Cleavage occurs at poly(A) and splicing along the black dashed lines.

Thyroid cells make calcitonin, a hormone involved in calcium and vitamin D metabolism. CGRP is produced in brain cells from the same gene and is

involved in taste sensations in the brain, a function completely different from that of calcitonin.

We get two different proteins because of **alternative splicing**. The first and the second introns are removed in both calcitonin and CGRP mRNAs. However, different portions are removed as the third intron. This results in two different mature mRNAs, which will result in two different proteins. The calcitonin exon is considered part of the third CGRP intron.

Polyadenylation is also different in these two mRNAs. There are two polyadenylation sites in this genes[poly-A site 1 and 2]. The poly-adenylation site right after the calcitonin exon(poly-A site 1) is used when the mRNA is produced in the thyroid cells (to make calcitonin), because the other polyadenylation site is cleaved. In the brain cell, however, the polyadenylation site after calcitonin is cleaved as part of the third intron in the CGRP mRNA, and the polyadenylation site(poly-A site 2) at the very end of the gene is used instead.

Sometimes, this can happen in the same cell, not in two different cells. The cell produces a certain version of the mRNA at one time, and another version at another time from the same gene.

RNA editing: the same mRNA, same gene. Apoprotein B is made of 4563 amino acids. The stop codon is at position 4563. However, intestinal cells go to a codon in the middle and change the C to a U. This results in a UAA codon (stop codon). In these cells, that's where transcription stops, resulting in a much shorter protein. This happened at the mRNA level, not at the DNA level (the DNA still CAA).

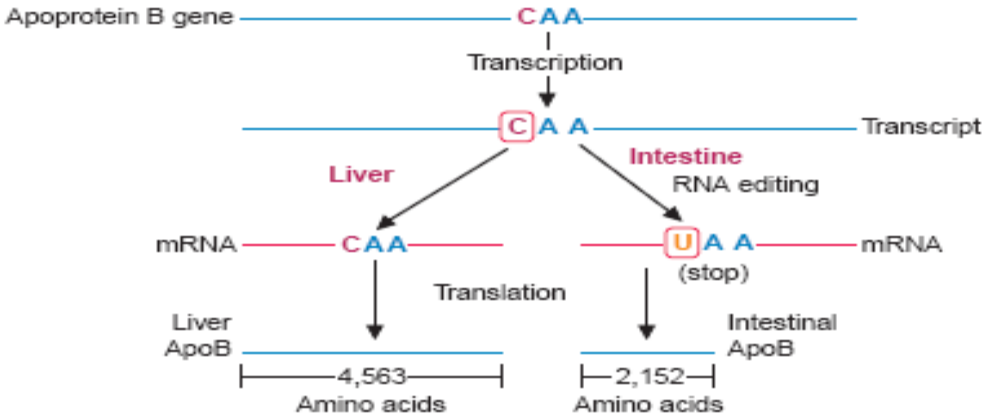


FIG. 16.19. RNA editing. In liver, the apoprotein B (*ApoB*) gene produces a protein that contains 4,563 amino acids. In intestinal cells, the same gene produces a protein that contains only 2,152 amino acids. Conversion of a C to a U (through deamination) in the RNA transcript generates a stop codon in the intestinal mRNA. Thus, the protein produced in the intestine (named apo B-48) is only 48% of the length of the protein produced in the liver (named apo B-100).

Regulation at a <u>translational level</u>: mRNA is synthesized, processed, and is out in the cytoplasm.

Examples:

1) Immature RBC's HRI Kinase:

This occurs at the last stage of differentiation. RBCs are concerned with producing only one protein, <u>hemoglobin</u>. In a short while, the RBC will lose all its DNA and will live for 120 days. It wants to synthesize enough hemoglobin to keep it alive for 120 days. However, it doesn't want to make more protein than what it has in heme; because that will waste energy and time. Don't underestimate the cell. It is very efficient with the use of its resources. It wants to terminally differentiate. How can it tell how much heme it has?

It does this via the heme kinase (HRI).

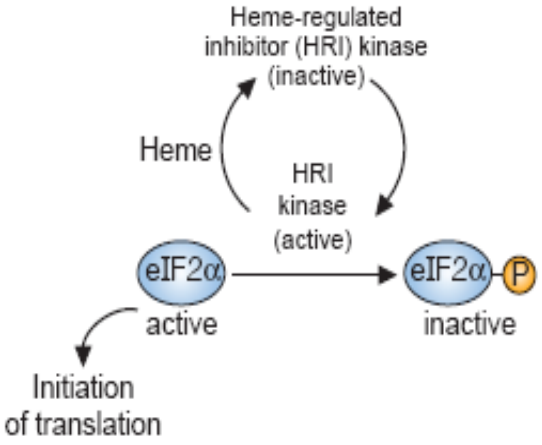


FIG. 16.20. Heme prevents inactivation of eIF2α. When eIF2α is phosphorylated by the heme-regulated inhibitor kinase, it is inactive and protein synthesis cannot be initiated. Heme inactivates the heme-regulated inhibitor kinase, thereby preventing phosphorylation of eIF2α and activating translation of the globin mRNA.

It's a kinase that phosphorylates a transcription initiation factor called IF2 (Figure 16.20).

When IF2 is phosphorylated, it's inactive. When IF2 is dephosphorylated, it's active. Initiation factors assemble the ribosome on the mRNA. In this case, it's the mRNA of globin. Without HRI kinase, globin mRNA won't be translated into protein. The heme affects the activity of HRI kinase. High levels of heme will inactive the HRI kinase, and so, it won't phosphorylate IF2, thus activating IF2 and translating the globin mRNA into proteins. Too little heme means no (HRI kinase inhibition) >HRI kinase active> phosphorylation of IF2 > inactive IF2 = no translation of globin mRNA into protein .

2) Ferritin synthesis:

Ferritin is a protein responsible for the storage of iron. A lot of iron in the cell means there is a need for a lot of ferritin, and vice versa. How does the cell know how much ferritin to make?

The mRNA of ferritin has a loop (hair-pin loop) at the 5' end (Iron Response Element). The **I**ron **R**esponse **E**lement (**IRE**) binds the **I**ron **R**esponse **E**lement **B**inding **P**rotein (**IRE-BP**). When IRE-BP is bound to IRE, translation is blocked because the loop is created. The ribosome can't move on and translate. Iron can take away IRE-BP.

Iron > no loop > ribosomes can translate

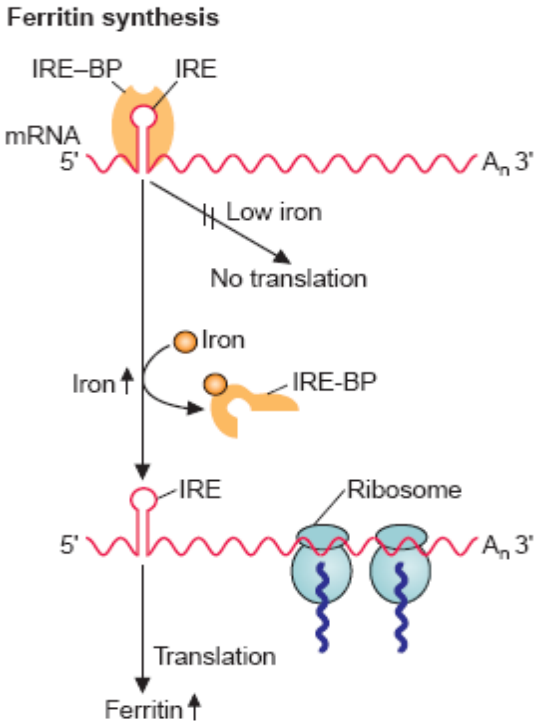


FIG. 16.21. Translational regulation of ferritin synthesis. The mRNA for ferritin has an IRE. When the IRE-BP does not contain bound iron, it binds to IRE, preventing translation. When IRE-BP binds iron, it dissociates, and the mRNA is translated.

-mRNA to make ferritin .

-No Iron > loop > ribosome can't move > no translation > low ferritin.

3) Transferrin Receptor:

Transferrin is responsible for the transport of iron. The transferrin receptor is a receptor on cell membranes that binds to transferrin, which has iron. So if cell needs iron, it makes the transferrin receptor so it binds to transferrin and gets iron. If the cell has enough iron, it should not make more transferrin receptors. How does this happen? It happens through a hair-pin loop, once again. However, in this case, it's on the 3' end. Instead of preventing ribosome movement , like it did with ferritin, its position at the 3' end prevents RNAse movement (RNAses break down mRNA). This prolongs mRNA survival and helps create more of the transferrin receptors.

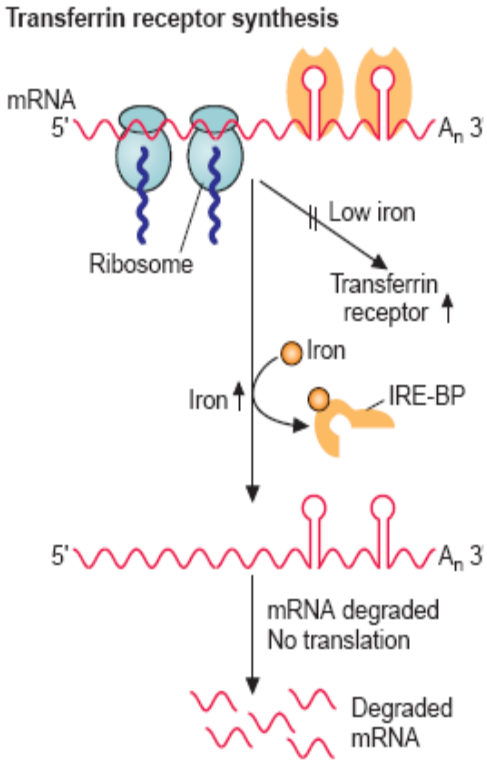


FIG. 16.23. Regulation of degradation of the mRNA for the transferrin receptor. Degradation of the mRNA is prevented by binding of the IRE-BP to IREs, which are hairpin loops located at the 3′-end of the transferrin receptor mRNA. When iron levels are high, IRE-BP binds iron and is not bound to the mRNA. The mRNA is rapidly degraded, preventing synthesis of the transferrin receptor.

-A lot of iron in the cell > iron will bind IRE-BP > IRE-BP removed from mRNA > no hair-pin loops (3' end exposed to RNAse activity) > RNAse activity not stopped > less translation > less tranferrin receptors > less iron brought into cell.

-Low iron→ IRE-BP not removed from mRNA→ hair-pin loops→less RNAse activity→more translation → more tranferrin receptors → more iron brought into cell.

**The doctor requested that you remind him on Sunday, April 5th, that there are only 2 slides left to cover from this chapter (chapter 16).

*****The End*****