# **16** Regulation of Gene Expression



The prokaryotic bacterium Escherichia coli can use a wide range of different nutrients to sustain growth. It accomplishes this partially by turning on and off expression of the genes required for synthesis of enzymes in different metabolic pathways. For example, enzymes encoded by the lac operon allow E. coli to use lactose for energy and the presence of lactose activates transcription of these genes. Thus, the nutrient itself is involved in regulating gene expression.

In contrast, the cells of the adult human are exposed to a nearly constant and safe environment composed of blood and interstitial fluid. Hormones maintain this nearly constant environment, in spite of changes in nutrient demand and availability, partially by regulating gene transcription. The immune system protects cells against foreign organisms, partially by controlling the expression of genes required for the immune response. However, some nutrient regulation of gene expression also occurs. For example, iron controls expression of the genes for its storage and transport proteins at the level of their mRNAs.

Gene expression, the generation of a protein or RNA product from a particular gene, is controlled by complex mechanisms. Normally, only a fraction of the genes in a cell are expressed at any time. Gene expression is regulated differently in prokaryotes and eukaryotes.

**Regulation of gene expression in prokaryotes.** In prokaryotes, gene expression is regulated mainly by controlling the **initiation of gene transcription**. Sets of genes encoding proteins with related functions are organized into operons, and each operon is under the control of a single promoter (or regulatory region). Regulatory proteins called *repressors* bind to the promoter and inhibit the binding of RNA polymerase (negative control), whereas activator proteins facilitate RNA polymerase binding (positive control). Repressors are controlled by nutrients or their metabolites, classified as **inducers** or corepressors. Regulation also may occur through attenuation of transcription.

Eukaryotes: Regulation of gene expression at the level of DNA. In eukaryotes, activation of a gene requires changes in the state of chromatin (chromatin remodeling) that are facilitated by acetylation of histones and methylation of bases. These changes in DNA determine which genes are available for transcription.

**Regulation of eukaryotic gene transcription.** Transcription of specific genes is regulated by proteins (called specific transcription factors or transactivators) that bind to gene regulatory sequences (called promoter-proximal elements, response elements, or enhancers) that activate or inhibit assembly of the basal transcription complex and RNA polymerase at the TATA box. These specific transcription factors, which may bind to DNA sequences some distance from the promoter, interact with coactivators or corepressors that bind to components of the basal transcription complex. These protein factors are said to work in "trans"; the DNA sequences to which they bind are said to work in "cis."

**Other sites for regulation of eukaryotic gene expression.** Regulation also occurs during the **processing** of RNA, during RNA transport from the nucleus to the cytoplasm, and at the level of translation in the cytoplasm. Regulation can occur simultaneously at multiple levels for a specific gene, and many factors act in concert to stimulate or inhibit expression of a gene.



#### WAITING ROOM тне

Arlyn Foma, a 68-year old man, complained of fatigue, loss of appetite, and a low-grade fever. An open biopsy of a lymph node indicated the presence of non-Hodgkin's lymphoma, follicular type. Computed tomography and other noninvasive procedures showed a diffuse process with bone marrow

involvement. He is receiving multidrug chemotherapy with AV/CM (doxorubicin [adriamycin], vincristine, cyclophosphamide, and methotrexate). His disease is not responding well to this regimen, and the follicular lymphoma appears to be evolving into a more aggressive process. Because recombinant interferon  $\alpha$ -2b has been reported to have synergistic or additive effects with these agents, it is added to the protocol. Although resistance to methotrexate is considered, the drug is continued as part of the combined therapeutic approach.



Mannie Weitzels is a 56-year-old male who complains of headaches, weight loss related to a declining appetite for food, and a decreasing tolerance for exercise. He notes discomfort and fullness in the left upper quadrant of his abdomen. On physical examination, he is noted to be pale and to have ecchymoses (bruises) on his arms and legs. His spleen is markedly enlarged.

Initial laboratory studies show a hemoglobin of 10.4 g/dL (normal = 13.5-17.5g/dL) and a leukocyte (white blood cell) count of 86,000 cells/mm<sup>3</sup> (normal = 4,500–11,000 cells/mm<sup>3</sup>). Most of the leukocytes are granulocytes (white blood cells arising from the myeloid lineage), some of which have an "immature" appearance. The percentage of lymphocytes in the peripheral blood is decreased. A bone marrow aspiration and biopsy show the presence of an abnormal chromosome (the Philadelphia chromosome) in dividing marrow cells.



Ann O'Rexia, who has anorexia nervosa, has continued on an almost meat-free diet (see Chapters 1, 3, 9, and 11). She now appears emaciated and pale. Her hemoglobin is 9.7 g/dL (normal = 12-16 g/dL), her hema-

tocrit (volume of packed red cells) is 31% (reference range for women = 36-46%), and her mean corpuscular hemoglobin (the average amount of hemoglobin per red cell) is 21 pg/cell (reference range = 26-34 pg/cell). These values indicate an anemia that is microcytic (small red cells) and hypochromic (light in color, indicating a reduced amount of hemoglobin per red cell). Her serum ferritin (the cellular storage form of iron) was also subnormal. Her plasma level of transferrin (the iron transport protein in plasma) was greater than normal, but its percent saturation with iron was below normal. This laboratory profile is consistent with changes that occur in an iron deficiency state.

#### **GENE EXPRESSION IS REGULATED FOR** Ι. **ADAPTATION AND DIFFERENTIATION**

Although most cells of an organism contain identical sets of genes, at any given time only a small number of the total genes in each cell are expressed (that is, generate a protein or RNA product). The remaining genes are inactive. Organisms gain a number of advantages by regulating the activity of their genes. For example, both prokaryotic and eukaryotic cells adapt to changes in their environment by turning the expression of genes on and off. Because the processes of RNA transcription and protein synthesis consume a considerable amount of energy, cells conserve fuel by making proteins only when they are needed.

In addition to regulating gene expression to adapt to environmental changes, eukaryotic organisms alter expression of their genes during development. As a fertilized egg becomes a multicellular organism, different kinds of proteins are synthesized in varying quantities. In the human, as the child progresses through adolescence and then into adulthood, physical and physiologic changes result from variations in gene expression and, therefore, of protein synthesis. Even after an organism has reached the adult stage, regulation of gene expression enables certain cells to undergo differentiation to assume new functions.

Each of the drugs used by Arlyn Foma inhibits the proliferation of cancer cells in a different way. Doxorubicin (adriamycin) is a large nonpolar molecule synthesized by fungi that intercalates between DNA bases, inhibiting replication and transcription and forming DNA with single- and double-stranded breaks. Vincristine binds to tubulin and inhibits formation of the mitotic spindle, thereby preventing cell division. Cyclophosphamide is an alkylating agent that damages DNA by covalently attaching alkyl groups to DNA bases. Methotrexate is an analogue of the vitamin folate. It inhibits folate-requiring enzymes in the pathways for synthesis of thymine and purines, thereby depriving cells of precursors for DNA synthesis.

E. coli is a facultative anaerobe, which means that it can grow in the presence or absence of oxygen. The switch to oxygen-requiring pathways for fuel metabolism is under control of arc, the aerobic respiration control gene. When arc is activated, transcription is increased by 1,000-fold or more for enzymes in the pathways that ultimately transfer electrons to oxygen (e.g., proteins of the respiratory chain, the TCA cycle, and fatty acid oxidation). In the absence of oxygen (i.e., anaerobic, without air), these proteins are not synthesized, an energy-saving feature useful for bacteria growing in the largely anaerobic colon. Most human cells, in contrast, express constant (constitutive) levels of respiratory enzymes and die without oxygen.



**Fig. 16.1.** *E. coli* cell. In prokaryotes, DNA is not separated from the rest of the cellular contents by a nuclear envelope; therefore, simultaneous transcription and translation occur in bacteria. Once a small piece of mRNA is synthesized, ribosomes bind to the mRNA, and translation begins.

# II. REGULATION OF GENE EXPRESSION IN PROKARYOTES

Prokaryotes are single-celled organisms and, therefore, require less complex regulatory mechanisms than the multicellular eukaryotes (Fig. 16.1). The most extensively studied prokaryote is the bacterium *Escherichia coli*, an organism that thrives in the human colon, usually enjoying a symbiotic relationship with its host. Based on the size of its genome ( $4 \times 10^6$  base pairs), *E. coli* should be capable of making several thousand proteins. However, under normal growth conditions, they synthesize only about 600 to 800 different proteins. Obviously, many genes are inactive, and only those genes are expressed that generate the proteins required for growth in that particular environment.

All *E. coli* cells of the same strain are morphologically similar and contain an identical circular chromosome (see Fig. 16.1). As in other prokaryotes, DNA is not complexed with histones, no nuclear envelope separates the genes from the contents of the cytoplasm, and gene transcripts do not contain introns. In fact, as mRNA is being synthesized, ribosomes bind and begin to produce proteins, so that transcription and translation occur simultaneously. The mRNAs in *E. coli* have very short half-lives and are degraded within a few minutes so that an mRNA must be constantly generated from transcription to maintain synthesis of its proteins. Thus, regulation of transcription, principally at the level of initiation, is sufficient to regulate the level of proteins within the cell.

# A. Operons

The genes encoding proteins are called structural genes. In the bacterial genome, the structural genes for proteins involved in performing a related function (such as the enzymes of a biosynthetic pathway) are often grouped sequentially into units called operons (Fig. 16.2). The genes in an operon are coordinately expressed; that is, they are either all "turned on" or all "turned off." When an operon is expressed, all of its genes are transcribed. A single polycistronic mRNA is produced that codes for all the proteins of the operon. This polycistronic mRNA contains multiple sets of start and stop codons that allow a number of different proteins to be produced from this single transcript at the translational level. Transcription of the genes in an operon is regulated by its promoter, which is located in the operon at the 5'-end, upstream from the structural genes.

# B. Regulation of RNA Polymerase Binding by Repressors

In bacteria, the principle means of regulating gene transcription is through repressors, which are regulatory proteins that prevent the binding of RNA polymerase to the promoter and, thus, act on initiation of transcription (Fig. 16.3). In general,



**Fig. 16.2.** An operon. The structural genes of an operon are transcribed as one long polycistronic mRNA. During translation, different start (AUG) and stop (shown in blue) codons lead to a number of distinct proteins being produced from this single mRNA.

regulatory mechanisms such as repressors that work through inhibition of gene transcription are referred to as negative control, and mechanisms that work through stimulation of gene transcription are called positive control.

The repressor is encoded by a regulatory gene (see Fig. 16.3). Although this gene is considered part of the operon, it is not always located near the remainder of the operon. Its product, the repressor protein, diffuses to the promoter and binds to a region of the operator called the operator. The operator is located within the promoter or near its 3'-end, just upstream from the transcription startpoint. When a repressor is bound to the operator, the operon is not transcribed because the repressor blocks the binding of RNA polymerase to the promoter. Two regulatory mechanisms work through controlling repressors: induction (an inducer inactivates the repressor), and repression (a co-repressor is required to activate the repressor).

#### 1. **INDUCERS**

Induction involves a small molecule, known as an inducer, which stimulates expression of the operon by binding to the repressor and changing its conformation so that it can no longer bind to the operator (Fig. 16.4). The inducer is either a nutrient or a metabolite of the nutrient. In the presence of the inducer, RNA polymerase can therefore bind to the promoter and transcribe the operon. The key to this mechanism is that in the absence of the inducer, the repressor is active, transcription is repressed, and the genes of the operon are not expressed.

Consider, for example, induction of the lac operon of E. coli by lactose (Fig. 16.5). The enzymes for metabolizing glucose by glycolysis are produced constitutively; that is, they are constantly being made. If the milk sugar lactose is available, the cells adapt and begin to produce the three additional enzymes required for lactose metabolism, which are encoded by the *lac* operon. A metabolite of lactose (allolactose) serves as an inducer, binding to the repressor and inactivating it. Because the inactive repressor no longer binds to the operator, RNA polymerase can bind to the promoter and transcribe the structural genes of the lac operon, producing a polycistronic mRNA that encodes for the three additional proteins. However, the presence of glucose can prevent activation of the lac operon (see "Stimulation of RNA polymerase binding," below).

#### 2. **COREPRESSORS**

In a regulatory model called repression, the repressor is inactive until a small molecule called a corepressor (a nutrient or its metabolite) binds to the repressor, activating it (Fig. 16.6). The repressor-corepressor complex then binds to the operator, preventing binding of RNA polymerase and gene transcription. Consider, for example, the trp operon, which encodes the five enzymes required for the synthesis of the amino acid tryptophan. When tryptophan is available, E. coli cells save energy by no longer making these enzymes. Tryptophan is a corepressor that binds to the inactive repressor, causing it to change conformation and bind to the operator, thereby inhibiting transcription of the operon. Thus, in the repression model, the repressor is inactive without a corepressor; in the induction model, the repressor is active unless an inducer is present.



If one of the lac operon enzymes induced by lactose is lactose permease (which increases lactose entry into the cell), how does lactose initially get into the cell to induce these enzymes? A small amount of the permease exists even in the absence of lactose, and a few molecules of lactose enter the cell and are metabolized to allolactose, which begins the process of inducing the operon. As the amount of the permease increases, more lactose can be transported into the cell.



Fig. 16.3. Regulation of operons by repressors. When the repressor protein is bound to the operator, RNA polymerase cannot bind, and transcription therefore does not occur.



Fig. 16.4. An inducible operon. In the absence of an inducer, the repressor binds to the operator, preventing the binding of RNA polymerase. When the inducer is present, the inducer binds to the repressor, inactivating it. The inactive repressor no longer binds to the operator. Therefore, RNA polymerase can bind to the promoter region and transcribe the structural genes.



**Fig. 16.5.** The protein products of the *lac* operon. Lactose is a disaccharide that is hydrolyzed to glucose and galactose by  $\beta$ -galactosidease (the Z gene). Both glucose and galactose can be oxidized by the cell for energy. The permease (Y gene) enables the cell to take up lactose more readily. The A gene produces a transacetylase that acetylates  $\beta$ -galactosides. The function of this acetylation is not clear. The promoter binds RNA polymerase and the operator binds a repressor protein. Lactose is converted to allolactose, an inducer that binds the repressor protein and prevents it from binding to the operator. Transcription of the lac operon also requires activator proteins that are inactive when glucose levels are high.

# C. Stimulation of RNA Polymerase Binding

In addition to regulating transcription by means of repressors that inhibit RNA polymerase binding to promoters (negative control), bacteria regulate transcription by means of activating proteins that bind to the promoter and stimulate the binding of RNA polymerase (positive control). Transcription of the *lac* operon, for example, can be induced by allolactose only if glucose is absent. The presence or absence of glucose is communicated to the promoter by a regulatory protein named the cyclic adenosine monophosphate (cAMP) receptor protein (CRP)



Fig. 16.6. A repressible operon. The repressor is inactive until a small molecule, the corepressor, binds to it. The repressor–corepressor complex binds to the operator and prevents transcription.



#### A. In the presence of lactose and glucose



**Fig. 16.7.** Catabolite repression of stimulatory proteins. The *lac* operon is used as an example. **A.** The inducer allolactose (a metabolite of lactose) inactivates the repressor. However, because of the absence of the required coactivator, cAMP-CRP, no transcription occurs unless glucose is absent. **B.** In the absence of glucose, cAMP levels rise. cAMP forms a complex with the cAMP receptor protein (CRP). The binding of the cAMP–CRP complex to a regulatory region of the operon permits the binding of RNA polymerase to the promoter. Now the operon is transcribed, and the proteins are produced.

(Fig. 16.7). This regulatory protein is also called a catabolite activator protein (CAP). A decrease in glucose levels increases levels of the intracellular second messenger cAMP by a mechanism that is not well understood. cAMP binds to CRP, and the cAMP-CRP complex binds to a regulatory region of the operon, stimulating binding of RNA polymerase to the promoter and transcription. When glucose is present, cAMP levels decrease, CRP assumes an inactive conformation that does not bind to the operon, and transcription is inhibited. Thus, the enzymes encoded by the *lac* operon are not produced if cells have an adequate supply of glucose, even if lactose is present at very high levels.

# D. Regulation of RNA Polymerase Binding by Sigma Factors

*E. coli* has only one RNA polymerase. Sigma factors bind to this RNA polymerase, stimulating its binding to certain sets of promoters, thus simultaneously activating transcription of several operons. The standard sigma factor in *E. coli* is

Nutrient regulation of gene expression may occur through the nutrient itself, or through a metabolite of the nutrient synthesized inside the cell. Sometimes both the nutrient and its metabolite are grouped into the term "catabolite," as in "catabolite repression."  $\sigma^{70}$ , a protein with a molecular weight of 70,000 daltons (see Chapter 14). Other sigma factors also exist. For example,  $\sigma^{32}$  helps RNA polymerase recognize promoters for the different operons that encode the "heat shock" proteins. Thus, increased transcription of the genes for heat shock proteins, which prevent protein denaturation at high temperatures, occurs in response to elevated temperatures.

# E. Attenuation of Transcription

Some operons are regulated by a process that interrupts (attenuates) transcription after it has been initiated (Fig. 16.8). For example, high levels of tryptophan attenuate transcription of the E. coli *trp* operon, as well as repress its transcription. As mRNA is being transcribed from the *trp* operon, ribosomes bind and rapidly begin to translate the transcript. Near the 5'-end of the transcript, there are a number of codons for tryptophan. Initially, high levels of tryptophan in the cell result in high levels of trp-tRNA<sup>trp</sup> and rapid translation of the transcript. However, rapid translation generates a hairpin loop in the mRNA that serves as a termination signal for RNA polymerase, and transcription terminates. Conversely, when tryptophan levels are low, levels of trp-tRNA<sup>trp</sup> are low, and ribosomes stall at codons for tryptophan. A different hairpin loop forms in the mRNA that does not terminate transcription, and the complete mRNA is transcribed.



**Fig. 16.8.** Attenuation of the *trp* operon. Sequences 2, 3, and 4 in the mRNA transcript can form base pairs (2 with 3 or 3 with 4) that generate hairpin loops. When tryptophan levels are low, the ribosome stalls at the adjacent trp codons in sequence 1, the 2–3 loop forms, and transcription continues. When tryptophan levels are high, translation is rapid and the ribosome blocks formation of the 2–3 loop. Under these conditions, the 3–4 loop forms and terminates transcription.

The tryptophan, histidine, leucine, phenylalanine, and threonine operons are regulated by attenuation. Repressors and activators also act on the promoters of some of these operons, allowing the levels of these amino acids to be very carefully and rapidly regulated.

# III. REGULATION OF PROTEIN SYNTHESIS IN EUKARYOTES

Multicellular eukaryotes are much more complex than single-celled prokaryotes. As the human embryo develops into a multicellular organism, different sets of genes are turned on, and different groups of proteins are produced, resulting in differentiation into morphologically distinct cell types able to perform different functions. Even beyond the reproductive age, certain cells within the organism continue to differentiate, such as those that produce antibodies in response to an infection, renew the population of red blood cells, and replace digestive cells that have been sloughed into the intestinal lumen. All of these physiologic changes are dictated by complex alterations in gene expression.

# A. Regulation of Eukaryotic Gene Expression at Multiple Levels

Differences between eukaryotic and prokaryotic cells result in different mechanisms for regulating gene expression. DNA in eukaryotes is organized into the nucleosomes of chromatin, and genes must be in an active structure to be expressed in a cell. Furthermore, operons are not present in eukaryotes, and the genes encoding proteins that function together are usually located on different chromosomes. Thus, each gene needs its own promoter. In addition, the processes of transcription and translation are separated in eukaryotes by intracellular compartmentation (nucleus and cytosol, or endoplasmic reticulum [ER]) and by time (eukaryotic hnRNA must be processed and translocated out of the nucleus before it is translated). Thus, regulation of eukaryotic gene expression occurs at multiple levels:

- DNA and the chromosome, including chromosome remodeling and gene rearrangement
- Transcription, primarily through transcription factors affecting binding of RNA polymerase
- Processing of transcripts
- Initiation of translation and stability of mRNA

Once a gene is activated through chromatin remodeling, the major mechanism of regulating expression affects initiation of transcription at the promoter.

# B. Regulation of Availability of Genes for Transcription

Once a haploid sperm and egg combine to form a diploid cell, the number of genes in human cells remains approximately the same. As cells differentiate, different genes are available for transcription. A typical nucleus contains chromatin that is condensed (heterochromatin) and chromatin that is diffuse (euchromatin)(Fig. 16.9) (see Chapter 12). The genes in heterochromatin are inactive, whereas those in euchromatin produce mRNA. Long-term changes in the activity of genes occur during development as chromatin goes from a diffuse to a condensed state or vice versa.

The cellular genome is packaged together with histones into nucleosomes, and initiation of transcription is prevented if the promoter region is part of a nucleosome. Thus, activation of a gene for transcription requires changes in the state of the chromatin, called chromatin remodeling. The availability of genes for transcription The globin chains of hemoglobin provide an example of functionally related proteins that are on different chromosomes. The gene for the  $\alpha$ -globin chain is on chromosome 16, whereas the gene for the  $\beta$ -globin chain is on chromosome 11. As a consequence of this spatial separation, each gene must have its own promoter. This situation is different from that of bacteria, in which genes encoding proteins that function together are often sequentially arranged in operons controlled by a single promoter.



**Fig. 16.9.** Inactivation of genes during development of red blood cells. Diffuse chromatin (euchromatin) is active in RNA synthesis. Condensed chromatin (heterochromatin) is inactive. As red blood cell precursors mature, their chromatin becomes more condensed. Eventually, the nucleus is extruded.



**Fig. 16.10.** Histone acetylation. Abbreviations: HAC, histone acetylase; HDAC, histone deacetylase.

also can be affected in certain cells, or under certain circumstances, by gene rearrangements, amplification, or deletion. For example, during lymphocyte maturation, genes are rearranged to produce a variety of different antibodies.

#### 1. CHROMATIN REMODELING

The remodeling of chromatin generally refers to displacement of the nucleosome from specific DNA sequences so that transcription of the genes in that sequence can be initiated. This occurs through two different mechanisms. The first mechanism is by an adenosine triphosphate (ATP)-driven chromatin remodeling complex, which uses energy from ATP hydrolysis to unwind certain sections of DNA from the nucleosome core. The second mechanism is by covalent modification of the histone tails through acetylation (Fig. 16.10). Histone acetyltransferases (HAT) transfer an acetyl group from acetyl CoA to lysine residues in the tail (the amino terminal ends of histones H2A, H2B, H3, and H4). This reaction removes a positive charge from the  $\epsilon$ -amino group of the lysine, thereby reducing the electrostatic interactions between the histones and the negatively charged DNA, making it easier for DNA to unwind from the histones. The acetyl groups can be removed by histone deacetylases (HDAC). Each histone has a number of lysine residues that may be acetylated and, through a complex mixing of acetylated and nonacetylated sites, different segments of DNA can be freed from the nucleosome. A number of transcription factors and co-activators also contain histone acetylase activity, which facilitates the binding of these factors to the DNA and simultaneous activation of the gene and initiation of its transcription.

### 2. METHYLATION OF DNA

Cytosine residues in DNA can be methylated to produce 5-methylcytosine. The methylated cytosines are located in GC-rich sequences (called GC-islands), which are often near or in the promoter region of a gene. In certain instances, genes that are methylated are less readily transcribed than those that are not methylated. For example, globin genes are more extensively methylated in nonerythroid cells (cells which are not a part of the erythroid, or red blood cell, lineage) than in the cells in which these genes are expressed (such as the erythroblast and reticulocyte). Methylation is a mechanism for regulating gene expression during differentiation, particularly in fetal development.

Methylation has been implicated in genomic imprinting, a process occurring during the formation of the eggs or sperm that blocks the expression of the gene in the fertilized egg. Males methylate a different set of genes than females. This sex-dependent differential methylation has been most extensively studied in two human disorders, Prader-Willi syndrome and Angelman syndrome. Both syndromes, which have very different symptoms, result from deletions of the same region of chromosome 15 (a microdeletion of less than 5 megabases in size). If the deletion is inherited from the father, Prader-Willi syndrome is seen in the child; if the deletion is inherited from the mother, Angelman's syndrome is observed. A disease occurs when a gene that is in the deleted region of one chromosome is methylated on the other chromosome. The mother methylates different genes than the father, so different genes are expressed depending on which parent transmitted the intact chromosome. For example, if genes 1, 2, and 3 are deleted in the paternal chromosome in the Prader-Willi syndrome, genes 1, 2 and 3 are deleted on the maternal chromosome and gene 1 is methylated on the paternal chromosome, only genes 2 and 3 would be expressed.



#### 3. GENE REARRANGEMENT

Segments of DNA can move from one location to another in the genome, associating with each other in various ways so that different proteins are produced (Fig. 16.11). The most thoroughly studied example of gene rearrangement occurs in cells that produce antibodies. Antibodies contain two light chains and two heavy chains, each of which contains both a variable and a constant region (see Chapter 7, section V.B, Fig. 7.19). Cells called B cells make antibodies. In the precursors of B cells, hundreds of V<sub>H</sub> sequences, approximately 20 D<sub>H</sub> sequences, and approximately 6 J<sub>H</sub> sequences are located in clusters within a long region of the chromosome. During the production of the immature B cells, a series of recombinational events occur that join one  $V_H$ , one  $D_H$ , and one  $J_H$  sequence into a single exon. This now encodes the variable region of the heavy chain of the antibody. Given the large number of immature B cells that are produced, virtually every recombinational possibility occurs, such that all VDJ combinations are represented within this cell population. Later in development, during differentiation of mature B cells, recombinational events join a VDJ sequence to one of the nine heavy chain elements. When the immune system encounters an antigen, the one immature B cell that can bind to that antigen (because of its unique manner in forming the VDJ exon) is stimulated to proliferate (clonal expansion) and to produce antibodies against the antigen.

#### 4. GENE AMPLIFICATION

Gene amplification is not the usual physiologic means of regulating gene expression in normal cells, but it does occur in response to certain stimuli if the cell can obtain a growth advantage by producing large amounts of a protein. In gene amplification, certain regions of a chromosome undergo repeated cycles of DNA replication. The newly synthesized DNA is excised and forms small, unstable chromosomes called "double minutes." The double minutes integrate into other chromosomes throughout the genome, thereby amplifying the gene in the process. Normally, gene amplification occurs through errors during DNA replication and cell division and, if the environmental conditions are correct, cells containing amplified genes may have a growth advantage over those without the amplification.



In fragile X syndrome, a GCC triplet is amplified on the 5'-side of a gene (FMR-1) associated with the disease. This gene is located on the X chromosome. The disease is named for the finding that in the absence of folic acid (which impairs nucleotide production and hence, the replication of DNA) the X chromosome develops single and double-stranded breaks in its DNA. These were termed fragile sites. It was subsequently determined that the FMR-1 gene was located in one of these fragile sites. A normal person has about 30 copies of the GCC triplet, but in affected individuals, thousands of copies can be present. This syndrome, which is a common form of inherited mental retardation, affects about 1 in 1,250 males and 1 in 2,000 females.



Although rearrangements of short DNA sequences are difficult to detect, microscopists have observed major rearrangements for many years. Such major rearrangements, known as translocations, can be observed in metaphase chromosomes under the microscope.

Mannie Weitzels has such a translocation, known as the Philadelphia chromosome because it was first observed in that city. The Philadelphia chromosome is produced by a balanced exchange between chromosomes 9 and 22.

Arlyn Foma has been treated with a combination of drugs that includes methotrexate, a drug that inhibits cell proliferation by inhibiting dihydrofolate reductase. Dihydrofolate reductase reduces dihydrofolate to tetrahydrofolate, a cofactor required for synthesis of thymine and purine nucleotides. Because Arlyn Foma has not been responding well, the possibility that he has become resistant to methotrexate was considered. Sometimes, rapidly dividing cancer cells treated with methotrexate amplify the gene for dihydrofolate reductase, producing hundreds of copies in the genome. These cells generate large amounts of difhydrofolate reductase, and normal doses of methotrexate are no longer adequate. Gene amplification is one of the mechanisms by which patients become resistant to a drug.



Fig. 16.11. Rearrangement of DNA. The heavy chain gene from which lymphocytes produce immunoglobulins is generated by combining specific segments from among a large number of potential sequences in the DNA of precursor cells. The variable and constant regions of immunoglobulins (antibodies) are described in Chapter 7.

#### GENE DELETIONS 5.

With a few exceptions, the deletion of genetic material is likewise not a normal means of controlling transcription, although such deletions do result in disease. Gene deletions can occur through errors in DNA replication and cell division and are usually only noticed if a disease results. For example, various types of cancers result from the loss of a good copy of a tumor suppressor gene, leaving the cell with a mutated copy of the gene (see Chapter 18).

# C. Regulation at the Level of Transcription

The transcription of active genes is regulated by controlling assembly of the basal transcription complex containing RNA polymerase and its binding to the TATA box of the promoter (see Chapter 14). The basal transcription complex contains the TATA binding protein (TBP, a component of TFIID) and other proteins called general (basal) transcription factors (such as TFIIA, etc.) that form a complex with RNA polymerase II. Additional transcription factors that are ubiquitous to all promoters bind upstream at various sites in the promoter region. They increase the frequency of transcription and are required for a promoter to function at an adequate level. Genes that are regulated solely by these consensus elements in the promoter region are said to be constitutively expressed.

The control region of a gene also contains DNA regulatory sequences that are specific for that gene and may increase its transcription 1,000-fold or more (Fig. 16.12). Gene-specific transcription factors (also called transactivators or activators) bind to these regulatory sequences and interact with a mediator protein, such as a coactivator. By forming a loop in the DNA, coactivators interact with the basal transcription complex and can activate its assembly at the initiation site on the promoter. These DNA regulatory sequences might be some distance from the promoter and may be either upstream or downstream of the initiation site.

#### 1. GENE-SPECIFIC REGULATORY PROTEINS

The regulatory proteins that bind directly to DNA sequences are most often called transcription factors or gene-specific transcription factors (if it is necessary to distinguish them from the general transcription factors of the basal transcription complex). They also can be called activators (or transactivators), inducers, repressors, or nuclear receptors. In addition to their DNA-binding domain, these proteins usually have a domain that binds to mediator proteins (coactivators, corepressors, or TATA binding protein associated factors-TAFs). Coactivators, corepressors, and other mediator proteins do not bind directly to DNA but generally bind to components of the basal transcription complex and mediate its assembly at the promoter. They can be specific for a given gene transcription factor or general and bind many different gene-specific transcription factors. Certain co-activators have histone acetylase activity, and certain corepressors have histone deacetylase activity. When the appropriate interactions between the transactivators, coactivators, and the basal transcription complex occur, the rate of transcription of the gene is increased (induction).

Some regulatory DNA binding proteins inhibit (repress) transcription and may be called repressors. Repression can occur in a number of ways. A repressor bound to its specific DNA sequence may inhibit binding of an activator to its regulatory sequence. Alternately, the repressor may bind a corepressor that inhibits binding of a coactivator to the basal transcription complex. The repressor may directly bind a component of the basal transcription complex. Some steroid hormone receptors that are transcription factors bind either coactivators or corepressors, depending on whether the receptor contains bound hormone. Furthermore, a particular transcription factor may induce transcription when bound to the regulatory sequence of one gene and may repress transcription when bound to the regulatory sequence of another gene.



The terminology used to describe components of gene-specific regulation varies somewhat, depending on the system. For example, in the original

terminology, DNA regulatory sequences called enhancers bound transactivators, which bound coactivators. Similarly, silencers bound corepressors. Hormones bound to hormone receptors, which bound to hormone response elements in DNA. Although these terms are still used, they are often replaced with more general terms such as "DNA regulatory sequences" and "specific transcription factors," in recognition of the fact that many transcription factors activate one gene while inhibiting another or that a specific transcription factor may be changed from a repressor to an activator by phosphorylation.

Histone acetylase activity has been associated with a number of transcription factors and co-activators. The proteins ACTR (activator of the thyroid and retinoic acid receptor) and SRC-1 (steroid receptor co-activator) are involved in activation of transcription by several ligand-bound nuclear receptors, and both contain histone acetylase activity. TAF250, a component of TFIID, also contains histone acetylase activity, as does the co-activator p300/CBP (CREB binding protein), which interacts with the transcription factor CREB.



**Fig. 16.12.** The gene regulatory control region consists of the promoter region and additional gene regulatory sequences, including enhancers and hormone response elements (shown in blue). Gene regulatory proteins that bind directly to DNA (regulatory DNA binding proteins) are usually called specific transcription factors or transactivators; they may be either activators or repressors of the transcription of specific genes. The specific transcription factors bind mediator proteins (co-activators or corepressors) that interact with the general transcription factors of the basal transcription complex. The basal transcription complex contains RNA polymerase and associated general transcription factors (TFII factors) and binds to the TATA box of the promoter, initiating gene transcription.

### 2. TRANSCRIPTION FACTORS THAT ARE STEROID HORMONE/THYROID HORMONE RECEPTORS

In the human, steroid hormones and other lipophilic hormones activate or inhibit transcription of specific genes through binding to nuclear receptors that are gene-specific transcription factors (Fig. 16.13A). The nuclear receptors bind to DNA regulatory sequences called hormone response elements and induce or repress transcription of target genes. The receptors contain a hormone (ligand) binding domain, a DNA binding domain, and a dimerization domain that permits two receptor molecules to bind to each other, forming characteristic homodimers or heterodimers. A transactivation domain binds the coactivator proteins that interact with the basal transcription complex. The receptors also contain a nuclear localization signal domain that directs them to the nucleus at various times after they are synthesized.

Various members of the steroid hormone/thyroid hormone receptor family work in different ways. The glucocorticoid receptor, which binds the steroid hormone cortisol, resides principally in the cytosol bound to heat shock proteins. As cortisol binds, the receptor dissociates from the heat shock proteins, exposing the nuclear localization signal (see Fig. 16.13B). The receptors form homodimers that are translocated to the nucleus, where they bind to the hormone response elements (glucocorticoid response elements–GRE) in the DNA control region of certain genes. The transactivation domains of the receptor dimers bind mediator proteins, thereby activating transcription of specific genes and inhibiting transcription of others. In a condition known as testicular feminization, patients produce androgens (the male sex steroids), but target cells fail to respond to these steroid hormones because they lack the appropriate intracellular transcription factor receptors. Therefore, the transcription of the genes responsible for masculinization is not activated. A patient with this condition has an XY (male) karyotype (set of chromosomes) but looks like a female. External male genitalia do not develop, but testes are present, usually in the inguinal region.



B. Transcriptional regulation by steroid hormone receptors



**Fig. 16.13.** Steroid hormone receptors. A. Domains of the steroid hormone receptor. The transactivation domain (TAD) binds coactivators; DNAbinding domain (DBD) binds to hormone response element in DNA; ligand-binding domain (LBD) binds hormone; NLS is the nuclear localization signal; the dimerization sites are the portions of the protein involved in forming a dimer. The inhibitor binding site binds heat shock proteins and masks the nuclear localization signal. B. Transcriptional regulation by steroid hormone receptors. Additional abbreviations: HSP, heat shock proteins; GRE, glucocorticoid response element; GIZ, glucocorticoid receptor.

Other members of the steroid hormone/thyroid hormone family of receptors are also gene-specific transactivation factors but generally form heterodimers that constitutively bind to a DNA regulatory sequence in the absence of their hormone ligand and repress gene transcription (Fig. 16.14). For example, the thyroid hormone receptor forms a heterodimer with the retinoid X receptor (RXR) that binds to thyroid hormone response elements and to corepressors (including one with deacetylase activity), thereby inhibiting expression of certain genes. When thyroid hormone binds, the receptor dimer changes conformation, and the transactivation domain binds coactivators, thereby initiating transcription of the genes.

The RXR receptor, which binds the retinoid 9-cis retinoic acid, can form heterodimers with at least eight other nuclear receptors. Each heterodimer has a different DNA binding specificity. This allows the RXR to participate in the regulation of a wide variety of genes, and to regulate gene expression differently, depending on the availability of other active receptors.

### 3. STRUCTURE OF DNA BINDING PROTEINS

Several unique structural motifs have been characterized for specific transcription factors. Each of these proteins has a distinct recognition site (DNA binding domain) that binds to the bases of a specific sequence of nucleotides in DNA. Four of the best-characterized structural motifs are zinc fingers, b-zip proteins (including leucine zippers), helix-turn-helix, and helix-loop-helix.

To regulate gene transcription, two estrogen receptors combine to form a dimer that binds to a palindrome in the promoter region of certain genes (see Figs. 16.12 and 16.13). A palindrome is a sequence of bases that is identical on the antiparallel strand when read in the opposite direction. For example, the sequence ATCGCGAT base-pairs to form the sequence TAGCGCTA, which when read in the opposite direction is ATCGCGAG. Each estrogen receptor is approximately 73 amino acids long and contains two zinc fingers. Each zinc is chelated to two cysteines in an  $\alpha$ -helix and two cysteines in a  $\beta$ -sheet region. The position of the nucleotide recognition sequence in an  $\alpha$ -helix keeps the sequence in a relatively rigid conformation as it fits into the major groove of DNA. The zinc finger that lies closest to the carboxyl terminal is involved in dimerization with the second estrogen receptor, thus inverting the nucleotide recognition sequence to match the other half of the palindrome. The dimer-palindrome requirement enormously enhances the specificity of binding, and, consequently, only certain genes are affected.

Zinc finger motifs (commonly found in the DNA binding domain of steroid hormone receptors) contain a bound zinc chelated at four positions with either histidine or cysteine in a sequence of approximately 20 amino acids (Fig. 16.15). The result is a relatively small, tight, autonomously-folded domain. The zinc is required to maintain the tertiary structure of this domain. Eukaryotic transcription factors generally have two to six zinc finger motifs that function independently. At least one of the zinc fingers forms an  $\alpha$ -helix containing a nucleotide recognition signal, a sequence of amino acids that specifically fits into the major groove of DNA (Fig. 16.16A).

Leucine zippers also function as dimers to regulate gene transcription (see Fig. 16.16B). The leucine zipper motif is an  $\alpha$ -helix of 30 to 40 amino acid residues that contains a leucine every seven amino acids, positioned so that they align on the same side of the helix. Two helices dimerize so that the leucines of one helix align with the

A wide variety of transcription factors contain the zinc finger motif, including the steroid hormone receptors, such as the estrogen and the glucocorticoid receptor. Other transcription factors that contain zinc finger motifs include Sp1 and polymerase III transcription factor TFIIIA (part of the basal transcription complex), which has nine zinc finger motifs.



**Fig. 16.14.** Activity of the thyroid hormone receptor–retinoid receptor dimer (TR-RXR) in the presence and absence of thyroid hormone ( $T_3$ ). Abbrev: HAC, histone acetylase; HDAC, histone deacetylase.



Leucine zipper transcription factors function as homodimers or heterodimers. For example, AP1 is a heterodimer whose subunits are encoded by the genes *fos* and *jun.* 



Fig. 16.15. Zinc fingers of the estrogen receptor. In each of the two zinc fingers, one zinc ion is coordinated with four cysteine residues, shown in blue. The region labeled  $\alpha$ -helix with NRS forms an  $\alpha$ -helix that contains a nucleotide recognition signal (NRS). This signal consists of a sequence of amino acid residues that bind to a specific base sequence in the major groove of DNA. Regions enclosed in boxed arrows participate in rigid helices.



**Fig. 16.16.** Interaction of DNA binding proteins with DNA. (A) Zinc finger motifs consist of an  $\alpha$ -helix and a  $\beta$  sheet in which 4 cysteine and/or histidine residues coordinately bind a zinc ion. The nucleotide recognition signal (contained within the  $\alpha$ -helix) of at least one zinc finger binds to a specific sequence of bases in the major groove of DNA. (B) Leucine zipper motifs form from two distinct polypeptide chains. Each polypeptide contains a helical region in which leucine residues are exposed on one side. These leucines form hydrophobic interactions with each other, causing dimerization. The remaining helices interact with DNA. (C) Helix-turn-helix motifs contain three (or sometimes four) helical regions, one of which binds to DNA, whereas the others lie on top and stabilize the interaction. (D) Helix-loop-helix motifs contain helical regions that bind to DNA like leucine zippers. However, their dimerization domains consist of two helices, each of which is connected to its DNA binding helix by a loop.

The helix-turn-helix motif is found in homeodomain proteins (proteins that play critical roles in the regulation of gene expression during development).



Many of the transcription factors containing the helix-loop-helix motif are involved in cellular differ-

entiation (such as myogenin in skeletal muscle, neurogenin in neurogenesis and SCL/tal-1 in hematopoeisis, blood cell development). other helix through hydrophobic interactions to form a coiled coil. The portions of the dimer adjacent to the zipper "grip" the DNA through basic amino acid residues (arginine and lysine) that bind to the negatively charged phosphate groups. This DNA binding portion of the molecule also contains a nucleotide recognition signal.

In the helix-turn-helix motif, one helix fits into the major groove of DNA, making most of the DNA binding contacts (see Fig. 16.16C). It is joined to a segment containing two additional helices that lie across the DNA binding helix at right angles. Thus, a very stable structure is obtained without dimerization.

Helix-loop-helix transcription factors are a fourth structural type of DNA binding protein (see Fig. 16.16D). They also function as dimers that fit around and grip DNA in a manner geometrically similar to leucine zipper proteins. The dimerization region consists of a portion of the DNA-gripping helix and a loop to another helix. Like leucine zippers, helix-loop-helix factors can function as either hetero or homodimers. These factors also contain regions of basic amino acids near the amino terminus and are also called basic helix-loop-helix (bHLH) proteins.

### 4. REGULATION OF TRANSCRIPTION FACTORS

The activity of gene-specific transcription factors is regulated in a number of different ways. Because transcription factors need to interact with a variety of coactivators to stimulate transcription, the availability of coactivators or other mediator proteins is critical for transcription factor function. If a cell upregulates or downregulates its synthesis of coactivators, the rate of transcription can also be increased or decreased. Transcription factor activity can be modulated by changes in the amount of transcription factor synthesized (see section 5, "Multiple regulators of promoters"), by binding a stimulatory or inhibitory ligand (such as steroid hormone binding to the steroid hormone receptors), and by stimulation of nuclear entry (illustrated by the glucocorticoid receptor). The ability of a transcription factor to influence the transcription of a gene is also augmented or antagonized by the presence of other transcription factors. For example, the thyroid hormone receptor is critically dependent on the concentration of the retinoid receptor to provide a dimer partner. Another example is provided by the phosphoenol pyruvate (PEP) carboxykinase gene, which is induced or repressed by a variety of hormone-activated transcription factors (see section 5). Frequently, transcription factor activity is regulated through phosphorylation.

Growth factors, cytokines, polypeptide hormones, and a number of other signal molecules regulate gene transcription through phosphorylation of specific transcription factors by receptor kinases.

For example, STAT proteins are transcription factors phosphorylated by JAK-STAT receptors, and SMAD proteins are transcription factors phosphorylated by serine-threonine kinase receptors such as the transforming growth factor- $\beta$ (TGF- $\beta$ ) receptor (see Chapter 11).

Nonreceptor kinases, such as protein kinase A, also regulate transcription factors through phosphorylation. Many hormones generate the second messenger **cAMP**, which activates protein kinase A. Activated protein kinase A enters the nucleus and phosphorylates the transcription factor CREB (cAMP response element binding protein). CREB is constitutively bound to the DNA response element CRE (cAMP response element) and is activated by phosphorylation. Other hormone signaling pathways, such as the MAP kinase pathway, also phosphorylate CREB (as well as many other transcription factors).

#### 5. MULTIPLE REGULATORS OF PROMOTERS

The same transcription factor inducer can activate transcription of many different genes if the genes each contain a common response element. Furthermore, a single inducer can activate sets of genes in an orderly, programmed manner (Fig. 16.17). The inducer initially activates one set of genes. One of the protein products of this set of genes can then act as a specific transcription factor for another set of genes. If this process is repeated, the net result is that one inducer can set off a series of events that result in the activation of many different sets of genes.



**Fig. 16.17.** 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.

Interferons, cytokines produced by cells that have been infected with a virus, bind to the JAK-STAT family of cell surface receptors. When an interferon binds, JAK (a receptor-associated tyrosine kinase) phosphorylates STAT transcription factors bound to the receptors (see Chap. 11). The phosphorylated STAT proteins are released, dimerize, enter the nucleus, and bind to specific gene regulatory sequences. Different combinations of phosphorylated STAT proteins bind to different sequences and activate transcription of a different set of genes. One of the genes activated by interferon produces the oligonucleotide 2'-5'oligo(A), which is an activator of a ribonuclease. This RNase degrades mRNA, thus inhibiting synthesis of the viral proteins required for its replication.

In addition to antiviral effects, interferons have antitumor effects. The mechanisms of the antitumor effects are not well understood, but are probably likewise related to stimulation of specific gene expression by STAT proteins. Interferon- $\alpha$ , produced by recombinant DNA technology, has been used to treat patients such as **Arlyn Foma** who have certain types of nodular lymphomas and patients, such as **Mannie Weitzels**, who have chronic myelogenous leukemia.

An example of a transcriptional cascade of gene activation is observed during adipocyte (fat cell) differentiation. Fibroblast-like cells can be induced to form adipocytes by the addition of dexamethasome (a steroid hormone) and insulin to the cells. These factors induce the transient expression of two similar transcription factors named C/EPB $\beta$  and C/EPB $\gamma$ . The names stand for CCAAT enhancer binding protein, and  $\beta$  and  $\gamma$  are two forms of these factors which recognize CCAAT sequences in DNA. The C/EPB transcription factors then induce the synthesis of yet another transcription factor, named the peroxisome proliferator-activated receptor  $\gamma$  (PPAR $\gamma$ ), which forms heterodimers with RXR to regulate the expression of two more transcription factors, C/EPB  $\alpha$  and STAT5. The combination of PPAR $\gamma$ , STAT5, and C/EPB $\alpha$  then leads to the expression of adipocyte-specific genes.

The enzyme PEP carboxykinase (PEPCK) is required for the liver to produce glucose from amino acids and lactate. Ann O'Rexia, who has an eating disorder, needs to maintain a certain blood glucose level to keep her brain functioning normally. When her blood glucose levels drop, cortisol (a glucocorticoid) and glucagon (a polypeptide hormone) are released. In the liver, glucagon increases intracellular cAMP levels, resulting in activation of protein kinase A and subsequent phosphorylation of CREB. Phosphorylated CREB binds to its response element in DNA, as does the cortisol receptor. Both transcription factors enhance transcription of the PEPCK gene (see Figure 16.17). Insulin, which is released when blood glucose levels rise after a meal, can inhibit expression of this gene.



Fig. 16.18. 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 phosphoenolpyruvate carboxykinase (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.

An individual gene contains many different response elements and enhancers, and genes that encode different protein products contain different combinations of response elements and enhancers. Thus, each gene does not have a single, unique protein that regulates its transcription. Rather, as different proteins are stimulated to bind to their specific response elements and enhancers in a given gene, they act cooperatively to regulate expression of that gene (Fig. 16.18). Overall, a relatively small number of response elements and enhancers and a relatively small number of regulatory proteins generate a wide variety of responses from different genes.

### D. Posttranscriptional Processing of RNA

After the gene is transcribed (i.e., posttranscription), regulation can occur during processing of the RNA transcript (hnRNA) into the mature mRNA. The use of alternative splice sites or sites for addition of the poly(A) tail (polyadenylation sites) can result in the production of different mRNAs from a single hnRNA and, consequently, in the production of different proteins from a single gene.

#### 1. ALTERNATIVE SPLICING AND POLYADENYLATION SITES

Processing of the primary transcript involves the addition of a cap to the 5'-end, removal of introns, and polyadenylation (the addition of a poly(A) tail to the 3'-end) to produce the mature mRNA (see Chapter 14). In certain instances, the use of alternative splicing and polyadenylation sites causes different proteins to be produced from the same gene (Fig. 16.19). For example, genes that code for antibodies are regulated by alterations in the splicing and polyadenylation sites, in addition to undergoing gene rearrangement (Fig. 16.20). At an early stage of maturation, pre-B lymphocytes produce IgM antibodies that are bound to the cell membrane. Later, a shorter protein (IgD) is produced that no longer binds to the cell membrane, but rather is secreted from the cell.

#### 2. RNA EDITING

In some instances, RNA is "edited" after transcription. Although the sequence of the gene and the primary transcript (hnRNA) are the same, bases are altered or nucleotides are added or deleted after the transcript is synthesized so that the mature mRNA differs in different tissues (Fig. 16.21).

# E. Regulation at the Level of Translation and the Stability of mRNA

# 1. INITIATION OF TRANSLATION

In eukaryotes, regulation of gene transcription at the level of translation usually involves the initiation of protein synthesis by eIFs (eukaryotic initiation factors), which are regulated through mechanisms involving phosphorylation (see Chapter 15, section V.B.). For example, heme regulates translation of globin mRNA in reticulocytes by controlling the phosphorylation of eIF2 (Fig. 16.22). In reticulocytes (red blood cell precursors), globin is produced when heme levels in the cell are high but not when they are low. Because reticulocytes lack nuclei, globin synthesis must be regulated at the level of translation rather than transcription. Heme acts by preventing phosphorylation of eIF2 by a specific kinase (heme kinase) that is inactive when heme is bound. Thus, when heme levels are high, eIF2 is not phosphorylated and is active, resulting in globin synthesis. Similarly, in other cells, conditions such as starvation, heat shock, or viral infections



**Fig. 16.19.** Alternative splicing of the calcitonin gene produces an mRNA for calcitonin in thyroid cells and an mRNA for CGRP in neurons. In thyroid cells, the pre-mRNA from the calcitonin gene is processed to form an mRNA that codes for calcitonin. Cleavage occurs at poly(A) site 1, and splicing occurs along the blue dashed lines. In the brain, the pre-mRNA of this gene undergoes alternative splicing and polyadenylation to produce calcitonin gene-related protein (CGRP). Cleavage occurs at poly(A) site 2, and splicing occurs along the black dashed lines. CGRP is involved in the sensation of taste.

may result in activation of a specific kinase that phosphorylates eIF2 to an inactive form. Another example is provided by insulin, which stimulates general protein synthesis by activating the phosphorylation of an inhibitor of eIF4E, called 4E-BP. When 4E-BP is phosphorylated, it dissociates, leaving eIF4E in the active form.

A different mechanism for regulation of translation is illustrated by iron regulation of ferritin synthesis (Fig. 16.23). Ferritin, the protein involved in the storage of iron within cells, is synthesized when iron levels increase. The mRNA for ferritin has an iron response element (IRE), consisting of a hairpin loop near its 5'-end, which can bind a regulatory protein called the iron response element binding protein (IRE-BP). When IRE-BP does not contain bound iron, it binds to the IRE and prevents initiation of translation. When iron levels increase and IRE-BP binds iron, it changes to a conformation that can no longer bind to the IRE on the ferritin mRNA. Therefore, the mRNA is translated and ferritin is produced.

### F. Transport and stability of mRNA

Stability of an mRNA also plays a role in regulating gene expression, because mRNAs with long half-lives can generate a greater amount of protein than can those with shorter half-lives. The mRNA of eukaryotes is relatively stable (with half-lives measured in hours to days), although it can be degraded by nucleases in the nucleus or cytoplasm before it is translated. To prevent degradation during transport from the nucleus to the cytoplasm, mRNA is bound to proteins that help to prevent its degradation. Sequences at the 3'-end of the mRNA appear to be involved in determining its half-life and binding proteins that prevent degradation. One of these is the poly(A) tail, which protects the mRNA from attack by nucleases. As mRNA ages, its poly(A) tail becomes shorter.

An example of the role of mRNA degradation in control of translation is provided by the transferrin receptor mRNA (Fig. 16.24) The transferrin receptor is a In addition to stimulating degradation of mRNA, interferon also causes elF2 to become phosphorylated and inactive. This is a second mechanism by which interferon prevents synthesis of viral proteins.



**Fig. 16.20.** Production of a membrane-bound antibody (IgM) and a smaller secreted antibody (IgD) from the same gene. Initially, the lymphocytes produce a long transcript that is cleaved and polyadenylated after the second stop codon. The intron that contains the first stop codon is removed by splicing between the 5'- and 3'-splice sites. Therefore, translation ends at the second stop codon, and the protein contains a hydrophobic exon at its C-terminal end that becomes embedded in the cell membrane. After antigen stimulation, the cells produce a shorter transcript by using a different cleavage and polyadenylation site. This transcript lacks the 3'-splice site for the intron, so the intron is not removed. In this case, translation ends at the first stop codon. The IgD antibody does not contain the hydrophobic region at its C-terminus, so it is secreted from the cell.



Heme (inactive) kinase (inactive) kinase (active) eIF2 active inactive inactive

**Fig. 16.21.** 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 (B-48) is only 48% of the length of the protein produced in the liver (B-100).







**Fig. 16.23.** Translational regulation of ferritin synthesis. The mRNA for ferritin has an iron response element (IRE). When the iron response element binding protein, 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.

protein located in cell membranes that permits cells to take up transferrin, the protein that transports iron in the blood. The rate of synthesis of the transferrin receptor increases when intracellular iron levels are low, enabling cells to take up more iron. Synthesis of the transferrin receptor, like that of the ferritin receptor, is regulated by the binding of the iron response element binding protein (IRE-BP) to the iron response elements (IRE). However, in the case of the transferrin receptor mRNA, the IREs are hairpin loops located at the 3'-end of the mRNA, and not at the 5'end where translation is initiated. When the IRE-BP does not contain bound iron, it has a high affinity for the IRE hairpin loops. Consequently, IRE-BP prevents degradation of the mRNA when iron levels are low, thus permitting synthesis of more transferrin receptor so that the cell can take up more iron. Conversely, when iron levels are elevated, IRE-BP binds iron and has a low affinity for the IRE hairpin loops of the mRNA. Without bound IRE-BP at its 3'end, the mRNA is rapidly degraded and the transferrin receptor is not synthesized. (Note: When IRE-BP is bound to IRE hairpin loops at the 3'-end of the mRNA, the IRE-BP prevents degradation of the mRNA)

#### **CLINICAL COMMENTS**

**Arlyn Foma.** Follicular lymphomas are the most common subset of non-Hodgkin's lymphomas (25–40% of cases). Patients with a more aggressive course, as seen in Arlyn Foma, die within 3 to 5 years after diagnosis if left untreated. In patients pretreated with multidrug chemotherapy (in this case AV/CM), a response rate of 50% has been reported when interferon- $\alpha$  is added to this regimen. In addition, a significantly longer event-free survival has been reported when using this approach.

**Fig. 16.24.** Regulation of degradation of the mRNA for the transferrin receptor. Degradation of the mRNA is prevented by binding of the iron response element binding protein (IRE-BP) to iron response elements (IRE), 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.



Ominipotent stem cells in the bone marrow normally differentiate and mature in a highly selective and

regulated manner, becoming red blood cells, white blood cells, or platelets. Cytokines stimulate differentiation of the stem cells into the lymphoid and myeloid lineages. The lymphoid lineage gives rise to B and T lymphocytes, which are white blood cells that generate antibodies for the immune response. The myeloid lineage gives rise to three types of progenitor cells: erythroid, granulocytic-monocytic, and megakaryocytic. The erythroid progenitor cells differentiate into red blood cells (erythrocytes), and the other myeloid progenitors give rise to nonlymphoid white blood cells and platelets. Various medical problems can affect this process. In Mannie Weitzels, who has chronic myelogenous leukemia (CML), a single line of primitive myeloid cells produces leukemic cells that proliferate abnormally, causing a large increase in the number of white blood cells in the circulation. In Anne Niemick, who has a deficiency of red blood cells caused by her  $\beta^+$  thalassemia (see Chapter 15), differentiation of precursor cells into mature red blood cells is stimulated to compensate for the anemia.



Ann O'Rexia has a hypochromic anemia, which means that her red blood cells are pale because they contain low levels of hemoglobin. Because

of her iron deficiency, she is not producing adequate amounts of heme. Consequently, elF2 is phosphorylated in her reticulocytes and cannot activate inititation of globin translation.

Mannie Weitzels. Mannie Weitzels has CML (chronic myelogenous leukemia), a hematologic disorder in which the proliferating leukemic cells are believed to originate from a single line of primitive myeloid cells. Although classified as one of the myeloproliferative disorders, CML is distinguished by the presence of a specific cytogenetic abnormality of the dividing marrow cells known as the Philadelphia chromosome, found in more than 90% of cases. In most instances, the cause of CML is unknown, but the disease occurs with an incidence of around 1.5 per 100,000 population in Western societies.

**Ann O'Rexia.** Ann O'Rexia's iron stores are depleted. Normally, about 16 to 18% of total body iron is contained in ferritin, which contains a spherical protein (apoferritin) that is capable of storing as many as 4,000 atoms of iron in its center. When an iron deficiency exists, serum and tissue ferritin levels fall. Conversely, the levels of transferrin (the blood protein that transports iron) and the levels of the transferrin receptor (the cell surface receptor for transferrin) increase.

#### **BIOCHEMICAL COMMENTS**

Regulation of transcription by iron. A cell's ability to acquire and store iron is a carefully controlled process. Iron obtained from the diet is absorbed in the intestine and released into the circulation, where it is bound by transferrin, the iron transport protein in plasma. When a cell requires iron, the plasma iron-transferrin complex binds to the transferrin receptor in the cell membrane and is internalized into the cell. Once the iron is freed from transferrin, it then binds to ferritin, which is the cellular storage protein for iron. Ferritin has the capacity to store up to 4,000 molecules of iron per ferritin molecule. Both transcriptional and translational controls work to maintain intracellular levels of iron (see Figs. 16.23 and 16.24). When iron levels are low, the iron response element binding protein (IRE-BP) binds to specific looped structures on both the ferritin and transferrin receptor mRNAs. This binding event stabilizes the transferrin receptor mRNA so that it can be translated and the number of transferrin receptors in the cell membrane increased. Consequently, cells will take up more iron, even when plasma transferrin/iron levels are low. The binding of IRE-BP to the ferritin mRNA, however, blocks translation of the mRNA. With low levels of intracellular iron, there is little iron to store and less need for intracellular ferritin. Thus, the IRE-BP can stabilize one mRNA, and block translation from a different mRNA.

What happens when iron levels rise? Iron will bind to the IRE-BP, thereby decreasing its affinity for mRNA. When the IRE-BP dissociates from the transferrin receptor mRNA, the mRNA becomes destabilized and is degraded, leading to less receptor being synthesized. Conversely, dissociation of the IRE-BP from the ferritin mRNA allows that mRNA to be translated, thereby increasing intracellular levels of ferritin and increasing the cells capacity for iron storage.

Why does an anemia result from iron deficiency? When an individual is deficient in iron, the reticulocytes do not have sufficient iron to produce heme, the required prosthetic group of hemoglobin. When heme levels are low, the eukaryotic initiation factor eIF2 (see Fig. 16.22) is phosphorylated, and inactive. Thus, globin mRNA cannot be translated because of the lack of heme. This results in red blood cells with inadequate levels of hemoglobin for oxygen delivery, and an anemia.

#### **Suggested Readings**

#### Regulation of gene expression in prokaryotic and eukaryotic cells:

Lewin B. Genes VII. Oxford: Oxford University Press, 2000:273–318, 649–684. Thalassemias:

Watson J, Gilman M, Witkowski J, Zoller M. Recombinant DNA. Scientific American Books. New York: WH Freeman, 1992:540–544.

Weatherall DJ, Clegg JB, Higgs DR, Wood WG. The hemoglobinopathies. In Scriver CR, Beaudet AL, Slv WS, Valle D. The Metabolic and Molecular Bases of Inherited Disease. 8th Ed. New York: McGraw-Hill, 2001:4571–4636.

#### Leukemias and Lymphomas:

Goldman L, Bennet JC. Cecil Textbook of Medicine. 21st Ed. Philadelphia: WB Saunders, 2000:945–977.

# REVIEW QUESTIONS-CHAPTER 16

- 1. Which of the following explains why several different proteins can be synthesized from a typical prokaryotic mRNA?
  - (A) Any of the three reading frames can be used.
  - (B) There is redundancy in the choice of codon/tRNA interactions.
  - (C) The gene contains several operator sequences from which to initiate translation.
  - (D) Alternative splicing events are commonly found.
  - (E) Many RNAs are organized in a series of consecutive translational cistrons.
- 2. In *E. Coli*, under high lactose, high glucose conditions, which of the following could lead to maximal transcription activation of the *lac* operon?
  - (A) A mutation in the lac I gene (which encodes the repressor)
  - (B) A mutation in the CRP binding site leading to enhanced binding
  - (C) A mutation in the operator sequence
  - (D) A mutation leading to enhanced cAMP levels
  - (E) A mutation leading to lower binding of repressor
- 3. A mutation in the I (repressor) gene of a "non-inducible" strain of *E. coli* resulted in an inability to synthesize any of the proteins of the *lac* operon. Which of the following provides a rational explanation?
  - (A) The repressor has lost its affinity for inducer.
  - (B) The repressor has lost its affinity for operator.
  - (C) A trans acting factor can no longer bind to the promoter.
  - (D) The CAP protein is no longer made.
  - (E) Lactose feedback inhibition becomes constitutive.
- 4. Which of the following double-stranded DNA sequences shows perfect dyad symmetry (the same sequence of bases on both strands)?
  - (A) GAACTGCTAGTCGC
  - (B) GGCATCGCGATGCC
  - (C) TAATCGGAACCAAT
  - (D) GCAGATTTTAGACG
  - (E) TGACCGGTGACCGG

# 5. Which of the following describes a common theme in the structure of DNA binding proteins?

- (A) The presence of a specific helix that lies across the major groove of DNA
- (B) The ability to recognize RNA molecules with the same sequence
- (C) The ability to form multiple hydrogen bonds between the protein peptide backbone and the DNA phosphodiester backbone
- (D) The presence of zinc
- (E) The ability to form dimers with disulfide linkages