Gene expression is the underlying force determining phenotype. For many years, gene expression in mature tissues such as the heart and lung was viewed as an unchanging constant; changes in gene expression were thought to be a factor only in development. It is now clear that there is a plasticity of gene expression in mature tissue, and that disease states can alter the expression of genes.\(^1\) For example, alteration in gene expression occurs in cardiac hypertrophy. We are just beginning to understand how modification of gene expression may play an important role in conditions such as cardiac hypertrophy, ischemic heart disease, and COPD. **Transcription**, the synthesis of RNA based on a sequence encoded for by the gene, and **translation**, the synthesis of protein based on a sequence encoded for by RNA, determine which genes are expressed as final protein products. An understanding of the regulation of transcription and translation, both of which alter gene expression, will lay the foundation for future therapeutic manipulation of gene expression. This review of the fundamentals of gene transcription and translation is targeted at the clinician, who does not perform basic research. The review will begin with an overview of gene structure and transcription, and proceed to translation of RNA into protein. Some examples of drugs and diseases that block transcription and translation are included at the end.

In the future, molecular medicine will become a reality and treatments derived from application of molecular biology to disease states will be used routinely. In time, as the medical community embraces the era of “gene therapy,” it will become necessary for clinicians to understand the basis of transcription and translation. With this in mind, this review is aimed at bringing clinicians a working knowledge of the control of gene expression in the mammalian cell—and thus in the mammalian organ. Selected references are included for the interested reader.

First, to refresh the reader on the basic structure of genes, DNA encodes all the information necessary to make an organism. DNA consists of only four nucleotides: adenosine (A), thymidine (T), guanine (G), and cytosine (C). Adenosine and guanine are purines, and thymidine and cytosine are pyrimidines. The four nucleotides (A, T, G, and C) consist of a deoxyribose molecule (a type of sugar) bound to the respective purine or pyrimidine. The structure of DNA was determined by Watson and Crick in the 1950s based in part on x-ray diffraction studies done by Rosalind Franklin. The basic structure of DNA is a double helix, which one can think of as a ladder coiled around its long axis. The rungs of the ladder are made of matching base pairs, which are adenosine and thymidine (AT) and guanine and cytosine (GC). The base pairs are joined by hydrogen bonds, two for AT and three for GC. Because GC forms three hydrogen bonds rather than two, GC pairs are “stickier” and more likely to bond to each other. The sides of the ladder are made of the deoxyribose regions of the molecules, which are linked by phosphodiester bonds between the methyl side chain on one sugar with the phosphate side chain on the next. The phosphodiester bonds of the backbone of the helix (the sides of the ladder) are strong bonds, but the hydrogen bonds of the base pairs are weak. This is

---

**Key words:** DNA binding proteins; gene expression; initiation factors; mRNA stability; promoters; protein synthesis; RNA splicing; transcription; translation

---

\(^1\)From the Cardiology Section, Baylor College of Medicine and the Veterans Affairs Medical Center, Houston. Supported in part by grant HL52910 from the National Heart, Lung, and Blood Institute, Bethesda, Md.

Reprint requests: Dr. Knowlton, Cardiology Research, 151C Medical Center, 2002 Holcombe, Houston, TX 77030
important, because in order for the DNA to be copied to replicate the cell or to make RNA, it is necessary to separate the two strands that make up the helix at the location being copied. Although there are two strands to the helix, only one strand encodes the sequence from which RNA and subsequently protein are synthesized.

Transcription refers to the synthesis of new RNA from a DNA template. All genes are encoded by DNA located on chromosomes in the nucleus, or for a few genes, located in the mitochondria. The current consensus is that gene expression is regulated primarily at the transcriptional level, though this may change as more is understood about posttranscriptional events. As illustrated in Figure 1, the typical mammalian gene consists of a promoter region, which regulates how frequently RNA copies are made, followed by the transcription start site. RNA polymerase II, the enzyme that synthesizes new messenger RNA (mRNA), binds to the promoter.2-4

Two key features of mammalian promoters are the TATA box and the CAAT box (the names derive from key sequences in these structures). These are important for the binding of RNA polymerase, which then synthesizes RNA in a 5' to 3' direction. It should be noted that all DNA and RNA has directionality; just as a sentence only makes sense read in one direction, DNA and RNA are read only in one direction, 5' to 3'. By consensus, the start of transcription (ie, RNA synthesis) is referred to as base 1, and the promoter region is referred to as negative bases. Thus, the TATA box is located at -25 to -40 or so in most genes. In addition to the almost universal TATA box and CAAT box, individual genes have additional specific sequences, known as enhancers, upstream of the start site, which generally promote binding of RNA polymerase and synthesis of RNA, though some of the regulatory elements are inhibitory.

Enhancers and Other Factors

Enhancers are short sequences of DNA that have an important influence on gene expression. Typically enhancers are more than 100 bases negative to the transcription start site and in fact can be found hundreds of base pairs negative to the start site; enhancers also occur within the introns (noncoding inserts within the gene) and are then plus to the start site. Protein factors bind to enhancers and influence the binding of RNA polymerase and the start of transcription. Some of the protein factors are constitutive, ie, expressed in all cells, and some are inducible, only expressed at certain times and in certain cells.

Inducible factors are proteins that bind to the promoter and increase (usually) transcription. All the intracellular hormone receptors, such as the glucocorticoid receptor and the thyroid receptor, are inducible factors. The glucocorticoid hormone crosses the cell membrane and binds to a receptor located in the cytoplasm. The receptor then moves to the nucleus where it binds to glucocorticoid-responsive elements (GREs), which are enhancers present only on some genes. After binding, the GRE may increase or inhibit transcription. Naturally, only if the receptor is synthesized by the cell will this binding take place. Thus, the glucocorticoid hormone directly influences the genes expressed by the cell. Thyroid hormone and progesterone act in a similar fashion. The thyroid hormone receptor binds to the promoters for myosin heavy chain and sarcomplasmpic Ca2+ ATPase among other genes, and increases transcription of both these proteins. The effects of these hormones on transcription are still being defined. In general, when a protein binds to the gene to influence transcription, it is referred to as a trans-acting factor, while when a sequence in the gene itself (usually in the untranslated region) influences transcription, it is referred to as a cis-acting sequence. Sometimes a protein factor will act only on its own gene, and then it is referred to as cis-acting.

DNA binding proteins are all the proteins that bind with specificity to regulatory regions such as the TATA box, the GRE, and the CAAT box.2 DNA binding proteins include inducible factors as well as proteins that are always present. Four common binding motifs, a defined pattern of protein structure that facilitates interaction with DNA, have been observed in eukaryotes for the DNA-binding proteins. These proteins must bind to the DNA helix, and help open it, undoing the base-pair interactions, so that transcription can occur on a single strand of DNA. Many DNA-binding proteins are also proto-oncogenes; in other words, many of the DNA-binding proteins are associated with tumors, when their synthesis is uncontrolled and in turn cell growth is unchecked. Oncogenes will be the topic of a future article in this series on basic science for the clinician.

Zinc fingers include the glucocorticoid receptor, the estrogen receptor, and SP-1, a protein required for initiation of transcription in many promoters. Their
name derives from the presence of a zinc atom combined with the finger-like appearance of their tertiary structure. **Helix-turn-helix proteins** use two \( \alpha \)-helices at angles to each other to open DNA. Many prokaryotic and eukaryotic genes have helix-turn-helix proteins regulating their expression. **Helix-loop-helix proteins** are amphipathic (possessing hydrophobic and hydrophilic regions) and include c-myc, a proto-oncogene, and myo-D, important in skeletal muscle differentiation. Investigators believe there must be a myo-D-like protein for heart muscle, but it has not yet been identified. **Leucine zippers** contain repetitive domains with a leucine at every seventh amino acid. These proteins form dimers, and include c-fos and c-jun. The dimer resembles a zipper. Heat shock factor (HSF), regulating the heat shock response, has a structure suggestive of a leucine zipper.

Enhancers and promoters are specific to individual genes and help determine specificity of expression. In other words, the set of genes expressed in a cardiac muscle cell is different from what is expressed in an alveolar cell. Regulation of expression is necessary for appropriate differentiation of cells. A muscle cell should not synthesize hemoglobin and an alveolus should not make atrial natriuretic factor (ANF, normally synthesized only by cardiac atrial cells). There are other mechanisms of inhibition of transcription. For example, DNA methylation prevents transcription.\(^6\) The tertiary structure of DNA, characterized by the formation of nucleosomes, compact units of protein and DNA, will also inhibit transcription.\(^7,8\) If the promoter for a gene is within the core of a nucleosome, the gene will not be transcribed; whereas if the promoter is located in an inter-nucleosome region, often referred to as a nucleosome-sensitive site (ie, exposed DNA that can be acted upon by a nuclease enzyme under experimental conditions), then the promoter can bind activating factors and RNA polymerase and the gene can be transcribed. Sometimes RNA polymerase binds to the TATA box, but then pauses, needing other factors for elongation. Multiple regulatory mechanisms result in specificity of expression.

**Modification of mRNA**

The initial RNA synthesized by RNA polymerase is referred to as **pre-mRNA** because it usually does not represent the final form for translation into a protein.\(^9,10\) Eighty percent of RNA synthesized in eukaryotic cells is degraded before it ever leaves the cell nucleus. It is not known why this seemingly inefficient process occurs. As shown in Figure 2, after mRNA is synthesized, it undergoes a number of modifications. With a few exceptions, most notably many of the heat shock genes, most genes consist of **exons** and **introns**. Only the exons include coding regions for the protein, and the introns must be spliced out while the RNA is still in the nucleus. Why there are introns and how they came to exist is not known. How the introns are recognized and how the splicing occurs is of great interest in basic research. Multiple small nuclear RNA molecules (snRNAs) and proteins (snRNPs) are involved in the recognition of exons and formation of a complex termed a **spliceosome.**\(^11\) An intermediate structure with a “lariat” or loop of RNA to be excised is formed, prior to the linking of the two exons.\(^12\) The details of how the exons are recognized are still being determined. Heat shock protein (HSP 70) contains no introns. It is thought that the simplicity of the HSP gene, with the absence of introns and very little secondary structure in its 5' untranslated region, contributes to its increased synthesis when the cell is stressed and possibly unable to splice mRNA or translate it efficiently.

In addition to splicing, several other modifications are made to the RNA. In most proteins, the initial portion of the first exon does not code for protein, but instead is modified to create a **cap site.** This cap site
is important because it assists in binding to the ribosome at the time of translation. A third modification of the RNA is the addition of a poly-A tail, and this is done by the poly-A polymerase enzyme. The 3' end of the pre-mRNA contains a signal for the addition of a poly-A tail, which consists in mammals of about 250 adenosines, one of the four nucleotides comprising RNA. This poly-A tail increases the half-life of the RNA when it is in the cytoplasm. From a research perspective, this poly-A tail has another advantage; it is relatively easy to isolate from the cell all RNA ending in poly-A, which means all RNA coding for protein. In the cell, 90 to 95% of the RNA is ribosomal RNA and t-RNA (transfer RNA) rather than mRNA.

**ALTERNATIVE SPLICING**

For some proteins, variation in protein sequence is achieved by modification of the pre-RNA, by a process in which exons are spliced into or out of the final mRNA molecule. A good example of this is the protein fibronectin, a large macromolecule involved in the early phases of wound repair. Fibronectin, which is synthesized both in the liver and in tissues, is found in relative abundance in the plasma. At least three regions in fibronectin are alternatively spliced, i.e., present in the final protein product only some of the time: the variable region, eIIa, and eIIb. The eIIa and eIIb sequences are not present in the plasma fibronectin, which is synthesized in the liver. However, when fibronectin synthesis increases in a tissue, such as in the heart after myocardial infarction and in the injured lung, there is a marked increase in the protein incorporating the eIIb and/or eIIa domain. The function of these protein domains (eIIa and eIIb) is not known, but increased incorporation or splicing of these regions into the final sequence has been seen in a number of models of tissue injury. Fibronectin also provides another perspective on the degree of splicing (not alternative) that can occur. The gene encoding fibronectin is 79 kilobases (kb) long and the final mRNA is in the range of 7.9 kb. Thus, only 10% of the gene is contained in the mRNA. Consequently, splicing performs two functions: it removes introns and in certain proteins it alters the coding sequence that is expressed by including or excluding additional domains.

**TRANSLATION**

Once the RNA has been completely modified, it is then transferred across the nuclear membrane to the cytoplasm. The nuclear membrane is known to have pores, which are involved in the transport of proteins into and out of the nucleus, but it is not known whether this is the path taken by mRNA to the cytoplasm. In the cytoplasm, the mRNA is bound to ribosomes, the synthetic engines of the cell for proteins, which assemble from 40S (S refers to Svedberg units, a measure of sedimentation rate) and 60S subunits as they bind the mRNA.

Translation has three phases: initiation, elongation, and termination of the nascent peptide (new protein). In initiation, as shown in Figure 3, multiple initiation factors (IFs) bind the 40S subunit consuming GTP (like ATP, a source of high energy phosphate) in the process. This IF-40S complex binds the mRNA and a special fmet-tRNA, a modified methionine bound to its transfer RNA (t-RNA), binds at the AUG initiation site for translation on the mRNA. The triplet codon AUG codes for the methionine, which is the first amino acid in almost all proteins. The 60S subunit of the ribosome then binds what is now the 40S initiation complex and the IFs are released.

**INITIATION AND CONTROL OF TRANSLATIONAL EFFICIENCY**

The 5’ untranslated region (5’UTR) strongly influences the efficiency of translation. Certain sequences in the vicinity of the start codon (AUG) favor initiation. This initiation site is typically preceded by a purine-rich sequence (the nucleotides compris-
ing DNA and RNA are purines and pyrimidines). The consensus sequence, known as the Kozak sequence, for optimal initiation is GCGG A/G AUUG (A/G means either of these bases are at this position), but many genes do not have this entire sequence. The most frequent finding is the presence of a purine (A or G) at the −3 position and the presence of a G at +4. A long 5′UTR favors translation as does the absence of secondary structure in the 5′UTR. High GC content (70 to 90%) tends to result in interference with initiation due to the formation of hairpin loops within the 5′UTR. The GC pairs are “stickier” than AU pairs (uridine rather than thymidine occurs in RNA) and more likely to interact because GC pairs form three bonds while AU pairs form only two. If the 5′UTR binds to itself, it is difficult for the IFs and the ribosome to bind to it. Some genes have multiple start sequences and a complex 5′UTR; this is particularly so with cytokines and other proteins that are typically transiently expressed for only a few minutes.

**ELONGATION**

Following initiation, elongation of the nascent peptide must occur. In this sequence of events, the hydrolysis of GTP by eukaryotic elongation factor 1 (eEF-1) is the rate limiting step. Hydrolysis of GTP, which is accompanied by a conformational change in eEF-1, must happen first for a change in the interaction between the amino acid-tRNA complex and the mRNA. As summarized in Figure 4, a complementary aminoacyl-tRNA (amino acid bound to its individual tRNA) is delivered to site A on the ribosome by eEF-1, which is bound to GTP. Once the amino acid-tRNA is at the A site, the GTP is hydrolyzed and eEF-1 dissociates. eEF-1 is recycled, binding first to eEF-2 (another eukaryotic elongation factor) and releasing the GDP remaining from hydrolysis of GTP. eEF-1 then binds a new GTP molecule and is ready to pick up another amino acid-tRNA and repeat the elongation step. After this occurs, peptidyl transferase (found in the 60S subunit) catalyzes formation of a peptide bond with the preceding amino acid, permanently linking the amino acid to the new protein. The whole protein must now move down (translocate) to vacate the site for addition of the next amino acid. Translocation requires yet another factor, eEF-G or translocase. The whole process repeats itself until the stop codon is reached. Release factors recognize the stop codon and release the new protein (termination).

The fidelity of translation depends on the accuracy of binding between the tRNA and the mRNA. Each amino acid is coded for by a triplet of nucleotides. With a given amino acid coded for by three nucleotides and four different nucleotides, a total of 64 possible combinations exist. The 20 amino acids are each specified by between one and six different sequences. Three sequences, UAA, UAG, and UGA, specify stop codons. The correct sequence is determined by the tRNA, which contains a matching antisense of the codon, and which binds to the mRNA. The error rate is 1 in 10,000, which is much more frequent than the rate of error in DNA replication.

**CONTROL OF THE RATE OF TRANSLATION**

Phosphorylation and dephosphorylation of select initiation factors markedly alter overall rates of translation. Phosphorylation of eIF-2α, a subunit of one of the initiation factors, inhibits translation by preventing initiation. Although it would appear that phosphorylation of initiation factors must affect all translation in the same manner, some mRNA is translated efficiently despite phosphorylation of eIF-2α. In contrast to eIF-2α, eIF-4α, required for initiation, must be phosphorylated to be active. Stresses such as heat shock and hypoxia inhibit protein synthesis. Despite the involvement of multiple factors in protein synthesis, it is a single event, phosphorylation of eIF-2α, that results in near-global inhibition of protein synthesis during cellular stress.
The other regulatory linchpin for translation, dephosphorylation of eIF-4, will also block translation. Recently it has been demonstrated that insertion into cells of a gene coding for a variant of eIF-2α, which had been altered so that it was resistant to phosphorylation, resulted in resistance of protein synthesis in those cells to heat shock, providing confirmation of the importance of phosphorylation of eIF-2α in inhibiting translation.

Protein synthesis, like RNA synthesis, is directional. The mRNA is read 5' to 3' and the resulting protein is synthesized from amino terminal end to carboxyl terminal end. The first amino acid is always a modified methionine. Usually, as translation proceeds, multiple ribosomes become bound to a single mRNA. The ribosomes close to the 3' end of the mRNA have longer newly synthesized peptides than the ribosomes close to the 5' end as translation progresses 5' to 3'. When multiple ribosomes are simultaneously translating a given mRNA molecule, the resulting structure is referred to as a polysome.

**RNA Turnover**

Messenger RNA stability also plays an important role in determining the levels of a particular protein present in a cell. The half-life of mRNA varies widely. The mRNA for growth factors and cytokines have short half-lives and the mRNA for housekeeping genes have long half-lives; for example, the mRNA for c-fos, a proto-oncogene and growth factor, has a half-life of 8 to 30 min, while in erythroid cells, the mRNA for β-globin mRNA has a half-life of greater than 24 h.

A number of variables influence mRNA stability. Virtually all mRNA has a poly-A tail, which is added as a result of the poly-A signal in the 3' untranslated region (3'UTR). The length of the poly-A tail varies, and this in itself can affect mRNA stability. Messenger RNA without a poly-A tail is rapidly degraded, though there are exceptions; the mRNA for histones, a protein group found only in the nucleus, have no poly-A tail. Certain sequences in the 3'UTR promote instability and favor degradation; however, when these sequences are placed in other mRNAs, they have little to no effect, suggesting that more factors are important. The 5'cap structure, added while the mRNA was in the nucleus, also protects mRNA from degradation. There appears to be a link between translation and degradation. If translation is blocked by a drug, all mRNA is stabilized. It may be that formation of the polysome exposes some sequences in the mRNA that are vulnerable to endonucleases and exonucleases, which destroy RNA.

Other factors affect translation and mRNA stability. The 3'UTR affects overall translational efficiency. Certain AU rich elements (usually AUUUA) influence the efficiency of translation and the stability of the mRNA. Responsive elements, regulated by hormones, such as glucocorticoid, occur in the 3'UTR and in the 5'UTR. In the 3'UTR, these responsive elements, when bound by the appropriate factor, increase mRNA half-life rather than transcript rate. Hence, some factors, such as the hormones with intracellular receptors (glucocorticoid, thyroid, estrogen, etc.) affect both transcription by activating a receptor that then binds to a responsive element in the promoter and translation by binding to the mRNA.

**Effects of Drugs and Disease on Transcription and Translation**

Gene expression is the result of many factors that influence the rate of transcription, the rate of initiation, the efficiency of translation, and the stability of mRNA as summarized in Figure 5. A number of disease states and drugs are known to affect transcription and translation. Amanita phalloides, the poisonous mushroom also known as the death cup, contains α-amanitin, which binds RNA polymerase II and blocks the elongation phase of mRNA synthesis, effectively shutting down transcription. Ingestion of this mushroom is fatal. Diphtheria toxin acts by inactivating eEF-2 (eukaryotic elongation factor 2), which is important in translation. One molecule of diphtheria toxin can interfere with protein synthesis sufficiently to kill the cell. Pseudomonas aeruginosa
produces PA toxin, which also inactivates elongation factor eEF-2. Many antibiotics inhibit translation, fortunately affecting the bacterial ribosome and not the human. Streptomycin blocks initiation. Erythromycin binds to the 50S subunit (the prokaryotic equivalent of the 60S subunit) of the ribosome, inhibiting translation. Chloramphenicol inhibits peptide transferase in the 50S unit, preventing elongation. Almost all of the aminoglycoside antibiotics block eukaryotic translation, but at much higher concentrations than used clinically. Tetracycline binds to the 30S subunit (40S in eukaryotes) and inhibits binding of the aminoacyl-tRNA complex, blocking translation. Tetracycline also binds to the eukaryotic ribosome, but fortunately it is not concentrated in the eukaryotic cells as it is in bacteria and therefore has little effect on protein synthesis in patients. Shiga toxin, from Shigella dysenteriae, blocks binding of amino-acyl-tRNA, inhibiting protein synthesis.

**CONCLUDING REMARKS**

Furthering our understanding of the regulation of transcription and translation will aid in the development of therapies to combat disease by manipulating gene expression. In some cases, this will entail inserting a new gene into host cells, which is already being tried using CFTR (cystic fibrosis transmembrane regulator protein) in patients with cystic fibrosis. It should be apparent now that the presence of the gene will not necessarily guarantee production of the protein. Initial research on gene expression used *reporter genes*, genes not normally found in the host cells, and which had an easily measured protein product. The reporter genes include LacZ, encoding β-galactosidase, which metabolizes a substrate, that is added to the preparation to produce a blue color, luciferase, which produces a luminescent product that can be quantitated, and chloramphenicol acetyl transferase, measured by assaying the acetylation of radioactive chloramphenicol by the cell containing the construct (in this case promoter plus reporter gene) of interest. These genes are not normally found in the eukaryotic cell. With expression of genes normally found in mammalian cells, we may encounter problems with inhibition of transcription, through feedback mechanisms involving, for example, an already present defective host protein or through lack of appropriate regulatory factors. Posttranscriptional events may also influence expression through changes in mRNA stability or efficiency of translation.

Gene expression can also be influenced by altering factors involved in transcription and translation without introducing new genetic material into the host. By understanding the complex regulatory controls on transcription and translation, we can potentially manipulate these controls to upregulate even a single gene. For example, heat shock protein (HSP) 70 is thought to protect the cell from injury. If we could upregulate this gene prior to surgery and increase expression of the protein product, this might decrease tissue injury, decrease some postoperative complications, and improve outcomes. For now this is speculative, but this type of therapy may become standard in the future and potentially provides very potent tools to combat inherited diseases and the chronic diseases we so often treat such as COPD, congestive heart failure, and angina. For a few diseases, such as cystic fibrosis, gene therapy has already become a reality in the experimental setting. The future holds many potentially exciting developments for medical treatments.

**ACKNOWLEDGMENTS**: The author thanks her colleagues who critically reviewed this manuscript. In particular the author thanks Dr. Gabriel Habib and Dr. Susan Wiegers (Hospital of the University of Pennsylvania School of Medicine) for helpful insights into the comprehensibility of the article from a clinician’s point of view.

**REFERENCES**


CHEST / 107 / 1 / JANUARY, 1995 247
22 Samuel CE. The eIF-2α protein kinases, regulators of translation in eukaryotes from yeasts to humans. J Biol Chem 1993; 268:7603-06
26 Sachs AB. Messenger RNA degradation in eukaryotes. Cell 1993; 74:413-21