BASIC REVIEW

Molecular biology: A brief overview

Vivian Gahtan, MD, Eric T. Olson, BA, and **Bauer E. Sumpio, MD, PhD,** *New Haven, Conn*

Molecular biology, the study of the mechanisms of transmission of genetic information, is integral to the understanding of vascular disease processes. Implicated in the initiation and progression of cardiovascular disease are single-gene defects and multiple-gene products interacting with environmental factors. Basic concepts and definitions for studying the mechanism of gene expression and an overview of commonly applied laboratory techniques are reviewed.

TRANSCRIPTION AND TRANSLATION OF THE GENETIC CODE

Deoxyribonucleic acid (DNA), a long polymer of linked nucleotides with deoxyribose as their sugar and a hydrogen-bonding base pair (guanine, cytosine, adenine, or thymine), can form double-stranded or helical structures and is the fundamental substance that forms genes. In vivo, DNA is complexed to DNA-binding proteins to form a three-dimensional supercoiled state. This condensed DNA structure is commonly known as chromatin. Three major processes involve the transmission of genetic information from DNA (Fig 1): (1) replication of the DNA molecule by a template mechanism (nucleus); (2) transcription, the process of ribonucleic acid (RNA) synthesis by DNA (nucleus); and (3) translation of messenger RNA (mRNA), an RNA molecule transcribed from the DNA of a particular gene, which results in protein synthesis (cytoplasm). These synthesized proteins are integral to cell structure and function.

A gene is a segment of DNA involved in production of an RNA chain and sometimes a polypeptide. Genes that encode proteins comprise several basic elements. mRNA-encoding sequences (exons) are usually discontinuous and

are separated by intervening non-coding sequences (introns). Some small human genes span 2000 or 3000 base pairs (bp). Genes encoding large polypeptides can have more than 100 exons and can be more than 1 million bp long. Intergenic regions do not code for protein, but indirectly can influence gene expression by affecting the formation of higher order chromatin structures. These include open chromatin (euchromatin), which is a minimal requirement for expression of nearby genes, or condensed chromatin (heterochromatin), which is transcriptionally silent.

Gene regulation. Each gene has a promoter, which is the recognition site used by the RNA polymerase enzyme complex as it begins to interpret the genetic code that it encounters along the chromosome. Regulatory proteins bind to promoter regions and guide the RNA polymerase complex to the start site; most stimulate initiation, but some suppress transcription. Gene promoters, composed of several short (5-10 bp) sequences, are found near genes expressed in a certain cell type or by a certain stimulus and are often identical. These common regulatory sequences give genes encoding unrelated proteins the capacity for co-expression. Some regulatory sequences can act as promoters for one gene, enhancers for another, and suppressors for other genes.

There are an estimated 10,000 to 100,000 genes coding for proteins along the 22 human autosomes and the X and Y sex chromosomes. Each cell type in an individual human has the same chromosome content, except for the rearrangements and deletions that occur naturally in the course of B lymphocyte immunoglobulin gene and T lymphocyte antigen receptor gene expression. The mRNAs produced by transcription of the genes in a particular cell type are unique for specific physiological or experimental conditions. Some proteins are maintained at constant concentrations in the cell in all circumstances and are encoded by the "housekeeping" genes (eg, actin, reduced nicotinamide adenine dinucleotide). Exceptions in specific cell types occur, and more sensitive or specific tests of gene function are necessary to reveal subtle regulatory differences. All cell types undergo changes in gene expression during their normal cellular functions. The expression of most proteins is regulated over tissue development, over the course of the cell cycle in normal growth, and with particular conditions, in response to acute stimuli for cellular activation.

The flux of intermediate forms of chromatin and specialized organelles within the nucleus likely represent addi-
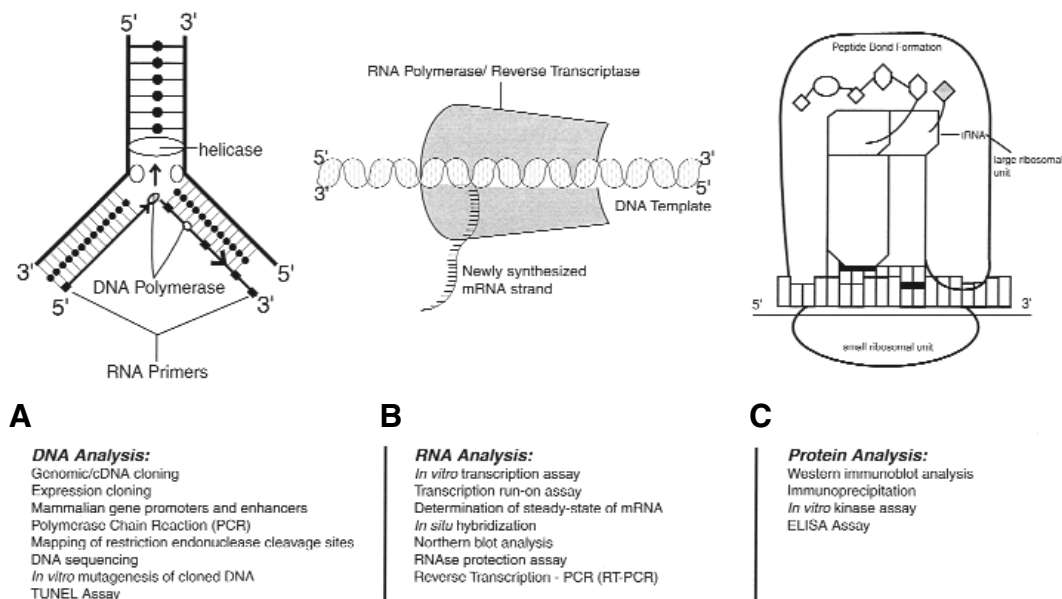
Fig 1. Transcription and translation of the genetic code. The three major processes involved in the transmission of genetic information are: **A,** DNA replication, which involves the synthesis of a complementary strand of an original DNA template; **B,** DNA transcription, in which a complementary mRNA strand is synthesized from an original DNA template; and **C,** translation of mRNA, resulting in protein synthesis. Each process is represented in this figure, with the commonly used molecular biology techniques applied to study these processes.

tional mechanistic components operating to regulate gene expression. These sites of interaction of multi-protein complexes with chromatin carry out DNA replication, repair, and transcription; sites of interaction of other multi-protein complexes with newly synthesized RNA carry out capping, polyadenylation, splicing, and transport.

Gene expression. Gene expression involves the transfer of coding information from DNA to mRNA (transcription; Fig 1*B*) and the subsequent decoding of the gene sequence resulting in the production of protein (translation; Fig 1*C*). Re-introducing the molecular construct into cells allows one to follow the production of the reporter mRNA, reporter protein, or both and to determine whether the control sequence is sufficient to induce expression. Conversely, in the preparation of a series of constructs (eg, 5'-deletion mutants) some of these artificial DNA sequences cannot promote the expression of an adjacent protein-encoding sequence when critical stimulatory regions are removed. The sequence of a promoter is often initially discernible because of its similarity to promoters from other previously characterized genes. The characteristics of a gene promoter can vary depending on the assay used: a promoter-reporter construct may be expressed at one level in vitro with cell extracts, at a different level when transferred into living cells, or still differently when engineered into the genome of a line of transgenic animals. Manipulation of the cell cycle has been used as a means of better understanding arterial disease processes.

Cell cycle. Most cells are in the G_0 phase and are non-cycling. For a cell to replicate, it enters the cell cycle. The four basic phases recognized in the cell cycle (Fig 2) are: (1) the Gap_1 (G_1) phase, involving most protein biosynthesis; (2) the Synthesis (S) phase, entailing most DNA synthesis and some S-phase–specific protein synthesis; (3) the Gap_2 (G_2) phase, involving some protein synthesis in preparation for mitosis; and (4) the Mitosis (M) phase, in which there is shut-down of protein and DNA synthesis and chromosome condensation for mitosis. The phases of the cell growth cycle except mitosis are called interphase, and it is during interphase that gene expression is observed. There are functionally significant links between the transcription-translation systems that yield proteins and the replication systems that yield DNA. Manipulation of the cell cycle has been used as a means of better understanding arterial disease processes.

Apoptosis. Apoptosis, defined as "programmed cell death," is recognized in both normal adaptive responses (tissue homeostasis) and pathologic states in which a cell actively participates in its own cell death. The morphological features associated with apoptosis are loss of cell membrane asymmetry and attachment, condensation of cytoplasm and nucleus, and internucleosomal cleavage of DNA. The dying cells then become fragmented into apoptotic bodies, which are eliminated by phagocytic cells without eliciting significant inflammatory damage to surrounding cells. Both morphological features and the lack of an inflammatory response distinguish apoptosis from necrosis. Inappropriate regulation of apoptosis has been

suggested in certain pathological states such as ischemia, stroke, atherosclerosis, and intimal hyperplasia.

METHODOLOGY AND EXPERIMENTAL DESIGN

The methodology currently available can precisely detect similarities and differences between biomolecules from changes of a single base to rearrangements of an entire gene region. The resulting differences in the proteins encoded can be predicted by means of comparison to known structural motifs and can also be observed in the test tube or experimental animal, with actual production of natural and mutant recombinant molecules. The ability to isolate genes and disassemble and reassemble them and to produce replenishable supplies of the DNA, encoded RNA, and recombinant protein from each of these constructions is largely caused by the commercial availability of hundreds of purified and well-characterized enzymes. With these powerful methods of handling DNA, naturally occurring mutants are no longer required to correlate structure and function.

The study of the intact cardiovascular system in animal models is a useful means of studying human disease processes at the molecular level. Animal studies are suitable for studying the development of pathological conditions at the molecular level and are needed to test procedures involving gene therapy that can then be extended to human clinical trials. Whole animal experiments also include the production of transgenic rodents or other animals, with an added gene construct, or the preparation of knockout mice, with a gene deleted by homologous recombination with a deletion construct.

In situ analysis of biomolecules provides positional information about the subcellular, cellular, or tissue level of activity. Histological sections can be analyzed for gene localization with chromosomal in situ hybridization, for mRNA localization to cell types in a tissue with in situ hybridization, and for antigen localization with immunocytochemistry. The orientation of polarized cells (eg, vascular endothelial cells) and their different cell membrane domains in intact tissues can provide experimental methods for manipulating molecular processes at one cell surface, for gene targeting to a cell facing the lumen, and for the preparation of specific cellular membranes, organelles, and molecules that would be randomly sampled in a total tissue preparation.

Continuously growing cell lines (human or animal) as models for the biology of normal cells are derived from a clone of cells recovered after the transforming events of tumorigenesis and subsequently selected biologically for survival. Continuously growing cell lines are convenient, reliable, and produce more easily reproducible results than studies with preparations of cells from whole organs. A basic disadvantage of using cell lines is the difference in control of growth and differentiation states seen in transformed cells. Therefore, in most in vitro studies, early passage vascular cells are used so that findings are more applicable to in vivo processes.
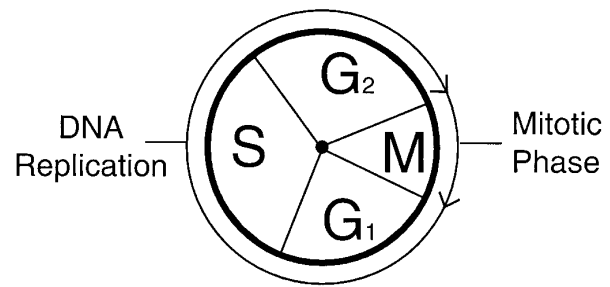

Fig 2. The cell cycle. The four basic phases of the eukaryotic cell cycle are shown. The G_1 phase involves most protein biosynthesis. The S-phase entails most of the cell's DNA synthesis and some additional protein synthesis. The G_2 phase is characterized by the protein synthesis necessary for mitosis. In the M phase, the chromosomes condense and cell replication (mitosis) is initiated.

Cell signaling pathways in human cells play critical roles in cell behavior and subsequent normal, adaptive, and pathologic responses. Signal transduction pathways extend into the cell from the extracellular environment through one or more domains of the cell membrane and the cytoplasm to the nucleus, where responses are carried out by gene expression. Specific membrane receptors are linked to signaling systems at the inside face of the cell membrane, and consist of glutamyl transpeptidase-binding proteins, second messenger generating enzymes, protein kinases, nucleotide, lipid, and cation substances, and their associated regulatory proteins. Responses potentially involve immediate cell activation, proliferation, differentiation, and/or migratory processes. The particular subcellular localizations of molecular events in the pathophysiology of the cardiovascular system are important for designing improved diagnostic and potentially therapeutic interventions involving the smooth muscle cells, endothelial cells, macrophages, T lymphocytes, and platelets. The modulation of gene expression in these cell types is intimately involved with the normal and pathological responses to external events, and the control points regulating gene expression provide numerous specific targets for novel diagnostic and therapeutic approaches to cardiovascular disease.

DNA ANALYSIS AND GENE STRUCTURE

Modeling of the genetic transmission of genetic information in living cells and in enzyme-catalyzed reactions generates molecular products. The desired products are called signals, and the unwanted materials are called noise. The experimental use of the enzymes' DNA-dependent DNA polymerases (DNA replication), RNA-dependent DNA polymerases (reverse transcription), and DNA-dependent RNA polymerases (gene transcription) provides specific information and the molecular production by various cellular and viral gene machines. Methods of DNA analysis are listed in Fig 1*A*.

The types of DNA molecules used in molecular biology can be conveniently categorized by means of size and
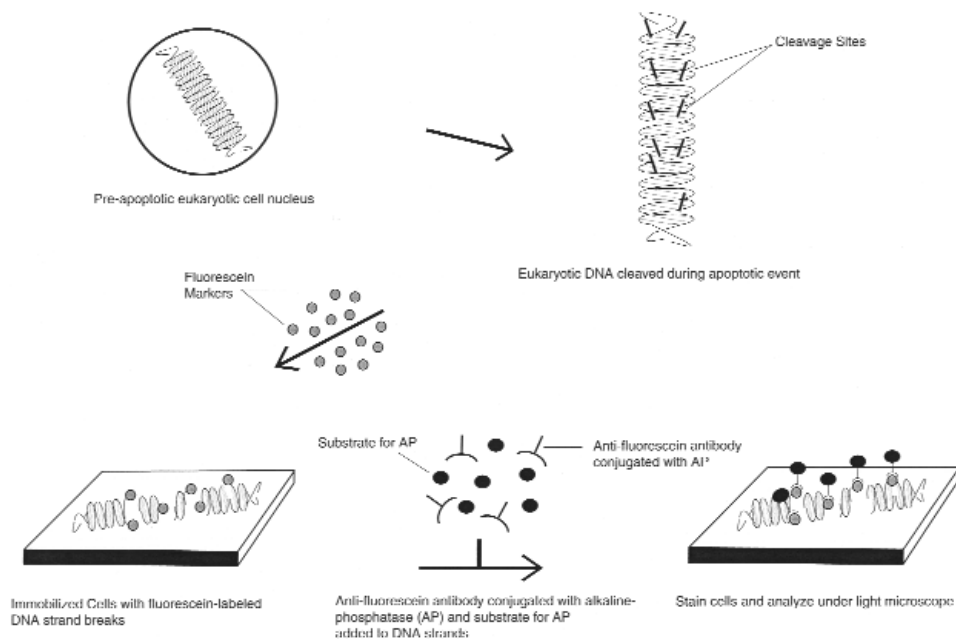
Fig 3. Terminal deoxynucleotidyl transferase-mediated dUTP nick end labeling (TUNEL) assay. DNA strand breaks with exposed carboxy (-OH) termini are labeled, conjugated, and studied to determine in vitro pre-programmed cell death or apoptosis.

are categorized as synthetic oligodeoxynucleotides or oligos (single-stranded linear sequence of nucleotides), plasmids (an extrachromosomal autonomously replicating circular DNA segment), bacteriophage DNA, single-stranded phage and phagemid DNA, linear fragments of DNA constructs, genomic DNA, linear fragments of genomic DNA, chromatin, and chromosomes. Each of these molecular forms of DNA can be used as substrates or analytical probes for specialized techniques. Selected techniques will be described in more detail.

Southern blot analysis. Southern blotting was developed by E.M. Southern as a means of providing preliminary information on gene structure. This technique involves transfer of denatured DNA from agarose gel to a nitrocellulose or nylon filter, where it can be annealed with a radio-labeled complementary nucleic acid.

Vectors for cloning. A vector is a self-replicating DNA molecule that transfers a DNA segment between host cells and provides for self-replication, with or without the expression of the foreign DNA as RNA and protein. Subsequently, many specialized vectors that can perform intricate molecular manipulations with DNA, RNA, or protein have been designed. These vector sequences encode a few important functions that give the cell containing the foreign DNA certain reproductive advantages.

Complementary DNA cloning. Complementary DNA (cDNA) is a single-stranded DNA copy of an mRNA made with the use of the enzyme reverse transcriptase; cDNA contains only the coding sequences of a gene. A cDNA library consists of a collection of DNA molecules comprising the chosen vector sequence and any one of the thousands or millions of cDNAs representing the genes expressed in the cells used for the original RNA source. The vector, also known as a vehicle, can be a plasmid or a bacteriophage. All cDNA vectors include a sequence that functions as a bacterial origin of DNA replication, and this allows the production in bacteria of a large number of copies of the entire plasmid construct, independent of bacterial cell growth.

Mammalian gene promoters and enhancers. Well-characterized promoters and enhancers of gene expression can be used in the design of an effective mammalian expression vector. In some instances, a promoter isolated from a genomic clone of a mammalian gene is useful for cloning genes functionally related to the promoter chosen. The functional characteristics of a particular region of a gene are commonly determined by their stimulation of gene expression in cells or in cell extracts, and the resulting RNA or protein gene products are detected by means of a specific assay. The structural characteristics of a gene promoter or enhancer can often be observed in experimental systems by the binding of particular proteins.

Polymerase chain reaction. Polymerase chain reaction (PCR) is an in vitro method for the enzymatic synthesis of specific DNA sequences by repetitive cycles, resulting in an exponential increase in product. This method uses two synthetic oligonucleotide primers flanking the region of interest in the target DNA and DNA
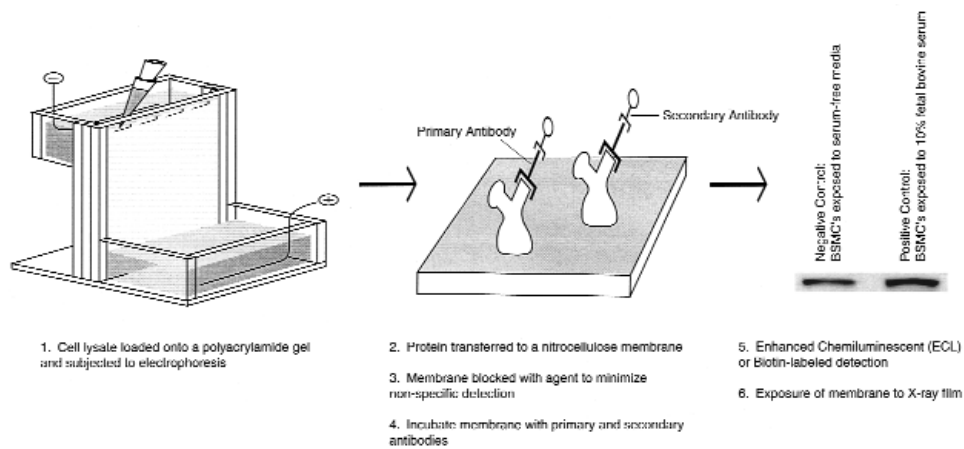
Fig 4. Western immunoblot analysis. Cell lysates are subjected to electrophoresis, transferred to a nitrocellulose membrane, incubated with protein-specific primary and marker-labeled (conjugated) secondary antibodies, and subsequently detected by means of chemiluminescence and x-ray film development.
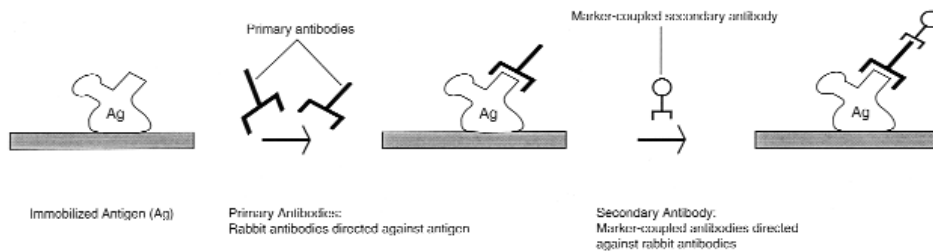

Fig 5. Enzyme-linked immunosorbent assay (ELISA). Immobilized antigens are incubated with protein-specific primary antibodies and marker-labeled (conjugated) secondary antibodies. Protein complexes are then quantified by means of colorimetric assay.

polymerase to amplify these sequences. At the completion of each cycle of this repetitive production of the DNA located between the primer-binding sites, the original DNA template and the DNA product separate, becoming substrates for each subsequent reaction. PCR can be applied to the routine preparation of DNA fragments of previously sequenced genes, to the detection and quantification of mutant alleles or variant alleles of a known gene, and to clinical assays for the detection of viral sequences, oncogene mutations, and other disease-associated or normal alleles of any gene that has been previously cloned and sequenced.

DNA sequencing. DNA sequencing is based on sequence information obtained by means of generation of DNA fragments of different lengths, which start at a fixed point and terminate at specific nucleotides. The DNA fragments are separated precisely by size with electrophoresis on a denaturing polyacrylamide gel. Fragmentation can be produced with chemical cleavage or enzymatic copying of single-stranded DNA. The results from the gel are read by detecting bands from the bottom of the gel in succession from the alternate lanes marked for the nucleotides A, C, G, or T and recording the sequence of these signals in a linear array, the DNA sequence.

In vitro mutagenesis of cloned DNA. The routine modifications of a single nucleotide, a short region of sequence, or kilobases of DNA are feasible by means of the approach of in vitro mutagenesis with a cloned gene as a template in a plasmid vector and a specifically designed synthetic oligodeoxynucleotide as the mutagenic primer. The oligonucleotide is designed to include perfect matches to sequences flanking the modified site. The products are the original sequence plus a mutated sequence. Several sophisticated variations provide for direct selection of mutated plasmid sequences. Essentially any variation of the original cloned gene can be readily prepared in large quantities. These types of preparations are critical for the precise mapping of DNA functions and are often required in the construction of multi-component DNA constructs for special uses.

Terminal deoxynucleotidyl transferase-mediated dUTP nick end labeling assay. The terminal deoxynu-

cleotidyl transferase-mediated dUTP nick end labeling (TUNEL) assay is used as a means of detecting apoptosis by preferentially labeling DNA strand breaks generated during apoptosis and can distinguish these cells from cells undergoing necrosis (Fig 3). DNA strand breaks with free 3'-OH termini are labeled in an enzymatic reaction. Then terminal deoxynucleotidyl transferase, which catalyzes polymeration of nucleotides to free 3'-OH DNA ends, is used to label DNA strand breaks. Incorporated fluorescein is detected by means of antifluorescein antibody fragments and conjugated with alkaline phosphatase. The apoptotic cells are stained and can be viewed with a light microscope.

RNA ANALYSIS AND GENE EXPRESSION

mRNA quantity, size, and cellular localization demonstrate the varying status of gene expression. Transcription initiation, in which most gene regulation apparently occurs, is the most difficult aspect of mRNA biosynthesis to monitor directly. The resulting steady-state levels of mRNA and polypeptide encoded by a gene are more readily detectable, and most analyses of gene expression deal with differences in steady-state levels of gene products in varying physiological or pharmacological conditions. Selected techniques will be discussed, and a more complete list is given in Fig 1*B*.

In situ hybridization. In situ hybridization is used as a means of localizing, within intact chromosomes, eukaryotic cells, or bacterial cells, nucleic acid segments complementary to specific labeled probes. To localize specific DNA sequences, specimens are treated so as to denature DNAs and to remove adhering RNAs and proteins. The DNA segments of interest are then detected by means of hybridization with labeled nucleic acid probes. The distribution of specific RNAs within intact cells or chromosomes can be localized by means of hybridization of sectioned specimens with an appropriate RNA or DNA probe.

Northern blot analysis. Northern blot analysis is used as a means of examining size and abundance of mRNA. This technique involves the transfer of RNA from agarose gel to a nitrocellulose or nylon filter on which it can be hybridized to a complementary nucleic acid. It is generally used as a means of examining size and abundance of mRNA.

TRANSLATION

Historically, the in vitro kinase assay was the primary method of quantifying phosphorylated (activated) proteins. Subsequently, immunoprecipitation was used for the isolation of specific proteins in solution, and then Western blot analysis was performed as a means of isolating the phosphorylated protein from the total protein. More commonly, phosphospecific antibodies are available and used as a means of directly detecting activated protein. Common methods used as a means of analyzing the products of translation are listed in Fig 1*C*. Selected techniques are described.

Western blot analysis. Western blotting is used as a means of immunodetecting specific proteins (Fig 4). It is a procedure for the transfer of protein products from an acrylamide gel to a nitrocellulose filter that is subsequently incubated with protein-specific antibodies. The protein-bound antibody is then detected with a secondary antibody. Signals are generated by means of an enzymatic reaction and detected by using an x-ray film.

Enzyme-linked immunosorbent assay. The enzyme-linked immunosorbent assay technique is used as a means of quantifying specific protein antigens in a solution (Fig 5). Although there are different protocols, generally the antigen is linked to an antibody-enzyme conjugate. The unbound conjugates are removed, and a chromogenic or fluorogenic substrate is added. As the substrate is hydrolyzed by bound enzyme conjugate, a colored or fluorescent product proportional to the amount of protein is generated and quantified with a microplate reader.

CONCLUSION

The technology of molecular biology has changed the way we study vascular disease. Identifying the mechanisms responsible for and the molecules involved in these processes should result in more accurate diagnostic testing and may facilitate the development of gene and cell therapy.

REFERENCES

1. Peyman JA, Sumpio BE. Molecular biology and the vascular surgeon. In: Sidawy AN, Sumpio BE, Depalma RG, editors. The basic science of vascular disease. 1st ed. Armonk, NY: Futura; 1997. p. 17-68.
2. Gahtan V, Peyman JA, Sumpio BE. Understanding molecular biology. Semin Vasc Surg 1998;11:125-33.
3. Alberts B, Bray D, Lewis J, et al. Molecular biology of the cell. 3rd ed. New York: Garland; 1994.
4. Darnell J, Lodish H, Baltimore D. Molecular cell biology. New York: Scientific American Books; 1990.
5. Berg P, Singer M. Dealing with genes: the language of heredity. Mill Valley, Calif: University Science Books; 1992.
6. Watson JD, Hopkins NH, Roberts JW, et al. Molecular biology of the gene. 4th ed. Menlo Park, Calif: Benjamin/Cummings; 1987.
7. Ausubel FM, Brent R, Kingston RE, et al, editors. Current protocols in molecular biology. New York: John Wiley & Sons; 1987.
8. Xing Y, Johnson CV, Dobner PR, Lawrence JB. Higher level organization of individual gene transcription and RNA splicing. Science 1993;259:1326-30.
9. Carter KC, Bowman D, Carrington W, et al. A three-dimensional view of precursor messenger RNA metabolism within the mammalian nucleus. Science 1993;259:1330-5.
10. Wansink DG, Schul W, van der Kraan I, et al. Fluorescent labeling of nascent RNA reveals transcription by RNA polymerase II in domains scattered throughout the nucleus. J Cell Biol 1993;122:283-93.
11. Kriegler M. Gene transfer and expression: a laboratory manual. New York: Stockton Press; 1990.
12. Russo VEA, Brody S, Cove D, Ottolenghi S, editors. Development: the molecular genetic approach. Berlin: Springer-Verlag; 1992.
13. Gavrieli Y, Sherman Y, Ben-Sasson SA. Identification of programmed cell-death in situ via specific labeling of nuclear-DNA fragmentation. J Cell Biol 1993;119:493-501.