

THE JOURNAL OF
ALLERGY
AND
CLINICAL IMMUNOLOGY

VOLUME 88

NUMBER 4

Postgraduate course

**Molecular biology and immunology:
An introduction**

John R. Rodgers, PhD, and Robert R. Rich, MD* *Houston, Texas*

The science of medicine has changed dramatically during the past decade as the power of recombinant DNA technology has been applied vigorously to virtually all the medical subdisciplines. Indeed, during the past several years, this technology has moved dramatically from the laboratory bench into the diagnostic and therapeutic armamentarium of practicing physicians.¹ Because the future promises a continuing avalanche of advances, it is important for clinicians as well as academicians to appreciate the fundamental concepts of molecular biology. We propose here to review salient aspects of molecular biology with an emphasis on certain aspects of molecular immunology. Highly readable but detailed accounts of molecular biology may be found in publications by Lewin (1987)² or Watson (1987).³

It is useful at the outset to understand the particular way in which scientists, and, more recently, the public, have come to use the term "molecular biology." This phrase at first referred to the interests of certain physicists and x-ray crystallographers who, in the decades bracketing the second World War, were turning their attention from particle physics and molecular structure to the study of life.⁴ This great first phase in the history of molecular biology led to the radical transformation of biology from a discipline derivative of natural history into a deliberate program of reduc-

Abbreviations used

MHC: Major histocompatibility complex

mRNA: Messenger ribonucleic acid

NF- κ B: Nuclear factor- κ B

ISRE: Interferon-stimulated response element

PCR: Polymerase chain reaction

TCR: T cell receptor

tRNA: Transfer ribonucleic acid

VDJ: Variable, diversity, and joining elements

Ab: Antibody

Ag: Antigen

A: Adenine

G: Guanosine

C: Cytosine or immunoglobulin constant domain

T: Thymine

U: Uracil

tionism. Just as Darwinism in the previous century (and neo-Darwinism in this century)⁵ had proposed the great null hypothesis that all of life stemmed from the random processes of mutation and selection, so the physicist's molecular biology proposed that the mechanisms subtending genetics, cell structure, morphogenesis, metabolism, and even behavior should now be traced to the angles and lengths of molecular bonds.

Contemporaneously, the extraordinary discoveries that the genes of bacteria⁶ and their viruses⁷ were composed of DNA led to equally heady postwar advances by geneticists and bacterial physiologists. In rapid succession, it was discovered that genes, like

From the Departments of Microbiology and *Immunology and *Medicine, Baylor College of Medicine, Houston, Texas.

Reprint requests: John R. Rodgers, PhD, Department of Microbiology and Immunology, Baylor College of Medicine, Houston, TX 77030.

1/1/31536

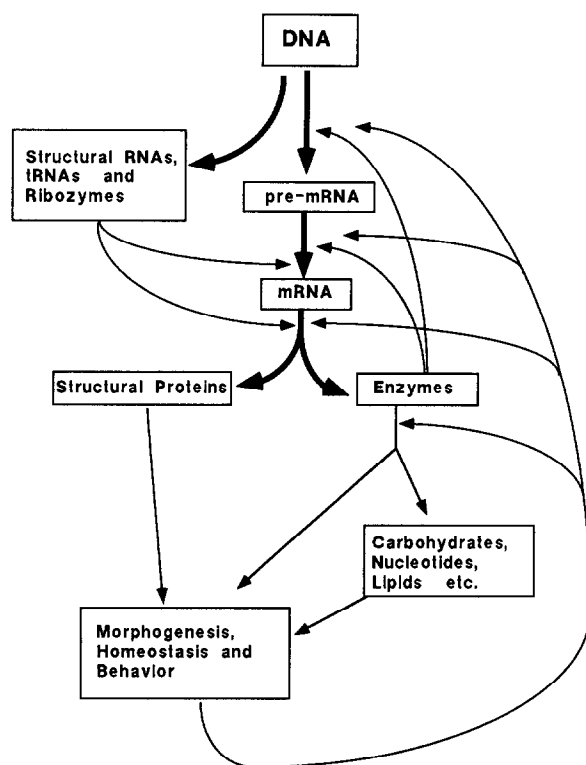


FIG. 1. The flow of genetic information. According to Crick,³⁹ genetic information can flow from DNA to RNA to protein (*heavy arrows*). However, the "meaning" of the encoded information depends on various levels of context (*light arrows*), which allows the genome to be responsive to the cellular and external environments.

atoms, could be subdivided, with the smallest division later recognized as the base pair. Regulatory elements ("promoters" and "operators") under the influence of the newly recognized class of regulatory proteins ("repressors" and "activators") controlled the expression of structural proteins and enzymes.

Much of the new terminology was adapted from the wartime language of information theory, cybernetics, and electronic communications. DNA itself carried the genetic information, but this information was somehow transcribed into an unstable molecular "messenger," demonstrated to be RNA. mRNA could be "translated" into protein, and from this discovery came the rapid deciphering of the genetic "code" in which a particular amino acid is specified by a trio of base pairs known as a "codon." The genetic code specifying a particular protein is analogous to the recorded program of a digital computer.

Molecular biology is now generally regarded as study of the cellular computer: how the information content of these molecular programs is stored in the form of genes and expressed as proteins. In addition, "molecular biology" refers to the technical prowess

that can be applied by myriad fields, from farming to forensics to the production of therapeutic factors, such as recombinant hormones. Thus, as commonly used, the term has a somewhat narrower meaning than terms such as biochemistry and cell biology, which are, of course, also directed to studies of molecules and biology. Nevertheless, because of the power of this technology to answer fundamental questions, academic departments of biochemistry and cell biology (not to mention allergy and immunology) are increasingly focusing research efforts on problems that depend on the concepts and techniques of molecular biology.

In this article we will first consider the flow of information within the cell: how the information content of DNA is expressed at the cellular level as protein and, to a certain extent, how information from the cellular environment is transmitted back to the genome. In the second part of the article, we will touch on technologic advances in molecular immunology.

Code and context: The flow of genetic information

A study of molecular biology begins with genes, which are the storehouses of heritable information. As with all information, the information contained in DNA can be "understood" in different ways, depending on its context.⁸ The process of the use of this information involves complex and carefully regulated activation of only those genes that are appropriate for expression in any particular cell. Thus, there are two kinds of information encoded in DNA. One kind of information provides a "genetic code" for specifying the exact linear sequence of RNA or proteins as decoded from the linear sequence of RNA. These linear sequences are the classic sense of "gene," and their information corresponds to the digital information stored in a computer memory. The flow of this information follows the "central dogma of molecular biology," which states that information flows steadily from DNA to RNA to protein (Fig. 1). In this sense the genome may be considered as the read-only memory of a cellular computer, encoding thousands of specific programs (RNAs, proteins, and, indirectly, metabolic pathways).

With the notable exception of the retroviruses, which are capable of writing to this "read-only" memory, this dogma holds true. However, for the expression of genetic information to be timely and appropriate to the livelihood of the organism, the cellular computer must be attentive to the flow of environmental and cellular information impinging on it; the computer must have mechanisms to control which program is expressed and when. The exact nature of these controls is an area of the most exciting research

currently but is largely beyond the scope of this discussion. This informational process can be bidirectional; certain sequences of DNA can encode information both to enhance and to diminish the expression of a particular gene, the decision as to which is made by a host of transcription factors, themselves proteins. It has been estimated that our protein-coding genes share 95% sequence identity with genes of our closest relatives, the chimpanzee. The information that specifies the significant (to our eyes) differences between us must lie within this second category of regulatory sequences.

Ultimately, however, both kinds of information are expressed through the process of transcription by which the sequence information of active genes is copied into RNA. Some RNAs are functional in later stages of information processing either as cofactors or directly as catalytic *ribozymes*.^{9, 10} Most RNAs are not directly functional but program the synthesis of specific proteins. These mRNAs carry messages in the form of nucleotide sequences, which are *translated* into the amino acid sequences of proteins in the cytoplasm.

A point of potential confusion is that DNA encodes only the RNA and protein constituents of a cell, although it is readily appreciated that other classes of molecules, such as carbohydrates and lipids, are equally essential for cellular function. This apparent paradox is readily resolved when it is understood that the biosynthesis and catabolism of such molecules as lipids and carbohydrates depend on enzymes (Fig. 1). Thus, the entire biochemical machinery of the cell can be tied directly or indirectly to its genes and the products of those genes.

We will focus on the mechanism by which the information content of DNA is ultimately expressed as protein, relying mostly on the example of the mammalian immunoglobulin heavy chain locus. There are few issues relevant to molecular immunology that do not also apply to this single locus.

The structure of RNA and DNA

The mechanisms of transfer of information from DNA → RNA → protein make use of a special chemical property of certain nucleic acids termed *complementarity*. Of the many discoveries in molecular biology for which Nobel prizes have been awarded, perhaps the most famous stemmed from the announcement in 1953 by Watson and Crick¹¹ of the fundamental structure of DNA, based on complementarity. In this structure, two strands are coiled about one another in a complementary fashion.

Each DNA strand consists of a polypentosephosphate backbone to which are covalently attached ni-

trogenous bases consisting of heterocyclic rings of carbon and nitrogen atoms (Fig. 2). The backbone consists of alternating phosphate residues and sugars in which the phosphate group forms a diester linkage between the 5' position of one pentose ring and the 3' position of the next (Fig. 3). In each strand, all the pentose sugars are aligned in the same 5' → 3' orientation, providing the sense or *polarity* of the strand. As both DNA and RNA are synthesized and "read" in the direction from 5' to 3', the beginning of a gene or RNA is often referred to as the "5'" end.

The nitrogenous bases are of two fundamental types, pyrimidines and purines; they differ in that pyrimidines consist of a six-membered ring and purines of joined five- and six-member rings. There are two purines and three pyrimidines of interest to our discussion (Fig. 2). The two purines, A and G, are contained in both DNA and RNA. The pyrimidines of DNA are C and T; the third pyrimidine, U, substitutes for T in RNA. RNA also differs from DNA in that its sugar is ribose rather than deoxyribose (hence, ribonucleic acid and deoxyribonucleic acid).

The structure of a DNA double helix is determined by two physical factors: (1) the "hydrophobic effect"¹² of excluding water from adjacent nucleotide bases stacked one on top of the other and (2) the noncovalent hydrogen bonding of purines and pyrimidines in one strand to the pyrimidines and purines, respectively, of a second strand (Fig. 3). Base stacking accounts for the stability of the double helix in water; the number and precise orientation of hydrogen bonding accounts for the specificity of complementary strand interactions. With rare exceptions, the purine G pairs only with the pyrimidine C; the purine A pairs only with the pyrimidine T in DNA or U in RNA. Thus, for example, a short strand of nucleotides, 5'-A-C-G-T-A-3', would bind to a complementary sequence, 3'-T-G-C-A-T-5'. In typical genes, only one strand actually encodes a protein product; the other strand is a noncoding complementary strand.

This two-stranded structure is critical to the process of DNA replication for cell division in which the two strands separate and then form templates for the synthesis of new noncoding and coding strands, respectively. Similarly, the coding strand of DNA provides the template for synthesis of a complementary RNA molecule, which will ultimately be translated into a protein.

The structure of genes and production of mRNA

One characteristic of gene organization distinguishing eukaryotes from prokaryotes is that the latter have achieved a meticulous, miniaturized economy of or-

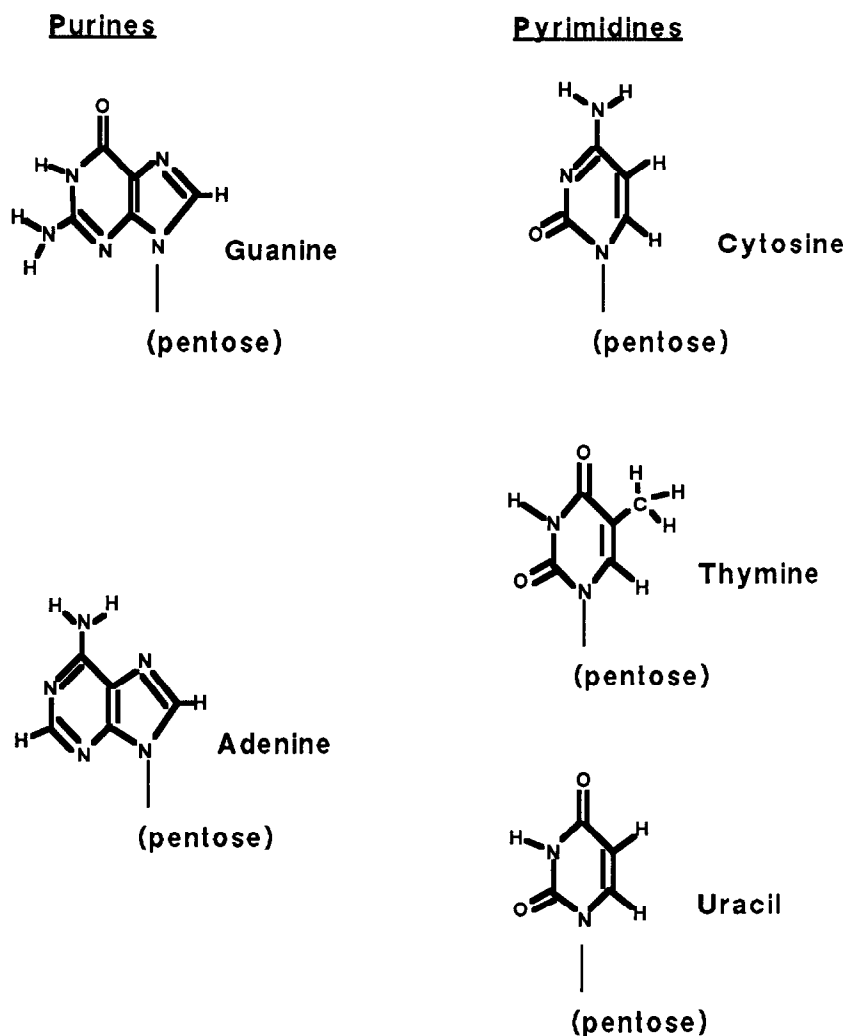


FIG. 2. The nitrogenous bases of DNA and RNA. The genetic code is composed of just four "letters." DNA and RNA differ in base usage in that uracil substitutes for thymine. Adenine forms exclusive base pairs with thymine/uracil, whereas guanosine forms base pairs with cytosine. The positions of attachment of the bases to ribose or deoxyribose sugars are indicated.

ganization. In contrast, eukaryotes are slovenly and would rather add new rooms than take out the trash. But out of this chaos, "higher" metazoans have evolved a regulatory and developmental complexity that depends on a patchwork of physical organization and genetic controls. The genetic controls, governing especially transcription, are scattered in and around the transcribed gene itself, often at great distances from the gene. Moreover, the actual coding sequence of the gene, those sequences that ultimately will be translated into a protein, are also scattered within the domain of transcription. These segments, termed "exons," are interspersed with stretches of DNA called intervening sequences or "introns," that do not encode protein and must be removed before the mature mRNA can be translated. Although regulatory sequences have

been identified within introns, the function of introns, generally, and, indeed, the fundamental reason for their existence, is essentially unknown. One possibility is that they are simply ancient fossils held over from the earliest times of the evolution of life.^{13, 14}

Eukaryotes may have escaped the extreme selective pressures that forced prokaryotes to streamline their genetic structure and therefore may have retained a primordial patchwork structure. Moreover, the concept of "genes in pieces" as an ancestral condition may help us understand the evolution, particularly of the immune system, in which both immunoglobulin and TCR genes must be pieced together in each Ag-specific clone of B or T cells (see below).

Interestingly, the exons of a gene are often related to the domain structure of the protein product.¹⁵ This

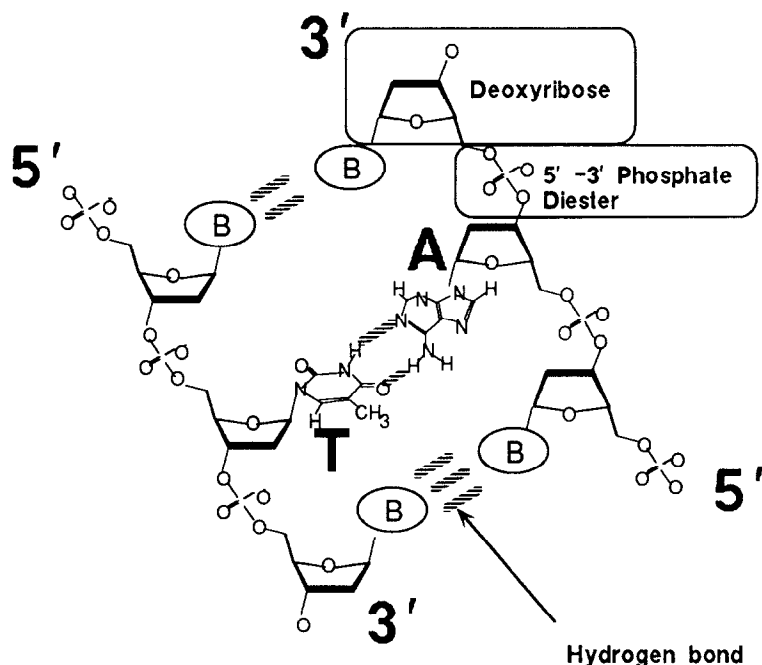


FIG. 3. The structure of DNA. The fundamental principle of complementarity depends on the "antiparallel" orientation of two strands of DNA and the formation of specific hydrogen bonds between nucleotide bases. The stability of the double-stranded structure is provided by "stacking" interactions between adjacent base pairs that can form when the two strands are twisted to form a helix. For simplicity, only a single base pair is illustrated in detail; a typical eukaryotic DNA molecule consists of two stands each containing *hundreds of millions* of nucleotides.

is observed, strikingly, for most of the members of the immunoglobulin superfamily of genes, including, among others, the Ab molecules themselves, TCRs for Ag, and MHC molecules. In each case, protein domains, defined by a single intrachain disulfide bond, directly correspond to individual exons within the gene.

Transcriptional organization and regulation

With fascinating but rare exceptions, the minimal size of a protein-coding gene is that required to code directly for the amino acid sequence of the protein. Since three nucleotides are required to encode each amino acid, the minimal length of an mRNA encoding a typical protein of 40,000 daltons (about 330 amino acids) is approximately 1000 base pairs. In fact, the average mRNA is about 1500 bases long, since typical mRNAs contain both 5' and 3' untranslated regions, some of which are involved in regulatory control of mRNA stability and translation (Fig. 4).¹⁶ However, the typical gene producing an average mRNA is five to 10 times larger, containing numerous and often large introns (Fig. 5). Just upstream and downstream from the transcriptional domain are the signals for initiating and terminating transcription.

This minimal size is still not sufficient to achieve

appropriate expression. This domain contains the sequence encoding the mature mRNA as well as regulatory sequences, called "promoters," "operators," and "terminators" in bacteria, that specify to the transcriptional apparatus of the nucleus where and when to start and stop transcription. Although many of the identified DNA sequences interact with protein factors regulating transcription per se, by interacting with the enzymatic machinery of transcription, other sequences may interact with the physical structure of chromatin, indirectly affecting the activity of the DNA template over long distances.

It has not been possible in eukaryotes to define contiguous promoter/operator sequences as were classically demonstrated by Monod and Jacob^{16a} for *Escherichia coli*. Instead, these functions in eukaryotes are distributed over a number of specific short nucleotide sequences, often called "boxes," including the "TATA" boxes, "CAAT" boxes, enhancers, silencers, and hormone response elements.¹⁷ These regulatory sequences or elements are often clustered in the region just 5' ("upstream") of the transcription start site but may also be found within the transcribed region and even downstream of a gene (Fig. 5). Moreover, many of these elements may be separated by considerable distance from the regulated gene. For example, the

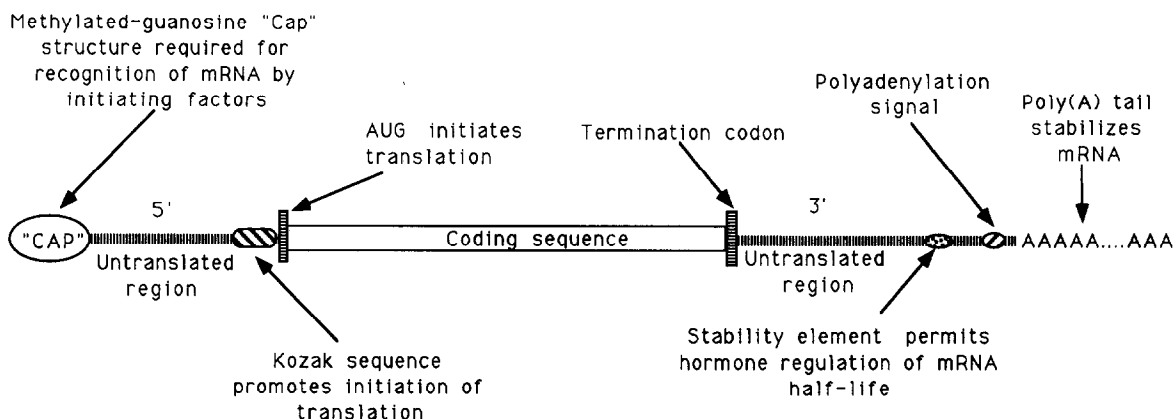


FIG. 4. Structure of a typical mRNA. In addition to the coding sequence, mRNAs contain additional sequences that are involved in the control of mRNA stability and translational activity. Most eukaryotic mRNAs contain considerable 5' and 3' untranslated sequences that subserve no presently known function.

tissue-specific expression of both κ and λ light chain genes is controlled by an enhancer located several thousand base pairs downstream of the gene.¹⁸⁻²⁰ All these genetic control elements function by binding special proteins, termed transcription factors, that recognize and bind to specific DNA sequences and are also capable of modifying the activity of the transcription complex, an assembly of proteins on the DNA that can initiate transcription in concert with RNA polymerase.²¹

The element most closely corresponding to the prokaryotic promoter is the TATA box. This element, with a consensus nucleotide sequence "TAATA," appears to specify the start of transcription 20 to 30 base pairs *further downstream* and, in most genes, is absolutely required for transcriptional activity. As with its prokaryotic analogue, this element has tight requirements for orientation and distance from the start site. Curiously, some genes function quite well without a TATA box; these genes often exhibit multiple transcriptional start sites.²² Many genes have a second orientation-specific signal, the CAAT box (similarly named for its consensus sequence), located several hundred base pairs upstream of the start site. These two positive elements respond to constitutive factors present in all tissues.

Enhancers, in contrast to promoters, stimulate transcription through a mechanism, presently unknown, that is independent of orientation and distance. In addition to these defining properties, the biology of enhancers encompasses several important general features. First, their activity depends on interactions with appropriate sequence-specific transcription factors. Second, some enhancers are active only in the context of other associated enhancer elements, and thus, their

activity may require cooperative binding of multiple enhancer-binding factors.^{23, 24} Third, more than one species of transcription factor may bind to a single enhancer; thus, the activity of the enhancer often reflects competition among multiple transcription factors.²⁵ Fourth, the acute activity of transcription factors may be regulated by association with inhibitory and/or modifying factors and/or by covalent post-translational modifications, often reflecting signal transduction from a hormone, cytokine, or other stimulus. In addition, the expression of gene-encoding transcription factor is often itself regulated. Finally, enhancer-binding transcription factors are often members of closely related gene families; different members of the same family may display distinct patterns of developmental and regulated expression, sequence specificities, and ability to interact with other transcription factors.

A classic example of hormonal regulation of immune responses is that of glucocorticoid. Glucocorticoids mediate their effects by binding to specific intracellular steroid receptors that then bind to the glucocorticoid response element, thereby controlling the rate of transcription.²⁶ Note, however, that glucocorticoids and other stimuli may also affect the levels of specific mRNA through posttranscriptional mechanisms (see below).¹⁶ Like other transcription factors, the glucocorticoid receptor is just one member of the family of steroid receptors.²⁷

Another well-known enhancer, the Ig κ enhancer, illustrates the several properties of enhancers listed above. The transcription factor first identified as binding to the immunoglobulin κ light-chain enhancer was NF- κ B.²⁸ Originally defined as a nuclear factor critically required for immunoglobulin- κ gene transcrip-

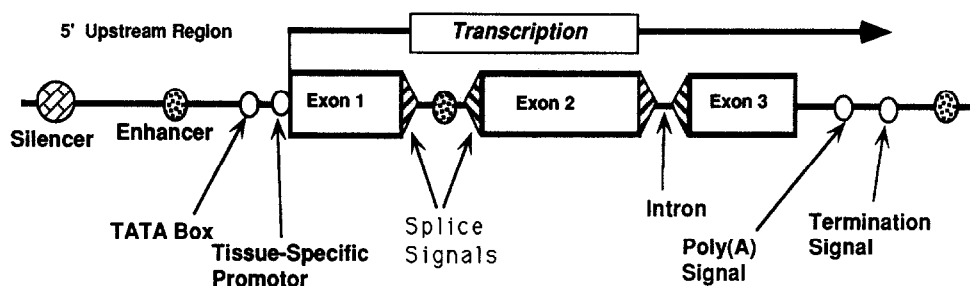


FIG. 5. Structure of a typical eukaryotic gene. Sequences that will appear in the mature mRNA (including 5' and 3' untranslated regions) are contained within exons. Signals required for the identification and removal of introns are contained at intron termini; additional sequences governing quantitative and developmental aspects of transcription are contained upstream, downstream, and sometimes within the gene itself. The structures are not drawn to scale.

tion in B cells,²⁹ κ -like enhancers have been found in a wide variety of genes expressed in many cell types and in several significant viruses, including cytomegalovirus³⁰ and HIV-1.^{31, 32} NF- κ B belongs to a family of closely related genes.³³⁻³⁷

The enhancing activity of NF- κ B is under complex regulation. Thus, NF- κ B is usually present in the cytoplasm as a 50 kd subunit complexed to both an inhibitory factor I- κ B,³⁸ and a modulating factor, p65.³⁹ NF- κ B can bind DNA in the absence of p65, whereas p65 appears to regulate which DNA sequences are recognized by NF- κ B. Moreover, p65 is required for I- κ B-inhibitory binding. Thus, lipopolysaccharide induction of B cells results indirectly in the release of I- κ B from the complex and transport of a (NF- κ B:p65)₂ tetramer to the nucleus in which it binds to DNA and enhances transcription of the target gene.

NF- κ B is also present in a constitutively active form in most cell types in which it is required for expression of class I MHC antigens.^{40, 41} This form of NF- κ B is also believed to interact with transcription factors binding to an adjacent ISRE required for interferon-mediated induction of class I MHC expression.⁴² A similarly juxtaposition of immunoglobulin- κ -like enhancers and ISRE sequences is found in the interferon- β ⁴³ and double-stranded RNA-dependent (2'5') oligoadenylate synthetase⁴⁴ genes. Similar sequences mediate the transcriptional induction of HLA class II antigen expression in endothelial tissues and most if not all other genes that respond to either type I or type II interferons.⁴⁵ Transcriptional regulation via the ISRE is believed to reflect the activity of another family of transcription factors.^{46, 47}

Silencers are negative regulatory factors that inhibit transcription of their associated genes. Like enhancers, silencers are effective in either orientation and relatively independently of distance. Since some en-

hancers can have the properties of silencers in certain situations, it is likely that a given DNA sequence can behave as either a silencer or an enhancer, depending on which transcription factors they are binding. Important silencers are present in immunoglobulin⁴⁸ and TCR⁴⁹ genes and are important in repressing immunoglobulin gene transcription in inappropriate tissues, such as T cells and fibroblasts. Silencers may also be involved in the phenomenon of allelic exclusion in which an active immunoglobulin gene represses the expression of other immunoglobulin genes in the same cell (see below).

A program of DNA rearrangements⁵⁰

Ab gene rearrangement is a special form of *site-specific recombination* in which enzymes called *recombinases* recognize specific DNA sequences and catalyze the physical rearrangement of often *unlike* sequences. Highly specialized systems of site-specific recombination have evolved in many organisms to handle host-parasite or sexual interactions. Examples include the integrative lysogeny of bacteria by bacteriophages,⁵¹ retrovirus integration into host DNA,⁵² mating-type (sex) switching in yeast,⁵³ antigenic phase variation in trypanosomes,⁵⁴ and the formation of active Ab and TCR genes.⁵⁰ All these systems involve a form of phase or clonal variation in which a gene or function is turned "on" or "off."

A mature Ab or TCR gene encodes N-terminal *variable* (V) domains required for Ag-binding, and C-terminal *constant* domains that carry out other functions, such as Fc receptor binding. As encoded in the sperm and egg and most somatic cells, immunoglobulin and TCR genes are defective. They are broken in three or more pieces or modules that, for a variety of reasons, cannot be joined through RNA splicing mechanisms and thus cannot, on their own, be expressed as functional mRNA. The germline contains

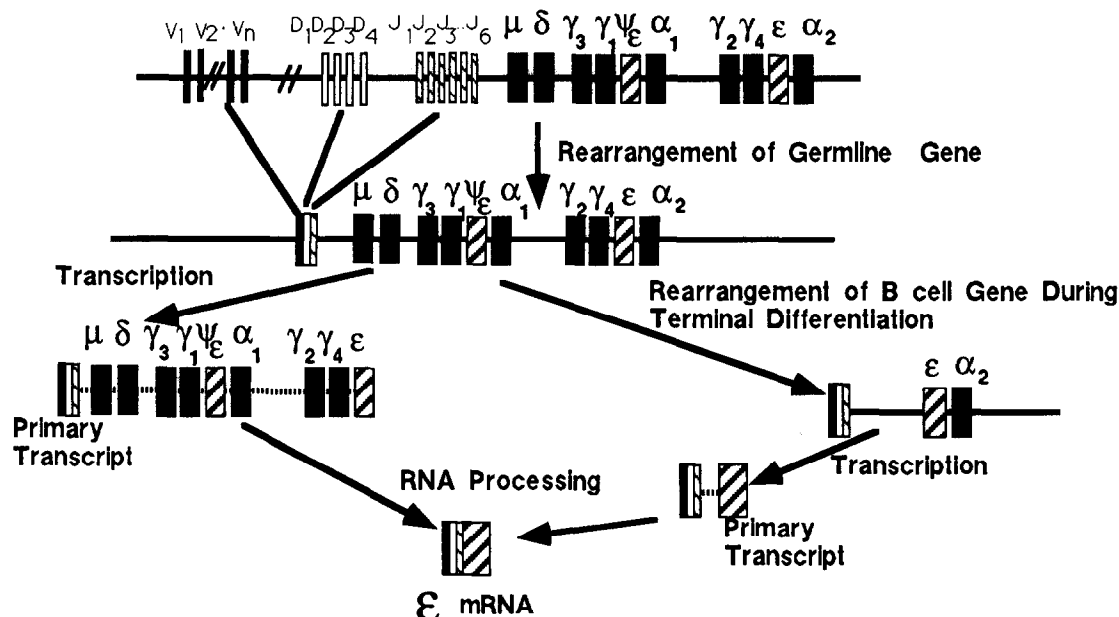


FIG. 6. Expression of IgE. Several decisive mechanisms lead to the expression of a given immunoglobulin heavy chain isotype. Rearrangements of the V, D, and J regions occur in pre-B cells. Isotype choice can occur at the level of splicing (*left pathway*) or DNA rearrangement (*right pathway*); these steps occur in mature and terminally differentiating B cells.

a large number of distinct but interchangeable V modules and a small number of functionally distinct but interchangeable C modules. In addition, between the V and C regions are two sets of interchangeable modules, termed D (for "diversity") and J (for "joining") segments, that must also be incorporated into a functional gene. To illustrate this process, we will follow the course of heavy chain Ab gene rearrangement; light chain and TCR gene rearrangements are quite similar.

An immature B cell expresses neither heavy chain nor light chain immunoglobulin proteins or mRNAs, and their genes are in the germline configuration (Fig. 6). Constant μ mRNA of dubious function is sometimes synthesized in these cells, indicating the presence of a cryptic promoter and limited transcriptional activity of μ region chromatin. As development progresses, chromatin containing the variable regions becomes accessible to DNA binding proteins, but variable region transcription is limited because of the fact that several important enhancers are located in the constant region, too far away to exert their "location independent" function.

At this juncture, specific recombinases recognize specialized sequences, *joining signals*, located at the 3' end of a germline V module and at the 5' end of one of several available D modules. The recombinases catalyze the fusion of the two modules by a mecha-

nism that removes the recombinogenic signals so that the new fused VD module has lost both original signals. The 3' end of the D module contains another recombination signal, however, which catalyzes a second fusion via a reciprocal signal at the 5' end of one of several available J modules, creating a VDJ module. Rearrangement ceases because the VDJ module contains no further recombinogenic signals. The choice of V, D, and J modules is essentially random, occurring before exposure to Ag. Two additional random processes, occurring at the time of rearrangement, contribute significantly to the generation of Ab: imprecise removal of the recombinogenic signals and insertion of extra base pairs into the DNA joint. The sequences of the V, D, and J modules all appear in the protein coding region of the mature mRNA, and all contribute to the Ag-binding sites of the Ab protein. Since there are hundreds of V modules, half a dozen D and J modules, and hundreds of possible modifications to the joints themselves, the combinatorial possibilities number in the tens of thousands, accounting for a considerable proportion of Ab diversity. The remainder of Ab diversity is largely contributed by the combinatorial effect of the Ab light chain, derived through a similar process.

After this round of rearrangement, the VDJ module is still not connected with the first available C region, encoding the μ heavy chain of IgM. The exon se-

quences of the VDJ modules are spliced together post-transcriptionally to the exon sequences of the C region, resulting in mature μ mRNA.

After a particular rearrangement is complete, the new gene can be transcribed, spliced, and translated into the V domain of a specific Ab. At this point the process of clonal selection can begin in which immature B cells expressing Ag-reactive surface IgM can be stimulated by T cells and cytokines to mature and proliferate.

Isotype switching⁵⁵

There is an additional complexity in the structure of the immunoglobulin heavy chain genes that helps in understanding the control of immunoglobulin isotype production. For example, an Ab response, initiated by some specific allergen, is likely to begin as an IgM response and then, by a process of immunoglobulin class switching, converts to synthesis of other isotypes, such as IgG or IgE. This switching of Ab heavy chain class does not involve any change in the specificity of the Ab molecule in question because the VDJ module of the heavy chain gene, encoding the Ag-binding portion of the molecule, remains intact. The exons that encode C domains of the various heavy chain isotypes are arrayed in a linear fashion downstream (or 3') from the V region exons, sequentially encoding IgM, IgD, each of the IgG subclasses, IgE, and IgA heavy chain C modules. Thus, a single lineage of B cells can produce IgM, IgG, IgA, or IgE Abs, all with identical VDJ modules, but bearing distinct heavy chain isotypes.

In most cases, however, any one cell can express only a single heavy chain at one time. To permit developmental expression of different heavy chain isotypes, B cells undergo another kind of DNA rearrangement that connects a functional VDJ module with a new C isotype module. Thus, in the production of a functional IgE heavy chain gene, *heavy chain switching* deletes all of the DNA from the beginning of the μ locus to the beginning of the ϵ locus. Switching is mediated by *homologous recombination* between *switch regions* located 5' of each heavy chain locus. Under the influence of various cytokines, B cells can undergo another kind of DNA rearrangement that connects a functional VDJ module with a new C module. The regulation of isotype switching is at least partially under the control of cytokines.⁵⁶⁻⁵⁸

Initiation of transcription

The initiation of transcription involves the identification of the transcription start site on a gene by the

initiating transcription complex and the commencement of transcription by RNA polymerase and associated transcription factors.⁵⁹ Eukaryotic mRNAs are initiated with a special methylated-guanosine nucleotide called the "CAP" (Fig. 4), which is required later for recognition by the protein factors responsible for recruiting mRNAs to the ribosome and initiating translation. Thereafter, the transcription complex traverses the entire gene, which is typically several thousand base pairs long, but may reach as high as several *hundred* thousand base pairs. As mentioned early, the bulk of the primary transcript consists of intron sequences that do not encode protein but may contain enhancer elements and other regulatory sequences.

Termination and polyadenylation

Transcription by the initiation complex continues past the end of the mature mRNA; termination is a poorly understood but active process. It appears that special terminator sites are specifically recognized by the transcription complex, at which point the complex disengages from the DNA template, and the primary transcript is released. Termination itself does not create the 3' end of the transcript. Rather, coupled with, or soon after termination, the transcript is cleaved just downstream of a *polyadenylation motif* on which a track of about 200 adenylate residues is added at the cleavage site.⁶⁰ This poly (A) "tail" is involved in stabilizing the mature transcript in the cytoplasm.⁶¹

Posttranscriptional processing

The completed transcript passes from the transcription complex into a much larger structure, consisting of proteins and special small nuclear RNAs. This organelle, termed the "spliceosome," carries out the complex and still mysterious steps of splicing, the removal of introns and ligation together of the ends of exons, all in the proper order, to produce a mature translatable mRNA.⁶²

The spliceosome scans the transcript to identify exons and then splices two exons together. As in so many other processes, special sequence signals are recognized by the spliceosome apparatus to identify the ends of exons. This process continues until no more introns remain, after which the mature mRNA is disgorged from the spliceosome and passed through a pore in the nuclear envelope into the cytoplasm.

The regulated production of μ_m and μ_s mRNA is dependent on the relative efficiencies of μ_s poly (A) site usage and the C μ 4-to-M1 splice.⁶³ Transcripts of many genes can be terminated, polyadenylated, or spliced in different ways to generate distinct mRNAs encoding proteins with different properties. The ex-

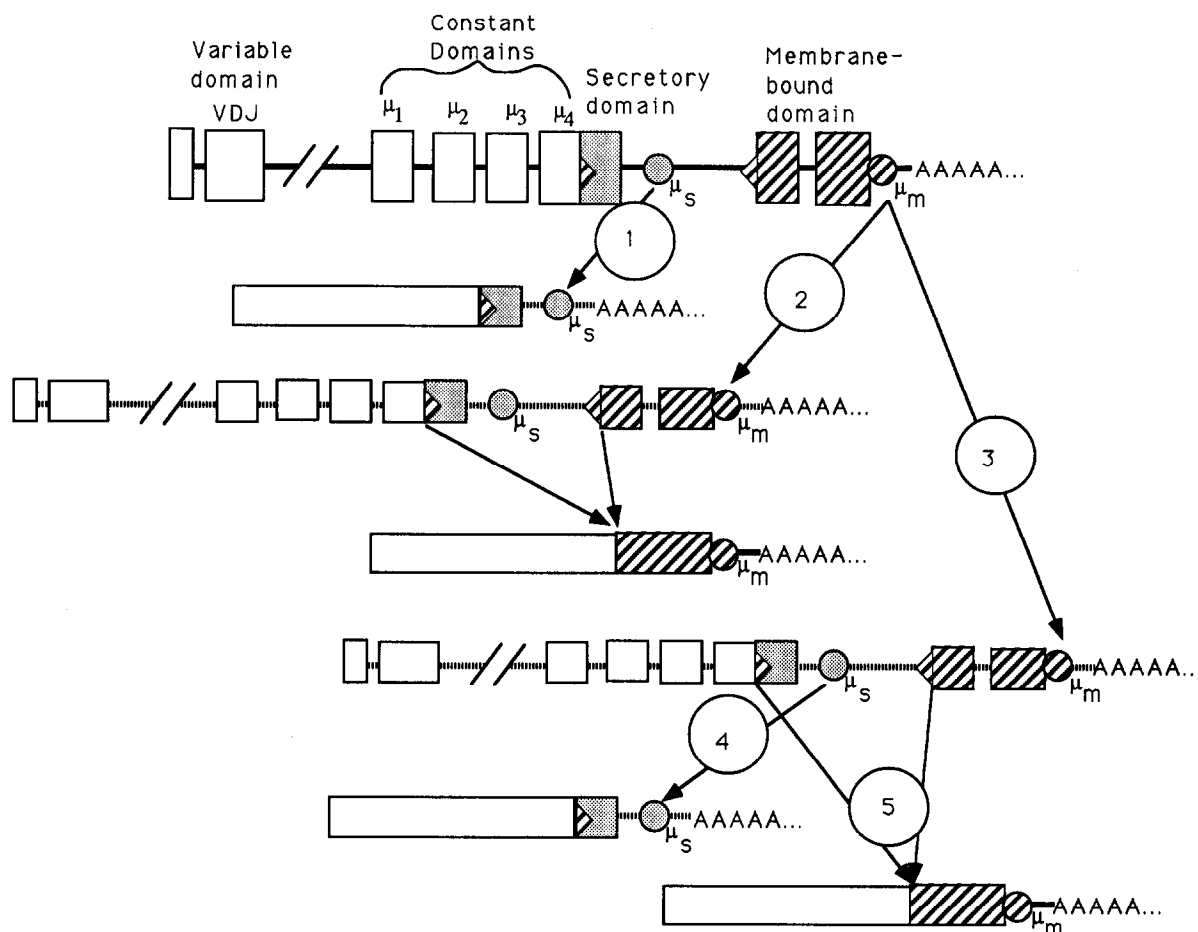


FIG. 7. The mechanism of regulating membrane-bound versus secreted immunoglobulin. The two forms of immunoglobulin can be generated from a single gene by one of two competitive mechanisms. In some cells, cleavage and polyadenylation at the μ_s site irreversibly generates secreted immunoglobulin (*pathway 1*). Cleavage and polyadenylation at μ_m generates a longer transcript still containing μ_s . In some cells the rate of secondary use of μ_s is much slower than the rate of splicing to remove the last intron (*pathway 2*). In most B cells, however, μ_s is inefficiently used to generate the primary transcript; most primary transcripts use μ_m . Production of secreted immunoglobulin then depends on competition between the rates of secondary cleavage at μ_s (*pathway 4*) and splicing to remove μ_s (*pathway 5*).

ample of alternate termination choice of IgM (μ) heavy chain transcripts has been well studied and most likely illustrates the general mechanism of regulating whether immunoglobulin heavy chains are secreted or membrane bound. The μ gene has two polyadenylation sites and two termination sites (Fig. 7). The μ_s site lies within the last intron of the gene, whereas the μ_m site lies within the last exon. The last exon encodes a peptide that anchors the immunoglobulin in the membrane. When transcripts terminate within the *intron* at the μ_s site, the last exon is not expressed and a secreted Ab is produced. In contrast, if transcription bypasses the μ_s site, a longer transcript is generated containing two potential polyadenylation sites.

The decision to produce secreted Ab is regulated;

the mechanisms by which the B cell makes this commitment is controversial. In a few plasmacytomas, termination downstream from the μ_m polyadenylation site assures that cleavage will occur at μ_m and that splicing will include the last exon in the mature mRNA. Termination and cleavage appear to be coupled in such cells, and the choice between producing membrane-bound versus secreted IgM depends largely on the activity of the μ_s site.

However, in most other plasmacytomas and B cell lines, secreted IgM can still be generated from transcripts that initially polyadenylate at μ_m (Fig. 7). Note that transcripts with μ_m still must be spliced to remove the last intron, and until they do so, μ_s still is present. Splicing the last intron out will commit the transcript to encoding membrane-bound IgM. A second cleav-

First Base	A	G	C	U	Last Base
A	Lysine	Arginine	Threonine	Isoleucine	A
	Lysine	Arginine	Threonine	Methionine/ (initiate)	G
	Asparagine	Serine	Threonine	Isoleucine	C
	Asparagine	Serine	Threonine	Isoleucine	U
G	Glutamate	Glycine	Alanine	Valine	A
	Glutamate	Glycine	Alanine	Valine/ (initiate)	G
	Aspartate	Glycine	Alanine	Valine	C
	Aspartate	Glycine	Alanine	Valine	U
C	Glutamine	Arginine	Proline	Leucine	A
	Glutamine	Arginine	Proline	Leucine	G
	Histidine	Arginine	Proline	Leucine	C
	Histidine	Arginine	Proline	Leucine	U
U	(terminate)	(terminate)	Serine	Leucine	A
	(terminate)	/Selenocysteine Tryptophan	Serine	Leucine	G
	Tyrosine	Cysteine	Serine	Phenylalanine	C
	Tyrosine	Cysteine	Serine	Phenylalanine	U

FIG. 8. The genetic code. Most codons are interpreted by the translation machinery in one way only; however, each amino acid may be encoded by more than a single codon. Note that the redundancy or "degeneracy" of the genetic code largely involves the third or "wobble" position of the codon. Context-dependent translations are shown in **bold**. Note that initiation always inserts a methionine (or formyl-methionine for bacteria, mitochondria, and chloroplasts). Minor differences in codon translations are found in some mitochondria, chloroplasts, and protozoans.

age/polyadenylation reaction can still occur at μ s, however, committing the transcript to encoding membrane-bound IgM. In this case, regulation appears to depend on competition between splicing and a second cleavage.

The genetic code and the genetic decoding ring

The mature mRNA is exported to the cytoplasm to be translated into protein; the latter may be the most complex process of all. The information contents of genes and mRNA are determined by the linear sequences of their bases. Thus, if the collection of all genes in an organism (the genome) is viewed as a library and the genes encoding specific proteins as the books, the actual information is determined by an alphabet of only four letters (the ribonucleotides A, C, G, and U), while triplets of nucleotides (called *codons*) are the words.

Each codon specifies a particular amino acid (Fig. 8). There are 64 possible combinations of the four

letters A, C, G, and U of RNA, and thus, in principle, 64 different amino acids could be specified with this genetic code. Until recently, only 20 amino acids were known that were actually encoded by mRNAs. Recently, it has been determined that a twenty-first amino acid, selenocysteine, is also encoded by mRNAs,⁶⁴ and minor variations in the code are found in mitochondria and certain protozoans.⁶⁵ Other amino acids, such as phosphotyrosine and phosphoserine, are produced posttranslationally by specific enzymes.

Since 64 codons encode 21 amino acids, it is not surprising that there is substantial redundancy or degeneracy in this genetic code. For example, four codons, CCU, CCC, CCA, and CCG, are all translated by the ribosome as the amino acid proline.

Punctuation marks are also important in understanding the code. The *initiator* codon is usually AUG, and occasionally GUG. As a consequence of the mechanism of peptide chain initiation, this codon also specifies the amino acid methionine (or formyl-methionine in bacteria and mitochondria). In eukaryotes, the ini-

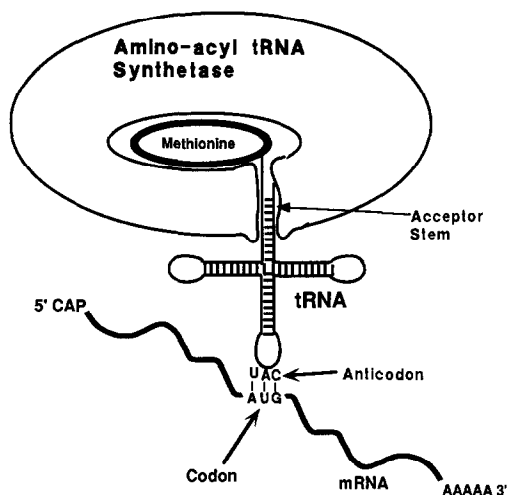


FIG. 9. The genetic decoding ring. Deciphering the genetic code requires two key steps. Codons are identified by the ribosome and are then paired by complementarity with the anticodons of amino-acylated ("charged") tRNAs; the ribosome is also responsible for joining the incoming amino acid to the nascent polypeptide chain. The proper charging of tRNAs, however, requires the exquisite specificity and proofreading abilities of amino-acyl tRNA synthetases. Each one of these enormous enzymes recognize a particular species of tRNA via the acceptor stem and joins it to the correct amino acid. Synthetases can also identify and destroy incorrectly paired amino-acyl tRNAs.

tiator codon is additionally signalled by the presence of a "Kozak" sequence just upstream of the initiator codon⁶⁶ and sometimes by additional sequences downstream of the initiator codon.⁶⁷ Three codons are often referred to as "nonsense" codons but can also be viewed as "termination" codons, and, for historical reasons, were named the *ochre* (UAA), *amber* (UAG), and *opal* (UGA) codons. Selenocysteine is also encoded by the *opal* (UGA) codon. There are no commas in the genetic code.

Any code, hence its information content, is indistinguishable from random noise unless an appropriate decoding device can meaningfully translate it. Perhaps even more central to life, and more ancient than the genetic code itself, is the genetic decoding ring. This decoding device or translation mechanism consists of a set of specialized tRNA molecules, a corresponding set of tRNA synthetase enzymes, and the ribosome. A tRNA is a tightly coiled structure of about 73 base pairs, often drawn as the familiar cloverleaf (Fig. 9). At the base of the tRNA is the anticodon, known to interact with the codon of an mRNA. At the top of the structure is the *acceptor stem*. A *charged* tRNA is covalently linked at its 3' hydroxyl group with the amino group of an amino acid.

Two features of tRNAs are essential to the processing of deciphering the 4 base/64 codon language of DNA/RNA into the 21 amino acid lexicon of proteins. The first of these features makes use of the same mechanism of nucleotide base complementarity used in replication and transcription. Thus, each kind of tRNA has a distinct *anticodon*, a triplet of bases complementary to a particular codon.

Second, tRNAs for each amino acid are recognized by a particular enzyme, amino-acyl tRNA synthetase. Each of the amino-acyl tRNA synthetases is a large, complex protein required to recognize a particular tRNA and a particular amino acid with great fidelity and mediate their correct attachment. Some synthetases can even recognize incorrectly charged tRNAs and uncharge them. It should be clear that tRNA synthetases provide the crucial span between the digitized information of the DNA/RNA world and the "analogue" information of the protein world. It has long been considered, with some experimental support, that tRNA synthetases mediate this information transfer by recognizing the anticodon itself. More recent evidence indicates that tRNA synthetases largely ignore the anticodon completely but instead recognize unique features of the acceptor stem⁶⁸ of each tRNA.

The process of protein translation occurs on cytoplasmic organelles termed ribosomes. (Mitochondria, the powerhouses of the eukaryotic cell, contain their own DNA and their own unique ribosomes.) The ribosome is a massive organelle consisting of two large and probably enzymatic RNAs and numerous ribosomal proteins.

The initiation of translation requires the activity of a cascade of initiation factors that recognize the 5' terminal CAP structure of mRNAs and a special tRNA, initiator tRNA, that is attached to the amino acid methionine. As the 5' end of the mRNA is fed into the reading head of the ribosome, the ribosome scans the mRNA sequence for a special sequence that serves as a warning flag, as it were, that the coding region is about to be read. Charged initiator tRNA, accompanied with its own special protein cofactors, enters the ribosome and is used by the ribosome to scan for the first available initiator codon. The initiating codon then defines the *reading frame* of the mRNA, and the initiating methionine becomes the N-terminus of the nascent peptide. The mRNA is ratcheted forward exactly three positions to bring the next codon into the reading head. As each codon enters the reading head, the ribosome samples available charged tRNAs until the correct anticodon is found. Then, a peptidyl-transferase activity concertedly cleaves the covalent bound between the amino nitro-

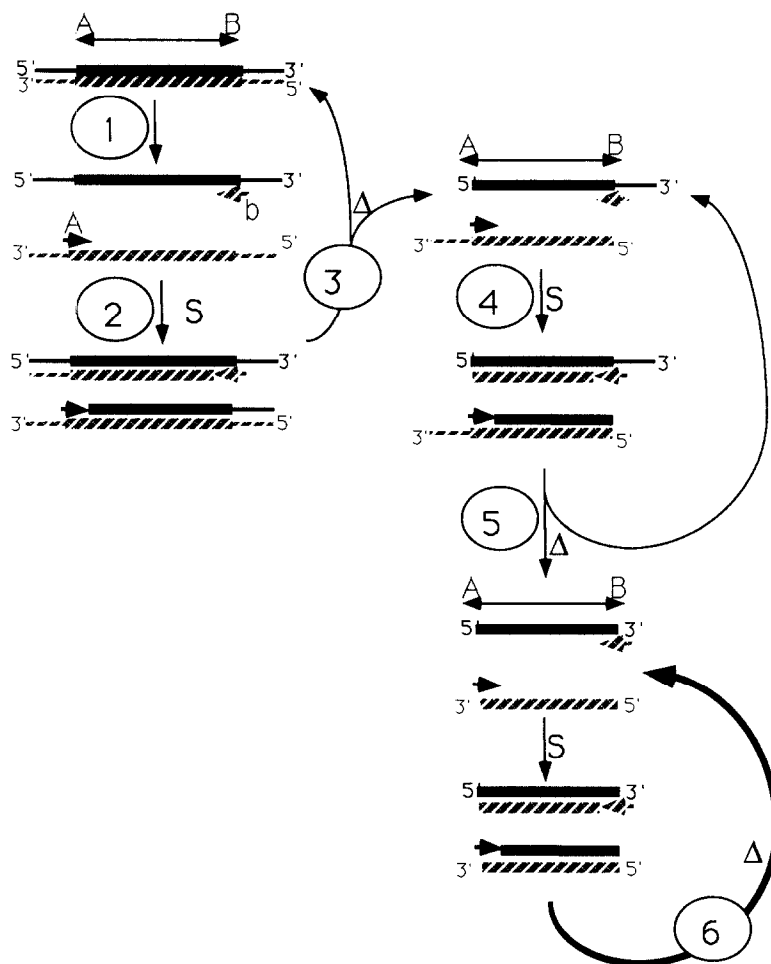


FIG. 10. The PCR. This powerful technique depends on the exponential replication in vitro of specific DNA sequences orchestrated by synthetic oligonucleotide primer sequences. The target sequence (*segment A ↔ B*) is often extremely rare and imbedded in irrelevant flanking sequences. In step 1, double-stranded DNA is heat denatured to separate the two strands and allowed to anneal to specific primers. Primer A is complementary to the sequences at the 3' end of the top strand, whereas primer b is complementary to the sequence the 3' of the bottom strand. In step 2, the heat-stable Taq polymerase enzyme is used to synthesize new DNA on both strands. Note that this enzyme will not synthesize new DNA in the absence of a correctly base paired primer. In step 3, the double strands are heat denatured again. After cooling and reannealing with primer, both the original strands and the new daughter strands are able to serve as templates for an additional round of synthesis. In step 4, the products of step 2 serve as templates. The new products of step 4 are short sequences corresponding exactly to the target segment A ↔ B. Thereafter, each cycle (step 6) doubles the number of A ↔ B segments; after 30 cycles, there are 2^{30} or one billion specific daughter molecules.

gen of the amino acid and the tRNA and forms a new peptide linkage with the carboxyl terminus of the nascent peptide. The used tRNA is released, and the process continues until a termination codon is found.

Termination and codon context^{69, 70}

Although the initiator codon interacts with a unique tRNA and encodes an amino acid, there are no termination tRNAs; when a termination codon is found,

the ribosome simply waits for a nonexistent tRNA to come in. A similar situation also occurs in cases of amino acid starvation, when a particular charged tRNA may become scarce. If no charged tRNA arrives, the ribosome may decide to terminate, to wait longer, or to signal an amino acid deficiency to the cell. The mechanism of this decision is not well understood but appears to depend on *codon context*. This is a subtle feature that is critical in certain situations

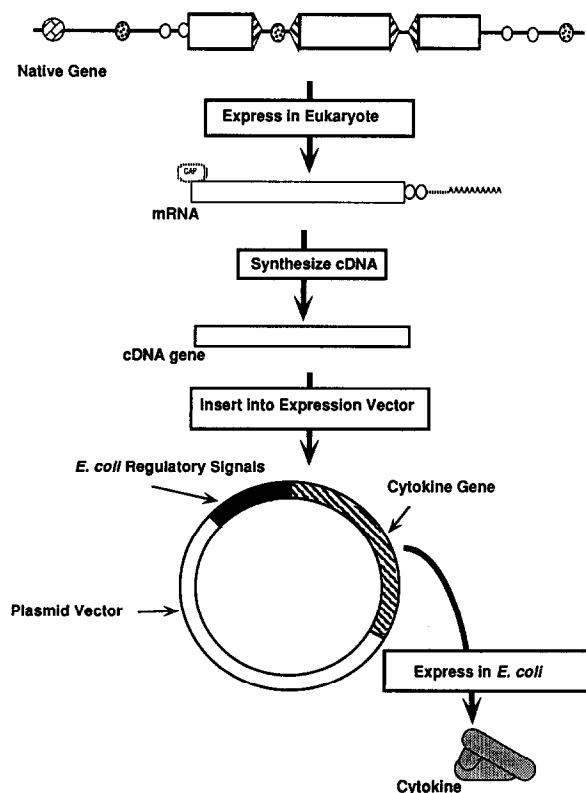


FIG. 11. Production of recombinant cytokines. The desired cytokine gene is expressed first in its eukaryotic source. This correctly removes intron sequences to generate a mature mRNA that is copied with the enzyme reverse transcriptase to generate a "copy DNA" (cDNA). The cDNA is then cloned into a new vehicle that contains regulatory sequences designed for efficient expression in the recombinant host (such as *E. coli*).

and emphasizes the role of the decoding device in making sense of genetic information. For example, the decision to translate the opal codon as selenocysteine instead of as a terminator is probably decided by sequences downstream of the opal codon itself, as detected by the ribosome. In addition, ribosomes are capable of shifting the reading frame at particular positions in some mRNAs to generate alternate proteins.

Posttranslational processing

As the protein passes from the ribosome into the endoplasmic reticulum, it also leaves the strict province of molecular biology and of this discussion. The protein undergoes posttranslational modification, such as phosphorylation, attachment of carbohydrate groups or other radicals, and cleavages before transport to its cellular (or extracellular) destination. Many of these modifications depend, as we should expect, both on the presence of specific amino acid sequence signals in the protein and on the context of each signal.

Genetic engineering^{71, 72}

Genetic engineering techniques have transformed molecular biology from being simply an academic discipline into a mainstream industry, with major current applications in agriculture and medicine. There are far greater potential applications in the future, for example, in the synthesis of "designer" enzymes.

The extraordinary power of recombinant DNA technology derives directly from our capacity to manipulate DNA and RNA in the test tube and to reexpress that information as proteins in vitro or in living cells of our choosing. Many of the important products at this point are themselves the reagents used in molecular biology, such as restriction enzymes and DNA polymerases. Other products include cytokines and other proteins, which can now be produced in great quantity and exceptional purity for therapeutic and other medical uses. Of great importance to allergists, especially, is the ability to study the function of cloned IgE receptors^{73, 74} and the recent isolation through molecular cloning techniques of specific allergens.^{75, 76} This development has permitted identification of specific IgE-reactive epitopes involved in the acute allergic reaction.⁷⁷⁻⁸⁴ Moreover, this approach will soon enable the molecular identification of allergenic epitopes reactive with T cells responsible for the initiation of allergy.⁸⁵

With the use of certain highly variable gene sequences and the use of appropriate enzymes that recognize and cleave specific DNA sequences (restriction endonucleases), it is possible to obtain DNA fingerprints of restriction fragments characteristic of any particular gene derived from a specific individual. By electrophoresis of such fragments and hybridization to complementary radioactive probes (Southern blotting), a restriction fragment length polymorphism analysis can be performed. This permits following the inheritance of a specific polymorphic gene within a family or, for example, in a cohort of patients with a particular disease.⁸⁶ Other enzymes (polymerases and transcriptases) make possible the in vitro replication of DNA and RNA based on a nucleotide template obtained from a natural gene or a synthesized segment of a gene.

The PCR is proving to be one of the most powerful new tools of the molecular biologist and deserves especial explanation in this forum.⁸⁷⁻⁸⁹ PCR makes use of the fundamental complementarity between two DNA strands and the mechanism of DNA synthesis in which DNA polymerase can copy one new daughter strand from each template strand during each round of synthesis (Fig. 10). To begin the reaction, the sample is heat denatured to separate the two strands of

DNA. The user then cools the sample and supplies DNA *primers*, usually obtained from automatic oligonucleotide synthesis, which can hybridize specifically to each of the 3' ends of the denatured target sequence. The primers are chosen carefully to avoid artifact-generating cross-hybridization with other sequences in the sample. In the presence of all needed nucleotide triphosphates, a heat-stable DNA polymerase generates a new DNA chain from the 3' end of each hybridized primer to complete a double-stranded DNA product. Thus, from one double-stranded starting molecule, two daughter molecules are generated. The two strands are denatured again, allowed to rehybridize with primers (which were provided in vast molar excess), and the process is repeated. With each round, the number of specific templates doubles. Multiple cycles of automated denaturation and polymerization yield a "chain reaction" of exponential increase; after 30 cycles (several hours on an automated thermocycling machine) there are 10^9 product molecules! Although PCR can easily amplify artifacts caused by trace contaminants or cryptic homologies between primers and unwanted DNA sequences, its simplicity of design and rapidity of execution has made this technique one of the premier tools of genetic engineering⁹⁰; because of its extreme sensitivity, PCR will be particularly useful in prenatal diagnoses,⁹¹⁻⁹³ biopsy, forensic medicine, and other applications in which tissue samples may be limiting.⁹⁴⁻⁹⁸

Together, these techniques enable the construction of entirely new genes ("designer genes"). This is a complex process; it requires not only isolation and reproduction of the coding portion of a gene but also assembly, usually within a bacterial or viral gene, in the company of the enhancers and/or promoters that will assure appropriate expression in a specific tissue or microorganism (Fig. 11). The novel gene constructs can then be transferred to other cells, a process termed transfection, in which their function may be expressed *in vitro* or they may be introduced into the genome of an intact organism. Many strategies for gene therapy make use of the former approach. Alternatively, the designer genes may be introduced into embryos or embryonic stem cells and incorporated into the genetic material of the developing embryo. The engineered gene may then be stably expressed in appropriate tissues throughout the life of the "transgenic animal" and later passed on to subsequent generations. As a consequence, new life forms, unknown in nature, are now regularly produced in the scientific laboratory, resulting in such dramatic results as the expression in transgenic mice of human genes and the

production of their corresponding human protein products. Moreover, the expression of human genes in bacteria presently offers the capability of synthesizing essentially unlimited quantities of potent new therapeutic agents, such as interleukins, interferons, and other polypeptide hormones and their derivatives. Thus, as medicine faces its future, the concepts and technologies of molecular biology offer opportunities and challenges unimaginable as late as 20 years ago and which are almost certain to impact directly on delivery of care to our patients in the very near future.

REFERENCES

1. Landegren U, Kaiser R, Caskey CT, Hood L. DNA diagnostics: molecular techniques and automation. *Science* 1988;242:229-37.
2. Lewin B. *Genes*. 3rd ed. New York: John Wiley & Sons, 1987.
3. Watson JD. *Molecular biology of the gene*. 4th ed. Menlo Park, Calif: Benjamin/Cummings, 1987.
4. Schrödinger E. *What is life?* Cambridge: Cambridge University Press, 1942.
5. Huxley J. *Evolution, the modern synthesis*. London: George Allen and Unwin, 1942.
6. Avery OT, MacLeod CM, McCarty M. Studies on the chemical nature of substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from *Pneumococcus* type III. *J Exp Med* 1944;79:137-58.
7. Hershey AD, Chase M. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J Gen Physiol* 1952;36:39-56.
8. Monod J. *Chance and necessity: an essay on the natural philosophy of molecular biology*. New York: Knopf, 1971.
9. Celander DW, Cech TR. Visualizing the higher order folding of a catalytic RNA molecule. *Science* 1991;251:401-7.
10. Cech TR. Ribozymes and their medical implications. *JAMA* 1988;260:3030-4.
11. Watson JD, Crick FHC. A structure for desoxyribose nucleic acids. *Nature* 1953;171:737-8.
12. Tanford C. *The hydrophobic effect: formation of micelles and biological membranes*. 2nd ed. New York: John Wiley & Sons, 1980.
13. Belfort M. Self-splicing introns in prokaryotes: migrant fossils? *Cell* 1991;64:9-11.
14. Rogers JH. The role of introns in evolution. *FEBS Lett* 1990;268:339-43.
15. Holland SK, Blake CC. Proteins, exons, and molecular evolution. *Biosystems* 1987;20:181-206.
16. Peppel K, Vinci JM, Baglioni C. The AU-rich sequences in the 3'-untranslated region mediate the increased turnover of interferon mRNA induced by glucocorticoids. *J Exp Med* 1991;173:349-55.
- 16a. Jacob F, Monod J. Genetic regulatory mechanisms and the synthesis of proteins. *J Mol Biol* 1961;3:318-56.
17. Nussinov R. Sequence signals in eukaryotic upstream regions. *Crit Rev Biochem Mol Biol* 1990;25:185-224.
18. Eccles S, Sarnar N, Vidal M, Cox A, Grosveld F. Enhancer sequences located 3' of the mouse immunoglobulin lambda locus specify high-level expression of an immunoglobulin

- lambda gene in B cells of transgenic mice. *New Biologist* 1990;2:801-11.
19. Meyer KB, Sharpe MJ, Surani MA, Neuberger MS. The importance of the 3'-enhancer region in immunoglobulin kappa gene expression. *Nuc Acids Res* 1990;18:5609-15.
 20. Pongubala JM, Atchison ML. Functional characterization of the developmentally controlled immunoglobulin kappa 3' enhancer: regulation by Id, a repressor of helix-loop-helix transcription factors. *Mol Cell Biol* 1991;11:1040-7.
 21. Latchman DS. Eukaryotic transcription factors. *Biochem J* 1990;270:281-9.
 22. Smale ST, Baltimore D. The "initiator" as a transcription control element. *Cell* 1989;57:103-13.
 23. Kern MJ, Woodward JG. The same CCAAT box-binding factor binds to the promotor of two coordinately regulated major histocompatibility complex class II genes. *Mol Cell Biol* 1991;11:578-81.
 24. Kanno M, Fromental C, Staub A, Ruffenach F, Davidson I, Chambon P. The SV40 TC-II (κ B) and the related H-2Kb enhancers exhibit different cell type specific and inducible proto-enhancer activities, but the SV40 core sequence and the AP-2 binding site have no enhancer properties. *EMBO J* 1989;8:4205-14.
 25. Israël A, Kimura A, Kieran M, et al. A common positive transacting factor binds to enhancer sequences in the promoters of mouse H-2 and β_2 -microglobulin genes. *Proc Natl Acad Sci USA* 1987;84:2653-7.
 26. Yamamoto K. Steroid receptor regulated transcription of specific genes and gene networks. *Annu Rev Genet* 1985;19:209-52.
 27. Carson-Jurica MA, Schrader WT, O'Malley BW. Steroid receptor family: structure and functions. *Endocr Rev* 1990;11:201-20.
 28. Srb AM, Owen RD, Edgar RS. General genetics. 2nd ed. San Francisco: WH Freeman, 1965.
 29. Atchison ML, Perry RP. The role of the kappa enhancer and its binding factor NF-kappa B in the developmental regulation of kappa gene transcription. *Cell* 1987;48:121-8.
 30. Sambucetti LC, Cherrington JM, Wilkinson GW, Mocarski ES. NF-kappa B activation of the cytomegalovirus enhancer is mediated by a viral transactivator and by T cell stimulation. *EMBO J* 1989;8:4251-8.
 31. Griffin GE, Leung K, Folks TM, Kunkel S, Nabel GJ. Activation of HIV gene expression during monocyte differentiation by induction of NF-kappa B. *Nature* 1989;339:70-3.
 32. Nabel G, Baltimore D. An inducible transcription factor activates expression of human immunodeficiency virus in T cells. *Nature* 1987;326:711-3.
 33. Baldwin AS Jr, LeClair KP, Singh H, Sharp PA. A large protein containing zinc finger domains binds to related sequence elements in the enhancers of the class I major histocompatibility complex and kappa immunoglobulin genes. *Mol Cell Biol* 1990;10:1406-14.
 34. Bours V, Villalobos J, Burd PR, Kelly K, Siebenlist U. Cloning of a mitogen-inducible gene encoding a κ B DNA-binding protein with homology to the rel oncogene and to cell-cycle motifs. *Nature* 1990;348:76-80.
 35. Fan C-M, Maniatis T. A DNA-binding protein containing two widely separated zinc finger motifs that recognize the same DNA sequence. *Genes Dev* 1990;4:29-42.
 36. Molitor JA, Walker WH, Doerre S, Ballard DW, Greene WC. NF-kappa B: a family of inducible and differentially expressed enhancer-binding proteins in human T cells. *Proc Natl Acad Sci USA* 1990;87:10028-32.
 37. Gilmore TD. NF-kappa B, KBF1, dorsal, and related matters. *Cell* 1990;62:841-3.
 38. Baeuerle PA, Baltimore D. I κ B: a specific inhibitor of the NF-kappa B transcription factor. *Science* 1988;242:540-6.
 39. Urban MB, Baeuerle PA. The 65-kd subunit of NF-kappa B is a receptor for I kappa B and a modulator of DNA-binding specificity. *Genes Dev* 1990;4:1975-84.
 40. Kimura A, Israël A, Le Bail O, Kourilsky P. Detailed analysis of the mouse H-2Kb promoter: enhancer-like sequences and their role in the regulation of class I gene expression. *Cell* 1986;44:261-72.
 41. Kieran M, Blank V, Logeat F, et al. The DNA binding subunit of NF- κ B is identical to factor KBF1 and homologous to the rel oncogene product. *Cell* 1990;62:1007-18.
 42. Israël A, Kimura A, Fournier A, Fellous M, Kourilsky P. Interferon response sequence potentiates activity of an enhancer in the promotor region of a mouse H-2 gene. *Nature* 1986;322:743-6.
 43. Fan C-M, Maniatis T. Two different virus-inducible elements are required for human beta-interferon gene regulation. *EMBO J* 1989;8:101-10.
 44. Cohen B, Peretz D, Vaiman D, Benech P, Chebath J. Enhancer-like interferon responsive sequences of the human and murine (2'-5') oligoadenylate synthetase gene products. *EMBO J* 1988;7:1411-9.
 45. Reich N, Evans B, Levy D, Fahey D, Knight E Jr, Darnell JE Jr. Interferon-induced transcription of a gene encoding a 15-kDa protein depends on an upstream enhancer element. *Proc Natl Acad Sci USA* 1987;84:6394-8.
 46. Harada H, Fujita T, Miyamoto M, et al. Structurally similar but functionally distinct factors, IRF-1 and IRF-2, bind to the same regulatory elements of IFN and IFN-inducible genes. *Cell* 1989;58:729-39.
 47. Driggers PH, Ennist DL, Gleason SL, et al. An interferon γ -regulated protein that binds the interferon-inducible enhancer element of major histocompatibility complex class I genes. *Proc Natl Acad Sci USA* 1990;87:3743-7.
 48. Pierce JW, Gifford AM, Baltimore D. Silencing of the expression of the immunoglobulin kappa gene in non-B cells. *Mol Cell Biol* 1991;11:1431-7.
 49. Winoto A, Baltimore D. $\alpha\beta$ lineage-specific expression of the α T cell receptor gene by nearby silencers. *Cell* 1989;59:649-55.
 50. Lewis S, Gellert M. The mechanism of antigen receptor gene assembly. *Cell* 1989;59:585-8.
 51. Weisberg R, Landy A. Site-specific recombination in phage lambda. In: Hendrix RW, Roberts JW, Stahl FW, Weisberg RA, eds. *Lambda II*. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory, 1983:211-50.
 52. Grandgenett DP, Mumm SR. Unraveling retrovirus integration. *Cell* 1990;60:3-4.
 53. Herskowitz I, Oshima Y. Control of cell type in *Saccharomyces cerevisiae*: mating type and mating-type interconversion. In: Strathern, Jones, Broach, eds. *The molecular biology of the yeast Saccharomyces*. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory, 1981:181-209.
 54. Borst P, Greaves DR. Programmed gene rearrangements altering gene expression. *Science* 1987;235:658-67.
 55. Cebra JJ, Komisar JL, Schweitzer PA. CH isotype "switching" during normal B-lymphocyte development. *Ann Rev Immunol* 1984;2:493-548.
 56. Callard RE, Turner MW. Cytokines and Ig switching: evolutionary divergence between mice and humans. *Immunol Today* 1990;11:200-3.

57. Gascan H, Gauchat JF, Roncarolo MG, Yssel H, Spits H, de Vries JE. Human B cell clones can be induced to proliferate and to switch to IgE and IgG4 synthesis by interleukin-4 and a signal provided by activated CD4⁺ T cell clones. *J Exp Med* 1991;173:747-50.
58. Kim PH, Kagnoff MF. Transforming growth factor beta 1 increases IgA isotype switching at the clonal level. *J Immunol* 1990;145:3773-8.
59. Sawadogo M, Sentenac A. RNA polymerase B (II) and general transcription factors. *Ann Rev Biochem* 1990;59:711-54.
60. Proudfoot N. Poly(A) signals. *Cell* 1991;64:671-4.
61. Bernstein P, Peltz SW, Ross J. The poly(A)-poly(A)-binding protein complex is a major determinant of mRNA stability in vitro. *Mol Cell Biol* 1989;9:659-70.
62. Sharp PA. RNA splicing and genes. *JAMA* 1988;260:3035-41.
63. Peterson ML, Perry RP. The regulated production of μ_m and μ_c mRNA is dependent on the relative efficiencies of μ , poly(A) site usage and the μ 4-to-M1 splice. *Mol Cell Biol* 1989;9:726-38.
64. Lee BJ, Rajagopalan M, Kim YS, You KH, Jacobson KB. Selenocysteine tRNA^{ser}:sec gene is ubiquitous within the animal kingdom. *Mol Cell Biol* 1990;10:1940-9.
65. Ninio J. The revised genetic code. *Orig Life Evol Biosph* 1990;20:167-71.
66. Kozak M. At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *J Mol Biol* 1987;196:947-50.
67. Kozak M. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc Natl Acad Sci USA* 1990;87:8301-5.
68. Hou YM, Shiba K, Mottes C, Schimmel P. Sequence determination and modeling of structural motifs for the smallest monomeric aminoacyl-tRNA synthetase. *Proc Natl Acad Sci USA* 1991;88:976-80.
69. Kozak M. Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. *Mol Cell Biol* 1989;9:5073-80.
70. Murgola EJ. Suppression and the code: beyond codons and anticodons. *Experientia* 1990;46:1134-41.
71. Ausubel FM, Brent R, Kingston RE, et al., eds. *Current protocols in molecular biology*. New York: John Wiley & Sons, 1989.
72. Sambrook J, Fritsch EF, Maniatis T, et al., eds. *Molecular cloning: a laboratory manual*. 2nd ed. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory, 1989.
73. Ishizaka K. IgE-binding factors and regulation of the IgE antibody response. *Ann Rev Immunol* 1988;6:513-34.
74. Blank U, Ra C, Miller L, White K, Mutzger H, Kinet JP. Complete structure and expression in transfected cells of high-affinity IgE receptor. *Nature* 1989;337:187-9.
75. Baldo BA, Donovan GR. The structural basis of allergenicity: recombinant DNA-based strategies for the study of allergens. *Allergy* 1989;44:81-97.
76. Heymann PW, Chapman MD. Recent progress in mite-allergen immunochemistry. *Clin Rev Allergy* 1990;8:51-68.
77. Chua KY, Doyle CR, Simpson RJ, Turner KJ, Stewart GA, Thomas WR. Isolation of cDNA coding for the major mite allergen *Der p* II by IgE plaque immunoassay. *Int Arch Allergy Appl Immunol* 1990;91:118-23.
78. Fang KS, Vitale M, Fehlner P, King TP. cDNA cloning and primary structure of a white-face hornet venom allergen. antigen 5. *Proc Natl Acad Sci USA* 1988;85:895-9.
79. Griffith IJ, Smith PM, Pollock J, et al. Cloning and sequencing of *Lol 1 p* I, the major allergenic protein of ryegrass pollen. *FEBS Lett* 1991;279:210-5.
80. Mohapatra SS, Hill R, Astwood J, et al. Isolation and characterization of a cDNA clone encoding an IgE-binding protein from Kentucky bluegrass (*Poa pratensis*) pollen. *Int Arch Allergy Appl Immunol* 1990;91:362-8.
81. Rafnar T, Griffith IJ, Kuo MC, Bond JF, Rogers BL, Klapper DG. Cloning of *Amb a* I (antigen E), the major allergen family of short ragweed pollen. *J Biol Chem* 1991;266:1229-36.
82. Singh MB, Hough T, Theerakulpisut P, et al. Isolation of cDNA encoding a newly identified major allergenic protein of ryegrass pollen: intracellular targeting to the amyloplast. *Proc Natl Acad Sci USA* 1991;88:1384-8.
83. Singh MB, Smith PM, Knox RB. Molecular biology of ryegrass pollen allergens. *Mono Allergy* 1990;28:101-20.
84. Tovey ER, Johnson MC, Roche AL, Cobon GS, Baldo BA. Cloning and sequencing of a cDNA expressing a recombinant house dust mite protein that binds human IgE and corresponds to an important low molecular weight allergen. *J Exp Med* 1989;170:1457-62.
85. Gurka G, Ohman J Jr, Rosenwasser LJ. Allergen-specific human T cell clones: derivation, specificity, and activation requirements. *J ALLERGY CLIN IMMUNOL* 1989;83:945-54.
86. McKusick VA. The new genetics and clinical medicine: a summing up. *Hospital Prac* 1988;23:79-93.
87. Remick DG, Kunkel SL, Holbrook EA, Hanson CA. Theory and applications of the polymerase chain reaction. *Am J Clin Pathol* 1990;93:S49-54.
88. Rodu B. The polymerase chain reaction: the revolution within. *Am J Med Sci* 1990;299:210-6.
89. Erlich HA. Polymerase chain reaction. *J Clin Immunol* 1989;9:437-47.
90. Gibbs RA. DNA amplification by the polymerase chain reaction. *Anal Chem* 1990;62:1202-14.
91. Eisenstein BI. The polymerase chain reaction: a new method of using molecular genetics for medical diagnosis. *N Engl J Med* 1990;322:178-83.
92. Husson RN, Comeau AM, Hoff R. Diagnosis of human immunodeficiency virus infection in infants and children. *Pediatrics* 1990;86:1-10.
93. Goldberg JD. Basic principles of recombinant DNA use for prenatal diagnosis. *Sem Perinatol* 1990;14:439-45.
94. Jackson JB. The polymerase chain reaction in transfusion medicine. *Transfusion* 1990;30:51-7.
95. Lynch JR, Brown JM. The polymerase chain reaction: current and future clinical applications. *J Med Genet* 1990;27:2-7.
96. Fox RJ, Dotan I, Compton T, Fei HM, Hamer M, Saito I. Use of DNA amplification methods for clinical diagnosis in autoimmune diseases. *J Clin Lab Anal* 1989;3:378-87.
97. Li HH, Cui XF, Arnheim N. Analysis of DNA sequences in individual gametes: application to human genetic mapping. *Prog Clin Biol Res* 1990;340C:207-11.
98. Wright PA, Wynford-Thomas D. The polymerase chain reaction: miracle or mirage? A critical review of its uses and limitations in diagnosis and research. *J Pathol* 1990;162:99-117.
99. Crick FHC. Central dogma of molecular biology. *J Mol Biol* 1968;38:367-80.