1

Structure and Function of the Gene

David Botstein

INTRODUCTION

A revolution in biology occurred during the second half of the twentieth century. The insights of modern genetics and biochemistry produced a new science, *molecular biology*, that provided unprecedented understanding of the molecular mechanisms at work in living organisms. This in turn has enabled scientists and physicians to intervene at the molecular level—an ability that has only begun to be appreciated and applied to the practice of medicine.

The revolution in biology is comparable in scale and impact to the revolution in physics that occurred during the first half of the twentieth century. The insights of relativity and quantum mechanics led to unprecedented understanding of the mechanisms underlying chemical, electronic, and nuclear phenomena. This understanding in turn led to a great variety of useful innovations, ranging from radioisotopes and nuclear power to the transistor and the laser.

GENES ARE INFORMATION ENCODED IN DNA

The central concept that drove the revolution in biology was the realization that the information stored in the genes, and the way in which this information is reproduced and expressed, is at the core of scientific understanding of all living things. The seminal paper by Avery, McLeod, and McCarty in 1944 began a trail of discovery that can be summarized by stating that essentially all the information that is passed from generation to generation, and that must be expressed in every cell of every organism, is encoded in DNA molecules. The discovery of the double-helical structure of DNA by Watson and Crick in 1953 and the elucidation of the mechanism of protein synthesis during the early 1960s resulted in the formulation of a *central dogma* that has proved to have extraordinary explanatory power. Stated in modern language, the central dogma consists of the following elements.

1. Information is passed from generation to generation through the replication of the double-stranded DNA polymer. The information is encoded in DNA as the sequence of four alternative nucleotide subunits, namely the bases adenine (A), guanine (G), cytosine (C), and thymine (T). In the DNA structure, an A on one strand is always paired with a T on the other, and a G on one strand is always paired with a C on the other. Thus, the double-stranded structure of DNA provides redundant coding, so that the genetic information is present twice—once on each complementary strand. As noted in the famous 1953 paper by Watson and Crick, the complementary strand feature makes it easy to see how replication can occur without loss of information, because the sequence in each strand suffices to determine, using the basepairing rule, the sequence of the complementary copy. During replication, the strands separate, and each serves as the template for assembling the subunits of its complement in the right order, producing two DNA copies with the same sequence. Thus, when one cell becomes two, each is endowed with the same information (Fig. 1-1).
2. Information encoded in the DNA is expressed within cells primarily by the specification of the sequence of amino acids in proteins. The sequence of bases in the DNA encodes the sequences of amino acids in the many different kinds of proteins. Each *gene* was found to encode a polypeptide chain such that the sequence of nucleotide bases determined a colinear sequence of amino acids, using a simple code. When this code was broken, in the early 1960s, it was learned that each amino acid is specified by a triplet *codon* of adjacent nucleotides. Significantly, it was discovered that the code relating nucleotide sequence to amino acid sequence is nearly universal among free-living organisms (Fig. 1-2). This eventually made possible the use of a convenient organism (e.g., a bacterium) to express and produce the protein specified by a gene of another (e.g., the human). During the late 1970s, this became the technology on which the biotechnology industry was founded.
3. The genetic information in DNA is not decoded directly into proteins when expressed in a cell. Instead, the DNA segment(s) encoding a polypeptide chain are first copied selectively into a single-stranded polymer called *messenger RNA* (mRNA). The mRNA copies are then used to direct the synthesis of proteins. The lifetime of these mRNA molecules in the cell is generally short, so that when their production ceases, production of protein ceases shortly thereafter.

   The decoding principle during protein synthesis directed by mRNA molecules is the use of an *adaptor* molecule called *transfer RNA* (tRNA) that has triplet sequences that recognize specific triplet codons at one end and the cognate amino acid residue attached to the other. The site of protein synthesis is large assemblies of RNA and protein molecules called *ribosomes* located in the cytoplasm; the DNA, of course, remains in the nucleus (Fig. 1-3). Thus, the mRNA molecules copied in the nucleus travel to the cytoplasm, where they are decoded.

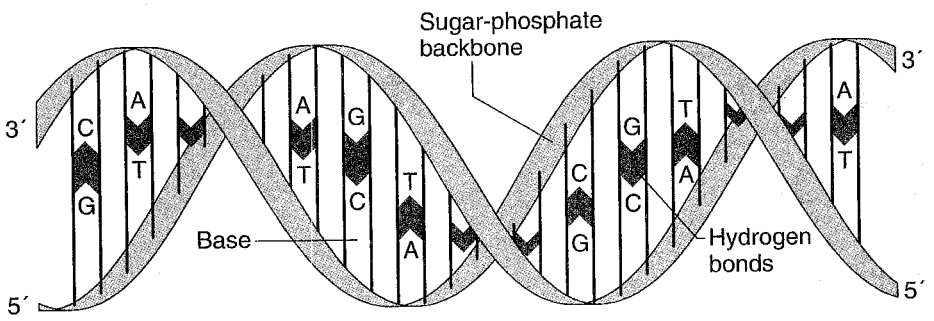


Fig. 1-1. Double-stranded DNA. Genomic DNA consists of two antiparallel strands carrying complementary sequences of nucleotide bases. The strands are held together by hydrogen bonding between the bases. Stacking of the bases results in a double-helical structure.

PROTEINS DO THE WORK OF THE CELL

Modern biology began at the end of the nineteenth century with the insight that all living things are made of cells. Some organisms consist of just a single self-reproducing cell. Others contain many millions of cells—cells that can be classified into thousands of different types based on their properties. A human body contains many different kinds of cells, yet a single individual begins life as a single cell, the zygote.

1st position (5′ end)	2nd position U	2nd position C	2nd position A	2nd position G	3rd position (3′ end)
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	STOP	STOP	A
	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Fig. 1-2. The genetic code. A unit of three nucleotides (codon) specifies each amino acid during translation of mRNA into protein. U stands for uracil, which takes the place of T in RNA. Note that most amino acids are encoded by more than one codon, but a single codon can specify only one amino acid.

What makes cells differ? All cells contain nucleic acids (DNA and RNA), proteins, polysaccharides, lipids, and other bodies. Virtually all cells of the same organism have DNA that is essentially identical. However, if one looks at proteins among different kinds of cells, one finds many obvious differences. Each cell type turns out to have a complement of proteins characteristic of that cell; virtually all other differences can be traced to differences in the protein complement. Possession of a unique characteristic set of proteins is also the underlying explanation of the unique functional differences among cells.

There are all types of proteins, proteins that have very diverse roles in the cell.

Enzymes

Enzymes are catalysts that allow chemical reactions to proceed that otherwise would not occur at a useful rate. Enzymes are used in cells to derive useful energy from the metabolism of nutrients;

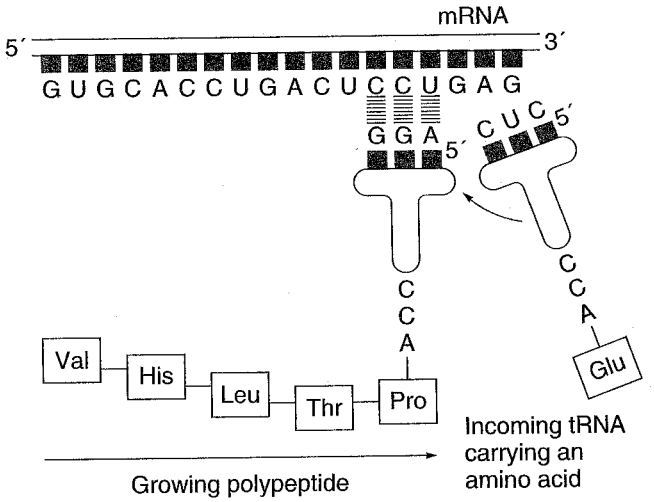


Fig. 1-3. Protein synthesis. An mRNA is read three bases at a time and paired with its complementary "anticodon" sequence in a tRNA. The tRNA carries the amino acid specified by the codon being read, which is linked to the growing polypeptide.

to synthesize the hundreds of compounds required as building blocks to make more cells; to carry out the assembly of polymers such as DNA, RNA, and protein; to modify chemically other molecules, including other proteins in order to regulate their function; and to form and destroy structures, such as blood clots, essential at one time or another for life.

Structural Proteins

Structural proteins are used to make a great variety of biologic structures. Examples include the capsids of viruses; the protein constituents of hair, nails, and bones; the microtubule proteins that form the mitotic spindle; and the fibrin that is the major constituent of blood clots that prevent an injured animal from bleeding to death.

Regulatory Proteins

Regulatory proteins bind to other macromolecular elements to regulate their function. Examples include hormone receptors; repressors and activators that regulate gene expression by binding to DNA to prevent or to stimulate transcription of genes; and inhibitors of proteases required to form blood clots until the clots are called for by injury.

Receptors

Receptors are proteins that span the membranes of cells that respond to molecules, large or small, called *ligands* that, when bound cause the receptor to carry out some intracellular reaction that sends a regulatory signal. This process, called *signal transduction* is at the heart of how the different cell types in multicellular organisms maintain their identity and communicate with each other. For instance, each cell type has receptors that, when bound by ligands called *growth factors*, transduce signals that stimulate cell growth. Cancer is frequently caused by mutations elements of growth factor signal transduction that permit, or even stimulate, growth under inappropriate circumstances.

Some proteins are made only by particular cell types, as described above. Others, in contrast, are ubiquitous because they carry out functions essential to all types of cells. All cells found in the human body, for instance, are the result of cell divisions that began with replication of the DNA by a group of enzymes, including especially one called *DNA polymerase*. As its name suggests, DNA polymerase is the protein that actually catalyzes polymerization of the individual nucleotide subunits into a DNA strand. DNA polymerases are associated with many other proteins whose role is to detect and repair replication mistakes. These *DNA repair* enzymes cooperate with the polymerases to keep the rate of mutation low. Some types of cancer are caused by failures in the function of DNA repair enzymes.

Biochemistry has come a very long way since the insight that proteins do the work of the cell. For virtually every process a living thing has been seen to do, one or more proteins have been detected (often purified and studied in detail) that account for the function. Some of these functions are very complex, yet proteins are the key players throughout.

STEPS IN THE INFORMATION PATHWAY

The central dogma is today frequently summarized as a pathway depicting the flow of information during gene expression:

$$\text{DNA} \rightarrow \text{mRNA} \rightarrow \text{protein}$$

The copying of a gene's information from DNA into mRNA is referred to as *transcription*; and the use of the information in the mRNA to specify the order of amino acids during protein synthesis is referred to as *translation*. Indeed, this nomenclature is appropriate, since the *language* of DNA (sequence of nucleotides, a four-letter alphabet) is the same as that of mRNA. It is only during protein synthesis that the information is translated into the language of proteins (sequence of amino acids, a 20-letter alphabet). The flow of information is thus

$$\text{DNA} \xrightarrow{\text{transcription}} \text{mRNA} \xrightarrow{\text{translation}} \text{protein}$$

This formulation led very quickly to the realization that a single gene's DNA must somehow not only encode the sequence of amino acids in a protein but also provide some indication of where to start and stop the process of transcription (i.e., begin and end the mRNA) and where to start and stop translation of the mRNA (i.e., begin and end the polypeptide chain). The translation start and stop signals are well understood. Translation generally begins with a triplet codon (ATG) that encodes the amino acid methionine—sometimes it is the first such codon in the mRNA. The end of translation is encoded as one of the three *stop codons* (TAG, TAA, or TGA). These do not specify an amino acid; the first stop codon encountered by the ribosomes terminates translation.

The signals that indicate where transcription of DNA into mRNA should begin are more complicated and less well understood than the signals in the mRNA that indicate where translation should begin. Nevertheless, much is known about the transcription start signals, which are referred to as *promoters*. There are also signals for ending transcription (*terminators*) that are likewise more complicated and less well characterized than the stop codons that cause translation to cease.

Most of the mRNAs in eukaryotic organisms and some bacteria are not simply the *primary transcript* (i.e., a copy of the DNA sequence from promoter to terminator). In most eukaryotes, the mRNA transcribed from all but a handful of genes is *polyadenylated*, meaning that a string of A nucleotide residues is attached to the end. In addition, the initial residue of most eukaryotic mRNAs is chemically modified, producing a structure called a *cap*.

More significantly, however, the primary transcript of most genes in most eukaryotic organisms (including humans) is much longer than the final mRNA. The transcript is then *spliced* so that the coding segments (*exons*) are joined together, leaving out the intervening noncoding sequences (*introns*). The exons are joined in the order in which they appear in the DNA, although in some cases some exons are selectively skipped under some circumstances or cell types and not in others. The signals that direct the splicing points are, like the promoter and terminator, encoded in the sequence.

These considerations result in the following pathway, which is also depicted in Figure 1-4:

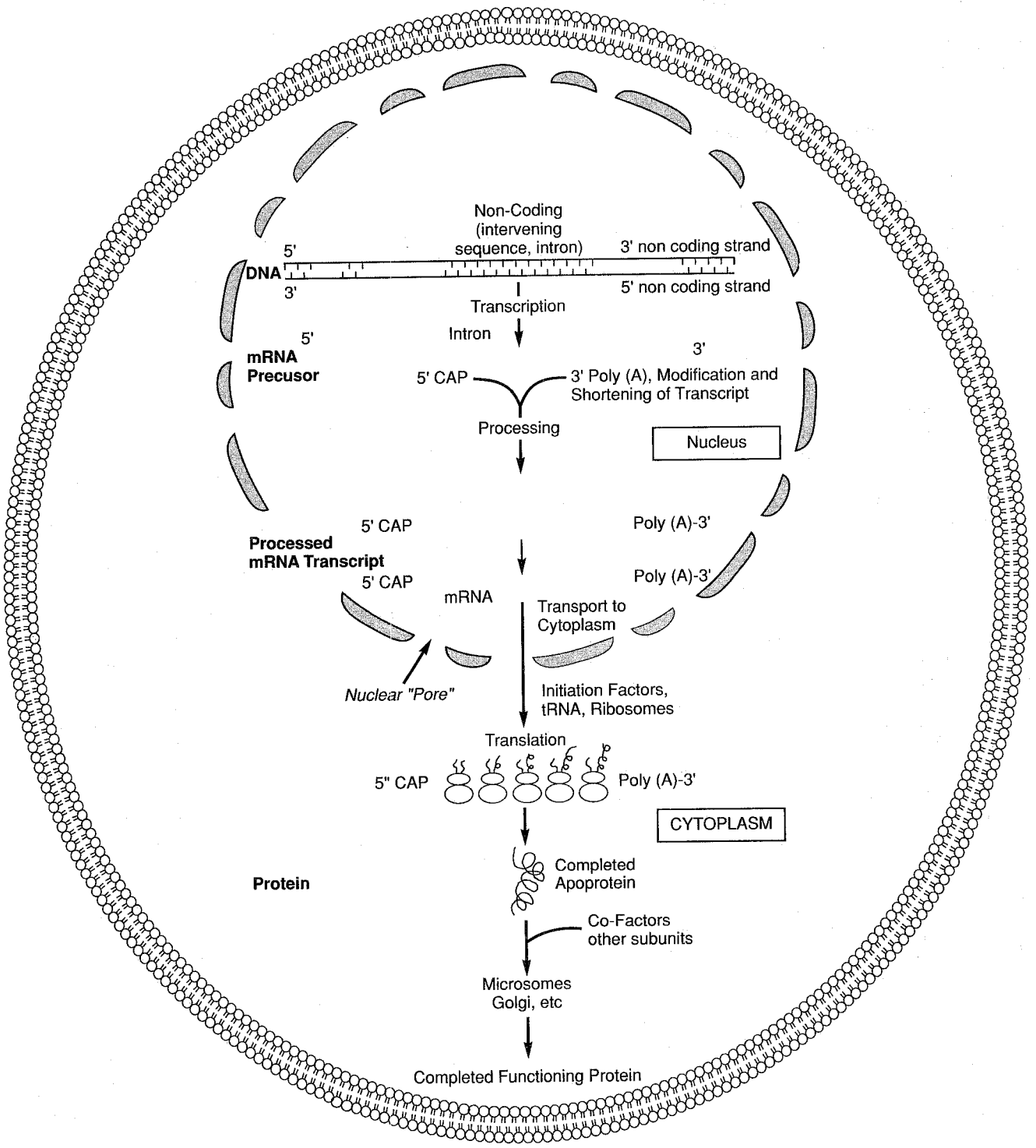
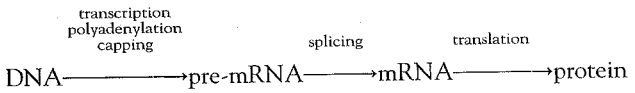


Fig. 1-4. Synthesis of mRNA and protein pathway of gene expression. The diagram of the DNA gene shows the alternating array of exons (red) and introns (pink) typical of most eukaryotic genes. (From Benz,[1] with permission.)



THE STRUCTURAL VIEW OF THE DEFINITION OF THE GENE

From the point of view of structure, a gene is the information encoded in DNA required to produce a protein. Since proteins sometimes consist of one polypeptide subunit and sometimes more than one, a more rigorous statement is to say that each gene encodes a single *polypeptide chain* (i.e., a continuous string of amino acids). A protein consisting of more than one kind of polypeptide subunit is encoded by more than one gene. The information in a gene thus includes not only the actual coding sequence that specifies the order of amino acids, but also the signals that start and end transcription, the signals that indicate where splicing is to occur, and the signals that indicate where translation into an amino acid sequence begins and ends. Any change in any of this information encoded in the DNA will have consequences for the production of the encoded polypeptide chain.

One of the most important generalizations to emerge from molecular biology is that the promoter region (i.e., sequences near the start of transcription in a gene) is usually the point at which regulation of gene expression is exerted. Thus the most common reason for the presence of a protein in one kind of cell and not in another is that the one kind of cell transcribes the gene encoding the protein and the other does not. Likewise, the most common reason for the presence of a protein in one environment, and not in another, is a regulatory mechanism that permits transcription in one circumstance and prevents it in the other. Regulation at other steps in the information pathway from DNA to protein have been observed, including important examples of regulation at the level of splicing, translation, stability of mRNA, and stability of the protein; however, regulation at the level of transcription remains the most common and important form of regulation of gene expression.

READING AND UNDERSTANDING DNA SEQUENCES

The complementary double-stranded DNA structure has an intrinsic symmetry. This means that a DNA sequence can be read in several ways. This is equally true for the cellular machinery, particularly the transcription apparatus, that expresses the information encoded in it. Molecular biologists have evolved nomenclature conventions to enable people and their computer programs to communicate about DNA molecules unambiguously. These conventions have become embedded in all writing about genes, and an understanding of these rules is essential. These rules are best illustrated by an example, such as a DNA molecule with the following sequence:

5′ GTAAATATGCGGGCAGTCCCGACCTGTAATCAGG 3′
3′ CATTTATACGCCCGTCAGGGCTGGTCATTAGACC 5′

Note first that the two complementary strands are always written so that the complementary base pairs are one above the other. The 5′ and 3′ indicate that the two strands run in opposite chemical direction. By convention, DNA molecules are always shown with the 5′ on the upper strand to the left. When only one strand is shown, the 5′ is always to the left. The sequence of DNA is written so that when the DNA is transcribed into mRNA, the *lower* strand is the one used as template for copying, meaning that the sequence that ends up in the mRNA is identical to the *upper* DNA strand; that is

5′ GUAAAUAUGCGGGCAGUCCCGACCTGUAAUCAGG 3′

(T's are U's in RNA, representing a trivial chemical difference in the base.) Again note that the mRNA sequence is always written with the 5′ to the left. This is convenient, because translation in cells always proceeds from the 5′ to the 3′ end of the mRNA.

In the example, the first AUG (methionine) codon would most probably initiate translation; the next triplet is CGG (arginine), then GCA (serine), followed by GUC (leucine), until the TAA (STOP, or ***) codon is reached.

To make this easier to comprehend, DNA sequences are often presented with spaces as follows:

5′ GTAAAT ATG CGG GCA GTC CCG ACC TGG TAA TCAGG 3′

3′ CATTTA TAC GCC CGT CAG GGC TGG TCC ATT AGACC 5′
met arg ser leu ser thr trp ***

The choice of the first methionine codon determines how the subsequent codons are read, the *reading frame*. If one does not know where translation begins, there are, for any DNA sequence, three potential reading frames for each strand. By convention, when the reading frame is known, the DNA, RNA, and protein sequences are aligned as above. In the sequences of most eukaryotic genes, the reading frame is maintained (i.e., there are long stretches of triplets with no stop codons) only in the exons. Introns, which are removed during splicing, generally do not have an *open reading frame*. The recognition of open reading frames in sequences is one of the most important steps in the interpretation of a DNA sequence.

The preceding examples show that the roles of the two complementary DNA strands differ: one contains the mRNA sequence that will be translated, while the complement is the actual template for transcription. Over the years, the DNA strand that has the same sequence as the mRNA is called the *sense strand*, while the complementary strand (which contains the same information, but the complementary sequence) is called the *antisense strand*.

MUTATIONS ARE ALTERATIONS IN DNA THAT HAVE HERITABLE FUNCTIONAL CONSEQUENCES.

Any change in the sequence of an organism's DNA is heritable. If the change occurs in a cell lineage that leads to the formation of *gametes* (i.e., egg or sperm), the change will be passed on to

offspring; such a mutation is sometimes said to affect the *germline*. If the change occurs in a *somatic* cell lineage (i.e., one not leading to the formation of gametes), the change cannot be passed to offspring, even though the change will be passed on to all the descendants of the mutant cell.

Mutations are classified in many ways. One set of classes is based on the structural change itself: major classes include *point mutations* (mutations that change a single base in the sequence), *deletions* or *insertions* (mutations that remove or add one or more bases), and *inversions* (mutations that invert a segment of DNA sequence). Another set of classifications is based on the consequences of the mutation for the encoded protein: major classes include *missense* (mutations that substitute one amino acid for another), *nonsense* (mutations that substitute a stop codon for codon specifying an amino acid), *frameshift* (mutations that add or delete a number of bases in a coding sequence that changes the reading frame, resulting in a mistranslation of all the amino acid sequence beyond the mutation), and *splicing* (mutations that alter a splice junction, usually causing major changes in the mRNA and thus the encoded protein). Obviously these different kinds of mutations have different likelihoods of eliminating a protein's function: missense mutations may alter the encoded protein severely, very subtly, or not at all. Large deletions, nonsense, and frameshift mutations, if they occur anywhere but the very end of a coding sequence, will most likely abolish the protein entirely.

GENOTYPE AND PHENOTYPE

From the point of view of biology, the most important ways of classifying mutations are not based on the structure of the mutation but on the consequences to the cell or organism. Geneticists use the word *phenotype* to mean any and all consequences of a mutation. The phenotype is anything one can see, measure, or detect in the organism that is caused by the gene.

By *genotype* geneticists mean the state of the gene(s) in an organism. For a single gene with two alternative sequences (*alleles*), there are two possibilities in a diploid organism like the human: the individual can have copies of the same allele (i.e., *homozygous*) or the individual can have one copy of one allele and one of the other (i.e., *heterozygous*).

Mutations Can Cause Their Characteristic Phenotypes Under Different Circumstances

If the mutation must be homozygous (i.e., present in both copies of a gene carried by an individual), the effect of the mutation is said to be *recessive*. A common example of a gene defect that causes a disease in a recessive manner is cystic fibrosis. Heterozygotes (i.e., individuals with one defective and one normal gene) are phenotypically completely normal, and the presence of the defective gene can be detected only by direct analysis of the DNA sequence. Individuals with two defective copies (i.e., homozygotes for the defective allele) invariably have the disease. Recessiveness of a defect generally means that the alteration in DNA sequence results is a loss of function.

If the presence of the mutation in one of the gene copies) suffices to produce the phenotype, the effect of the mutation is said to be *dominant*. An example of a gene defect that causes a disease in a dominant fashion is Huntington disease. Virtually all individuals with even a single defective copy of the gene (i.e., heterozygotes) are eventually affected. Dominance of defect generally is taken to mean that the alteration in DNA sequence results in a gain of (an inappropriate) function.

Dominance and recessiveness are related. In the case of a gene defect with recessive effect, the normal form of the gene (referred to as the normal *allele*) is said to be dominant over the disease-causing allele. Similarly, in the case of a gene defect with dominant effect, the normal allele is said to be recessive to the disease-causing one.

Sometimes the phenotype of an individual reflects both alleles present. Two alternative alleles of the same gene produce a *codominant* effect when both are qualitatively reflected in the genotype. An example is the AB blood type, which is observed if, and only if, an individual has one copy of the ABO gene that specifies the A surface antigen and one copy that specifies the B antigen. In some cases, the phenotype of an individual reflects both alleles quantitatively, rather than a qualitatively, in which case they are said to show *incomplete dominance*. A common example of incomplete dominance is the gene for familial hypercholesterolemia (FH). The blood level of cholesterol in heterozygotes carrying one normal and one defective FH allele is significantly elevated but is not nearly as extreme as the level seen in homozygotes for the defective gene.

Dominance and Recessiveness Apply to the Phenotype

Depending on the level of analysis, a mutation can have a great variety of phenotypes. Consider, for instance, the mutation that changes the sixth codon (from the N-terminus) of the human gene encoding the β-subunit of hemoglobin from . . . GAG . . . to . . . GTG. . . . The codon table confirms that this will change the glutamate residue at position 6 of the amino acid sequence to valine.

At the level of the whole organism, this mutation results in a debilitating (in earlier times, fatal) disease, sickle cell anemia. The effect of the altered gene sequence is recessive, since heterozygotes are not anemic, whereas homozygotes invariably are.

Still at the level of the whole organism, this mutation results in resistance to *Falciparum* malaria. Indeed, this is the reason for the high prevalence of the disease gene in the tropics, where malaria is endemic. The effect of the altered gene here is dominant, since heterozygotes are resistant.

At the level of the erythrocyte, the characteristic feature of the presence of this mutation is the tendency of the cells to deform, producing a characteristic sickle shape. Under conditions of low to moderate oxygen tension, the erythrocytes of homozygotes sickle completely and rapidly, whereas the erythrocytes from heterozygotes sickle more slowly. The effect of this mutation is incompletely dominant.

At the level of the hemoglobin protein molecule, measurement of the electrophoretic mobility of the hemoglobin extracted from red cells shows a single band in homozygotes (for either normal or mutant alleles) but two bands in heterozygotes. Thus, the altered gene and the normal are codominant with respect to this phenotype.

At the level of the DNA sequence, the polymerase chain reaction (PCR) can readily detect both sequences in heterozygotes. Here again the mutation shows codominance. Actually, all gene mutations are codominant at the DNA level.

To conclude, then, the dominance characteristics of mutations and alleles depend on the phenotype being observed. The mutation at position 6 of β-globin shows many phenotypes, and each has its characteristic dominance relationship with the normal allele.

FUNCTIONAL VIEW OF THE DEFINITION OF THE GENE

The structural view of the gene given above can be complemented by a more biologically oriented view, based not on the mechanics of gene expression but on the effect of changes in genes. Thus, the functional view of the gene is that each gene is responsible for one or more biologic functions whose nature is generally inferred from the phenotypes of mutations causing changes (generally loss) in function. Molecular methods have made it possible to bring together the two views of the gene so that we can actually attribute all the phenotypes associated with the sickle cell disease mutation to the one C ! T change in one letter (out of 3 billion) of the genomic sequence.

The same blending of functional and structural genetics has made it possible to attribute many (if not all) cancers to mutations in DNA sequences of genes. From study of the phenotypes, we are now beginning to understand how growth is regulated. Careful studies of the failures of that regulation show that each is caused by one, or at most a few, mutation(s).

MUTATIONS IN ONCOGENES ARE THE PROXIMAL CAUSE(S) OF CANCER

During the past decade it has become clear that mutations are causes of cancer. More than 100 *oncogenes* have been identified in which mutations contribute to the cause of one or another type of tumor. The mutations that have been found include virtually all the types of mutations found in other genes. Some mutant oncogenes contribute to the tumor phenotype(s) in a dominant fashion, others in a recessive fashion. At the cellular level, more than one gene must mutate to produce a fully malignant tumor phenotype, although a mutation in a single gene sometimes suffices to start a series of events that leads inexorably to the formation of tumors.

Sometimes the mutations occur somatically and are found only in the cells of the tumor or their immediate ancestor lineages. In other cases, the mutant gene can be inherited. As with all mutations, the relationship between genotype and phenotype can be quite complex, depending on both the level of analysis and the details of the phenotypes being studied.

The classic oncogenes (e.g., *src* and *ras*) were originally discovered as the cancer-causing elements of viruses responsible for cancer in birds and animals. They are often point mutations in genes specifying normal components of signal transduction pathways that cause the signals to be transduced inappropriately (e.g., a growth signal in the absence of the growth factor). As one might expect from this mechanism, these mutant genes exert their effect in a dominant fashion.

The so-called *tumor suppressor genes*, or *recessive oncogenes*, were originally discovered through the study of inherited strong predispositions to cancer, such as retinoblastoma. These genes confer upon cells homozygous for the mutant allele the ability to grow inappropriately. At the cellular level, they thus are recessive. However, at the level of the organism, they are dominantly acting. Individuals who inherit one copy of a defective retinoblastoma gene eventually lose, in at least a few cells in each eye, the other (normal) copy through mutation, chromosome loss, or mitotic recombination. This results in homozygosity and the expression of the tumor phenotype. Thus, the pattern of inheritance is that of a single dominant autosomal gene.

Like the mutation that causes sickle cell anemia, the retinoblastoma mutation is dominant or recessive, depending on the phenotype one is studying and the level of analysis. The mutant retinoblastoma gene is dominant at the organism level, where a single copy means (eventually) cancer, and recessive at the cellular level, where the normal allele has to be lost before the series of events leading to the tumor phenotype is initiated.

Genes in which mutations predispose to cancer by raising the rate of mutation have been discovered, again through the study of inherited strong dispositions to cancer, in this case colon cancer. Here the gene can be inherited and acts as a dominant at the organismal level and a recessive (or partial dominant) at the cell level. The significant point is that this mechanism is indirect, raising the rate at which other more direct-acting oncogenes are mutated.

CONCLUSIONS

Our understanding of basic genetic mechanisms has made possible a detailed description of what genes are and how they work. This in turn has led to the discovery that cancer is caused by mutations in genes, mutations that can be inherited through the germline, or mutations that occur somatically in one or another tissue of the body.

REFERENCES

1. Benz EJ Jr: Anatomy and physiology of the gene. p. 1. In Hoffman R, Benz EJ Jr, Shaltil GJ et al: (eds) Hematology: Basic Principles and Practice. 2nd Ed. Churchill Livingstone, New York, 1995

SUGGESTED READINGS

Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson J (eds): Molecular Biology of the Cell. Garland Publishing, Inc., New York, NY, 1983

Benz EJ Jr (ed): Molecular Genetics Methods. Methods in Hematology Series. Churchill Livingstone, Edinburgh, 1989

Eng C, Ponder BAJ: The role of gene mutations in the genesis of familial cancer. FASEB J 7:910, 1993

High KA, Benz EJ Jr: ABC's of molecular genetics: a haematologist's introduc-

tion. p. 25. In Hoffbrand AV (ed): Recent Advances in Hematology. Churchill Livingstone, Edinburgh, 1985
High KA, Benz EJ Jr: Recombinant DNA technology in antenatal diagnosis. Lab Manage 24:31, 1986
Jeffreys AJ, Wilson V, Thein SL: Hypervariable "minisatellite" regions in human DNA. Nature 314:67, 1985
Nowell PC: Cytogenetic approaches to human cancer genes. FASEB J 8:408, 1994
McKusick VA: The morbid anatomy of the human genome: a review of gene mapping in clinical medicine. Medicine (Baltimore) 66:1, 1987
Mount SM, Steitz JA: RNA splicing and the involvement of small ribonucleoproteins. Med Cell Biol 3:249, 1984
Nevins JR: The pathway of eukaryotic mRNA formation. Annu Rev Biochem 52:441, 1983
Loda M: Polymerase chain reaction-based methods for the detection of mutations in oncogenes and tumor suppressor genes. Hum Pathol 25:564, 1994
Pardee AB: Multiple molecular levels of cell cycle regulation. J Cell Biochem 53:375, 1994
Rabbitts TH: Chromosomal abnormalities in cancer cells: Identification of molecules important for tumor development. Biochem Soc Transactions 21:809, 1993
Radin AL, Benz EJ Jr: Antenatal diagnosis of the hemoglobinopathies. Hematol Pathol 2:199, 1988
Rowley PT, Benz EJ Jr, Nienhuis AW: Molecular genetics for the hematologist. Curr Hematol Oncol 4:1, 1986
Varmus H, Weinberg RA (eds): Genes and the Biology of Cancer. Scientific American Library, New York, NY, 1993
Weatherall DJ: The New Genetics and Clinical Practice. 2nd Ed. Oxford University Press, Oxford, 1985
Weinberg RA: Tumor suppressor genes. Science 254:1138, 1991
Zambetti GP, Levine AJ: A comparison of the biological activities of wild-type and mutant p53. FASEB J 7:855, 1993