2

The Biological Basis for Human Variation

Humans share similar modes of reproduction with most of the other mammals, and inheritance mechanisms are the same—the combination of certain material in the germ cells of male and female parent to produce a fertilized egg. These mechanisms of inheritance are the source of much of the vast diversity seen in the biological world. Though the variety may seem to be extremely random and unlimited, there are, in fact, limits to the extent and degree of variation in each species. *Homo sapiens*, the species with which we are most concerned, contains as much or perhaps more variation than any other mammalian group, but its diversity is also limited by certain processes. For many centuries, natural scientists had sought to comprehend and explain the processes of reproduction and the transmission of traits between generations. The explanations varied from a description of a "blending" of parental bloodlines, favored by animal husbandry, to a theory of "preformism," the idea that the individual, in miniature form, existed in either the ovum or sperm awaiting stimulation by fertilization to begin its development. None of these explanations could account for the ranges of individual similarities or differences among offspring and their parents. A thorough understanding of the mechanisms of inheritance was slowly gained through the accumulated work of many investigators from the nineteenth century into the middle of this century. A significant—and perhaps the initial—advance was made in the middle of the last

century by a botanist experimenting with plant hybridization. The discovery of the laws of biological inheritance by Mendel eventually led to the understanding of these mechanisms and provided the answer to a crucial question—the source of individual variation, a question that had plagued Charles Darwin.

PRINCIPLES OF INHERITANCE

Johann Gregor Mendel (1822–1884), often described as the founder of the science of genetics, spent most of his life as a member of the Augustinian order in a monastery in Brunn, Czechoslovakia. He had been an excellent student but had been forced to discontinue his studies because of ill health and poverty. On entering the priesthood he was able to continue his education, in part as preparation for teaching in the local secondary schools. Mendel studied in Vienna under leading natural scientists of the period, and far from being an isolated, obscure, ill-trained monk as has been described, he was well educated for the period. Most important for the future of genetics, Mendel came under the influence of Franz Unger, a botanist whose theory on the importance of varieties in natural populations probably was the stimulus that caused Mendel to begin work on the problem of inheritance (see Mayr 1982).

Whatever the influence, Mendel spent years studying plant hybridization, and he is best known for his extensive experiments on cross-pollination of common varieties of garden pea (*Lathyrus*). Mendel was fortunate in his choice of characteristics because they happened to be traits of simple inheritance: The plants bred true without intermediate traits—that is, each succeeding generation possessed traits like the parental generation. He cross-pollinated these plants for color, shape, size, and form of seed pod. Analysis of these multiple crosses led Mendel to derive the hypothesis that an organism's characteristics were inherited as discrete units or elements and not through a blending of parental traits, as was assumed in Mendel's day.

In some of his earliest experiments Mendel crossed plants that had violet-red blossoms with plants that had white (colorless) blossoms and produced hybrids that all had violet-red blossoms. But when these hybrids were crossed they produced a mixture of white and violet-red (Figure 2-1). Also, plants of different stem length were crossed (tall with dwarf), and the F_1 (first filial) generation were all of the tall variety. Crossbreeding of plants of this hybrid generation (the F_2) produced a mixture of tall and short plants. Mendel sought to explain these results by hypothesizing that these traits were determined by a pair of elements. One of the elements, or heredity particles, was dominant over the other, and they segregated independently in each generation—the *Law of Independent Segregation.* Mendel continued these kinds of experiments many times, and his results were close to a certain ratio of traits in the F_2 generation as diagrammed in Figure 2-1. The relative frequency of these traits is known as the *Mendelian ratio.*

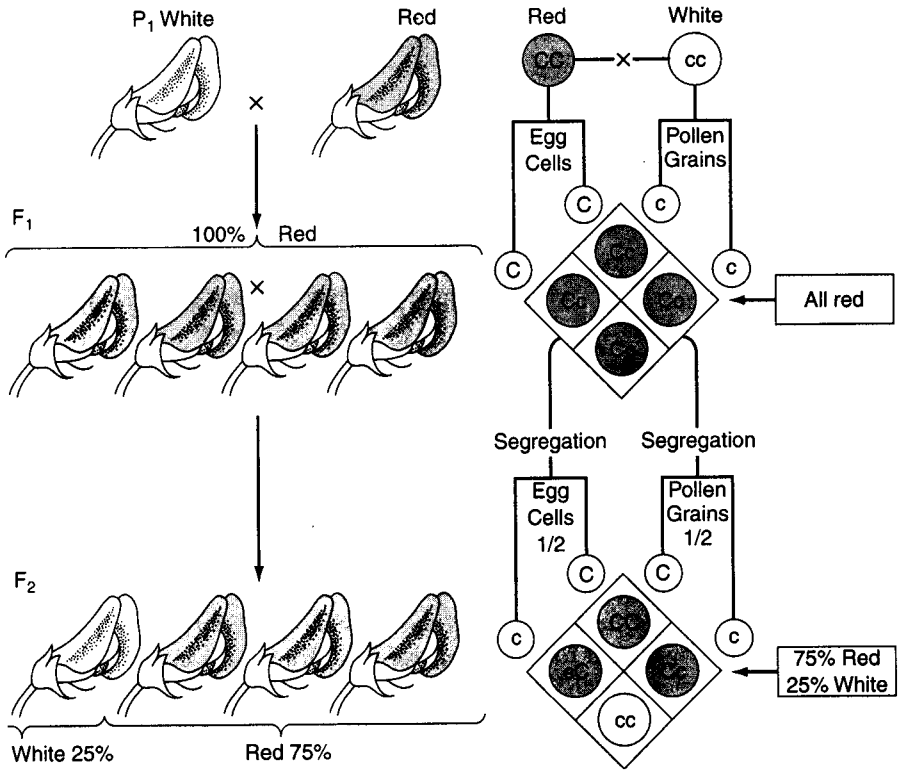


FIGURE 2-1 Diagram of Mendel's Experiments and the Results.

Experiments crossing plants selected for a difference of two traits produced dihybrids with a certain ratio of these traits among the F_2 generation, as was the case with crosses of single traits. These results demonstrated that traits such as seed shape and color were determined by paired elements that independently assorted in ovule and pollen (diagram in Figure 2-2). The repetition of such experiments produced results that could be predicted because they always fell within close range of the expected, establishing the *Law of Independent Assortment.* Thousands of experimental crosses of plants, selecting for single or paired traits, proved the correctness of Mendel's hypothesis and demonstrated the mechanisms of inheritance.

Mendel's work remained unappreciated during his lifetime, partly because cellular structures and their functions were just being discovered and partly because of a choice of traits that just happened to have a simple mode of inheritance. It was not until 1900 that particulate inheritance was recognized as the mode of transmission of characteristics between generations. Three botanical researchers (de Vries, Correns, and Tschermak), working independently on plant hybridization, provided experimental support. Within

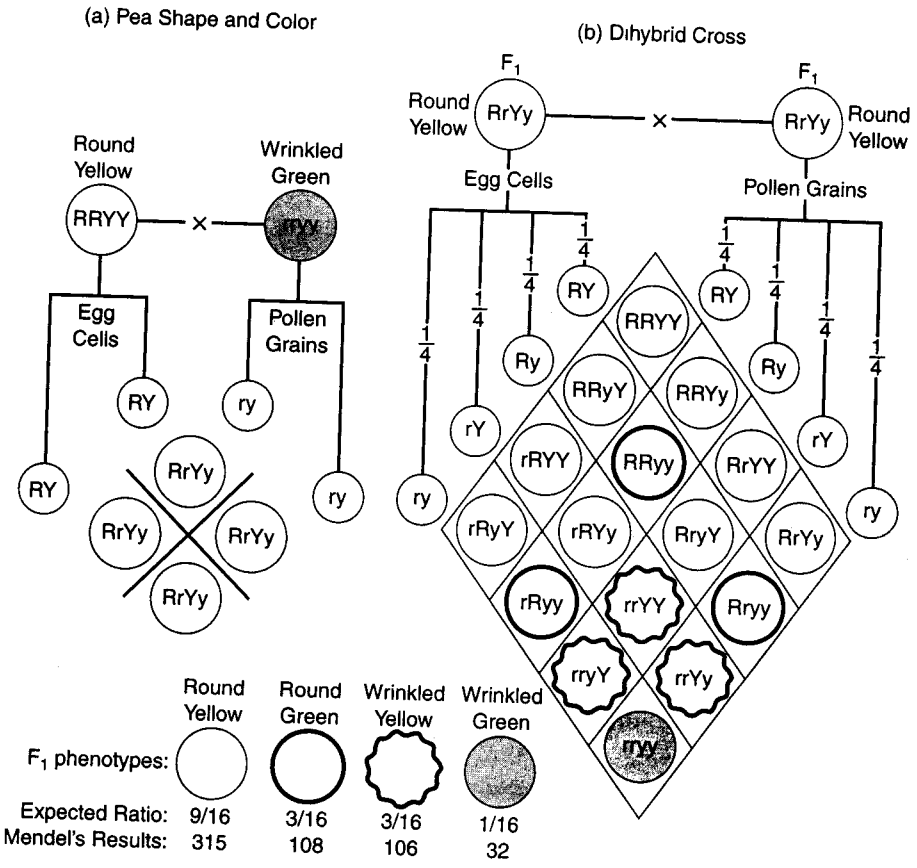


FIGURE 2-2 Independent Assortment: Mendel's Second Law.

When Mendel crossed plants, selecting for two characteristics at a time, he found that the paired factors for each character assorted independently. Diagram A shows the cross of plants to produce a "dihybrid" generation that, when crossed, produces four different characteristics, shown in diagram B.

less than a decade other scientists showed that inheritance of traits in animals also followed Mendel's laws. These studies and the thousands of experiments that followed during the early decades of this century established the foundations of modern genetics.

The Gene

In 1909, the element described by Mendel became known as the *gene,* a unit of inheritance—a term derived from the Greek root *gen* (to become or to grow out of). Each species has a specific quantity of genes numbering in the thousands or tens of thousands. It is estimated that our species, for example, has between 50,000 and 100,000 genes. Geneticists recognized early that

genes segregated just as did chromosomes, the darkly staining, threadlike bodies in a cell's nucleus. Groups of these genes are arranged lineally along the chromosomes. The locus, or position, of each gene in this linear sequence has a special significance for determination of a trait. For example, the reproduction of dihybrids for color and seed shape, as in Mendel's experiments, suggested that the locus for seed color is on a different chromosome than the locus that carries the gene for seed shape. In addition, there may be more than one form of gene for each locus—for example, one that determines that the seed is green or one that determines that the seed is yellow. These alternate forms of the gene for a particular characteristic were called *allelomorphs*, from which the term *allele* is derived as used today to describe the variety of gene forms of a trait. Again, in reference to Mendel's study, we see that some alleles are dominant to others, as was the case with plant color shown in Figure 2-1. The paired combination of alleles, one carried at a locus on each of the chromosomes of the pair, is called the *genotype*. Hence, the genotype, or heredity type, for color may be CC, Cc, or cc. The trait that is the result of the genotype combination is the *phenotype* (the visible type or trait).

Chromosomes and Cell Division

Each cell of an organism contains several pairs of chromosomes within its nucleus. When the cell grows and eventually divides, as in cell reproduction, these chromosomes undergo several changes that alter their shapes prior to division. They reorganize from the irregular threadlike bodies of darkly staining material to form shorter, thicker structures. The chromosomes are recognized as independent bodies at this stage and each appears as two joined strands. These strands are called *chromatids* and are held together at a point along their length called the *centromere*. During cell division, or *mitosis*, the chromatids of each chromosome are pulled apart and each is attracted to opposite poles of the cell, which become the center for the formation of the daughter cells (see Figure 2-3). The end result of mitosis is to double the number of cells with an even distribution of chromosome materials between the daughter cells. During the interphase, or resting stage, the missing halves of chromosomes (the chromatids) are duplicated from materials in the cells' cytoplasm, and the chromosomes are then completed and will be ready for the next cell division. The splitting of each chromosome in half and the movement of the chromatids into the daughter cell ensures that each cell has a full and identical complement of genes. Such a process allows tissues to grow and still maintain their identity and special functions.

The number and sizes of chromosomes of the body's cells, or *diploid* number, is fixed for each species. For example, this distinctive array, or karyotype, in *Homo sapiens* has forty-six, whereas the chimp and gorilla have forty-eight and the gibbon forty-four. The forty-six chromosomes in our species are arranged as twenty-three pairs, of which twenty-two are known as *auto-*

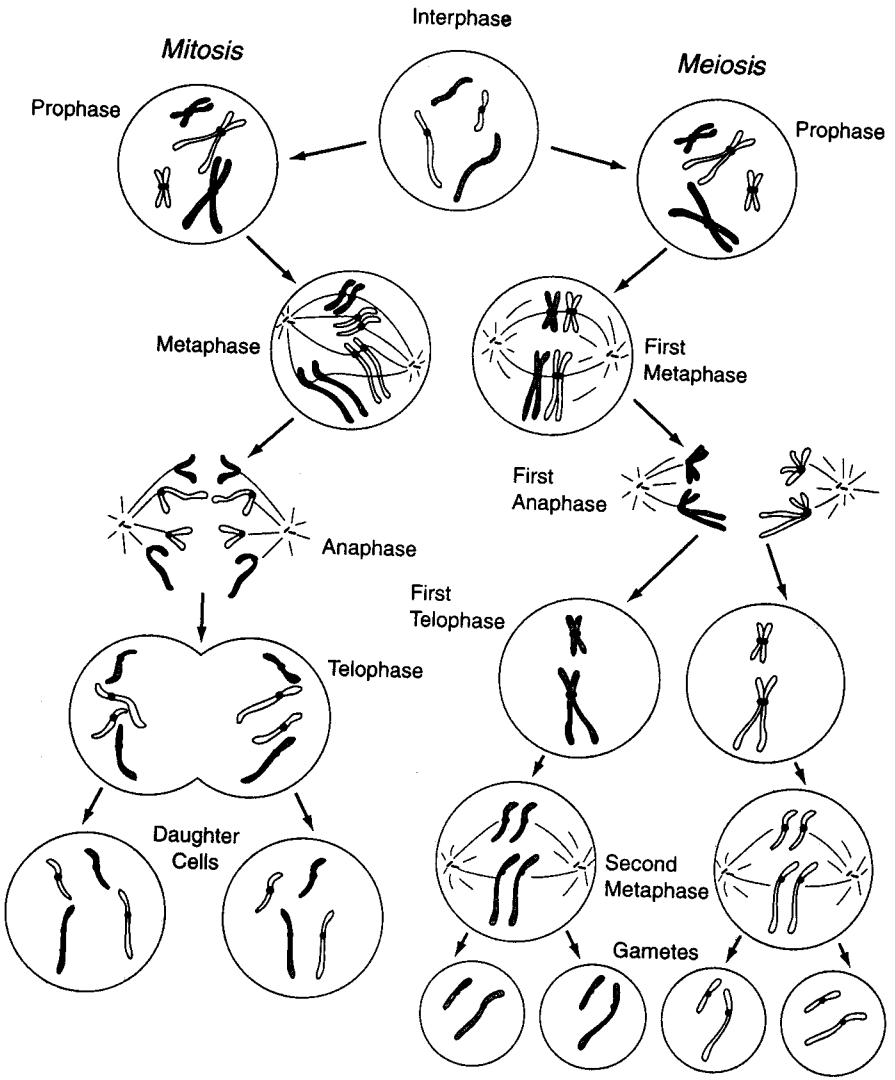


FIGURE 2-3 Stages of Cell Division.

These diagrams show the major steps that occur in cell division. A comparison between *mitosis* and *meiosis* illustrates the organization of the chromosomes at each stage.

Mitosis: This simple cell division starts with chromosomes as pairs of chromatids joined at some point along their length by a centromere. Through metaphase and anaphase stages the chromosomes are arranged in the equatorial plane of the dividing cell and, finally, the chromatids are pulled apart by the end of the anaphase. Telophase is the stage during which the cell membrane grows and eventually separates into two daughter cells.

Meiosis: The major distinction of meiosis is that through a process of reduction and division, daughter cells are produced that have one-half of the chromosomes of the parent cell. The chromosomes are aligned side by side along the equatorial plane in the first metaphase rather than end to end as in mitosis. By the end of the first anaphase the pairs have been separated into cells at a telophase stage, but these cells continue to divide. The second metaphase separates the chromatids into the germ cells called *gametes*.

somes and one pair, the sex chromosomes, are responsible for initiating sex determination. The sizes of the chromosomes of the same pair are identical, and, though there is some similarity between certain pairs, the structural uniqueness of each pair of autosomes sets them apart and prevents the combining of chromosomes from different pairs, though occasionally a fragment from one chromosome will attach to a chromosome of another pair (nonhomologous). This *translocation* frequently causes severe disruption of the cellular functions, which almost always leads to destruction of the cell. If abnormal chromosome form or number occurs, the *zygote* (fertilized ovum) will not grow and divide beyond a few divisions, and only rarely will it reach the embryo stage. There are, however, occasional abnormal combinations of chromosomes. The best-known example is *Down syndrome,* formerly called Mongolism.[1] An individual with this affliction has forty-seven chromosomes instead of forty-six because of an extra chromosome at the twenty-first pair, or *trisomy.* Down syndrome includes a group of abnormal physical traits in addition to a varying degree of mental retardation. Several other syndromes, described below, are due to abnormal numbers of sex chromosomes.

Because the number of chromosomes is critical and must remain constant from one generation to the next, a basic problem of sexual reproduction is how to ensure that an equal number of chromosomes are passed on to the next generation. Because sexual reproduction involves a combination of materials from two individuals in order to produce the offspring, this problem of the maintenance of the species' chromosome number is solved by a process of cellular reduction and division known as *meiosis.* Meiosis is, to a certain extent, comparable to mitosis of somatic cell division, with several important exceptions (see Figure 2-3). A major distinction is one of chromosome number: The dividing cell separates the homologous chromosomes shortly before division. The cells, now with only twenty-three chromosomes or one from each pair, continue to divide and, during the second metaphase, the chromatids are pulled apart. The final stage produces cells, the *gametes,* with one half the number of chromosomes, the *haploid* number.

The germ cells, whose major function is production of the gametes (eggs in the case of a female or sperm in the case of a male), are formed in specialized tissues found in the gonads. These cells undergo *meiosis,* which divides the chromosome pairs to form a gamete able to combine with a gamete of the opposite sex in order to form the fertilized egg or *zygote.* This fertilized egg pairs up chromosomes from each parent in order to duplicate the proper number of chromosomes for the species. This process of sexual

[1]Langdon Down, a nineteenth-century London physician, described patients with a particular type of congenital mental deficiency as "typical mongols." These individuals, because of their general appearance of a broad, flattened face, epicanthic eye folds, and other features, were likened to the Mongoloid race.

reproduction is one of the most fundamental and important factors in the introduction of new varieties because it combines materials from two individuals. During meiosis each chromosome segregates independently from all of the others. Therefore, chromosomes that were provided to the individual at conception by the gamete from the male or female parent are often separated, so it is highly improbable that a person's gametes will contain an even distribution of the chromosomes that were inherited from each of the parents. Of the twenty-three individual chromosomes contained in a particular gamete, for example, fifteen may have been derived from those inherited from one parent and the remaining eight from the other parent. This independent assortment of chromosomes during meiosis is one kind of *recombination* that occurs during meiosis and contributes to diverse combinations of genes in each gamete. The mixing of proportions of one's maternal and paternal chromosomes during meiosis generates a variety of gametes; the total number of gamete types that can be produced by humans is 2^{23} or over eight million.

Another type of recombination, and one that is of primary importance in its influence on gamete diversity, is *crossover* during an early stage of meiosis. Crossover refers to an exchange of parts of non-sister chromatids of homologous chromosomes. The homologous chromosomes align in pairs, or *synapsis.* The chromatids of the pair of chromosomes are closely bound into a tetrad bundle, and when they begin to separate to opposite poles of the dividing cells there is a swapping of parts of the non-sister chromatids, as illustrated in Figure 2-4. This breakage and rejoining after an exchange of corresponding parts is called *chiasma,* which causes a realignment of the linear arrangement of genes along each chromosome, and the frequency of this occurrence or the chance that it will happen depends on the distance between gene loci (Figure 2-4).

Neither type of recombination adds new genetic information into the population. It merely reassorts the genes so that individuals in each generation will have different gene arrangements and combinations, causing each person to be a unique creation. Because these gene arrangements or genotypes, discussed earlier, influence the characteristics, *recombination* is an important source of individual variability.

The Sex Chromosomes and Sex Determination

Sex determination is a complicated multistep process beginning at conception with the fertilization of the egg by the sperm. If an X chromosome is provided by the sperm, then the zygote will have an XX pair and will usually possess the genetic equipment to develop into a female. In the early weeks after conception, the embryo tissues begin to differentiate, and the region that will become the urinary tract and reproductive organs reaches a level of development with sex undifferentiated. There is a potential for becoming female should certain

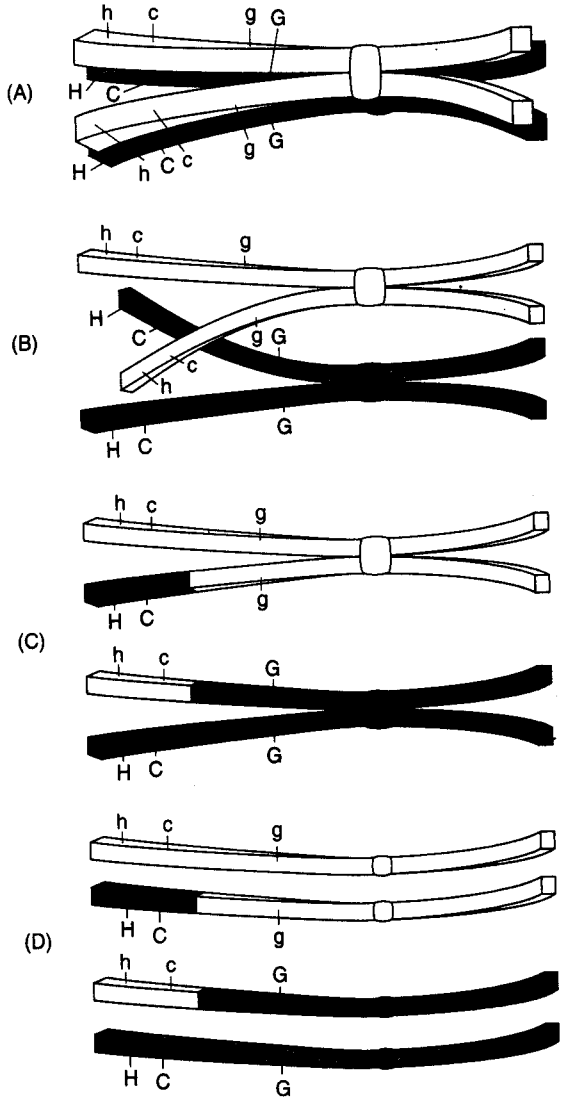


FIGURE 2-4
Crossover of X-Chromosome Fragments during Meiosis.

Occasionally, there is an exchange of fragments between homologous chromosomes (members of a pair). This exchange may take place due to breakage of chromatids during the prophase stage (shown in Figure 2-3). The parts are then rejoined to the other chromosome. The model of the X-chromosome illustrates the swapping of the section with the h allele (hemophilia) and c allele (colorblindness) with the fragment carrying H (normal allele) and C (normal visual). This event occurs during prophase, steps A and B. Step C shows a separation of the chromosomes—each is broken apart into its chromatids, step D. The final result is a realignment of genes.

conditions continue to prevail—that is, the embryo is a presumptive female. The ducts (Mullerian) that give rise to the ovaries and reproductive organs of the female will develop, while those ducts (Wolfian) that are precursors of male reproductive organs will regress. By the twelfth week the female sex will be established as the embryo enters the fetal stage of development.

If, however, the sperm carries a Y chromosome, then a complicated process of differentiation begins, initiating a series of steps leading to maleness of the embryo and hence of the fetus. The short region of the Y chromosome

carries a gene essential for maleness—the testes determination factor (TDF). The products of this gene will stimulate the development of testes from the gonadal ridge by the sixth embryonic week. As this development proceeds, a potent form of testosterone is secreted that initiates a series of steps that leads to sex identification. One of the major changes initiated is the alteration of the neural pathways within the hypothalamus of the brain, the controlling center for endocrine function. This pathway regulates ovarian activity and controls menstrual cycling in the postpubertal female. However, the testosterone secretion of the male embryo changes this pathway in the male direction. The testosterone secretions also stimulate differentiation of the Wolfian ducts and formation of the male reproductive system consisting of the prostate gland, seminal vesicles, and vas deferens of the testis; they also organize the shaping and growth of the external genitalia. Another gene (the H-Y), carried on the long region of the Y, has been described as a possible controlling factor in sperm production. In addition, there are controls for cellular receptors of the male hormone, or for maintenance of androgen–estrogen ratios. Other hormones, under the control of genes carried on the Y chromosome, are being discovered even at this writing, and these add to the complexity of biochemical and physical sex identity. In summary, the early embryo stage follows a basic development plan that is female unless altered by the action of the products of certain Y-linked genes.

Chromosome determination of sex usually proceeds as expected and sex identity is established, but on rare occasions the opposite turns out to be the case. About once in 20,000 male births, there is an individual born with a pair of XX chromosomes. What has occurred is that during meiosis in the male parent a fragment of the Y carrying the gene for testes development (the TDF part) breaks and attaches itself to an X, which then is passed on in the sperm that by chance fertilizes the maternal parent's ovum. This means that even though the offspring has the XX pair, he still possesses the TDF critical for stimulating the products that cause the development of a male embryo. These rare individuals are, however, sterile since they lack the region of the Y chromosome necessary for sperm production. By contrast, XY females have occasionally been born, and here the explanation is similar in that the TDF was also involved, but the effect was opposite since it involved the lack of the TDF gene. During meiosis in the male parent, the TDF-bearing fragment of the Y was lost and a sperm was produced with this deficiency. Hence, an embryo bearing the X plus defect Y would develop, following the basic female body plan because it lacked the hormonal stimulation to determine the male sex.

There are numerous other variations in sex determination recorded, but these mostly involve differences in the number of the sex chromosomes. Such deviation from the sex chromosome pair usually yields an individual with abnormal developmental characteristics. Occasionally (about once in every 400 male births) an extra X chromosome is combined with the XY pair. The individual has a diploid number of 47 and is an XXY male with poorly developed sexual characteristics together with some female ones as well (*Klinefelter*

syndrome). Males with an extra Y chromosome (47, XYY) have also been recorded. These are normal males, with the exception of their greater-than-average height. Early studies of this condition described a possible association with certain behavioral pathologies and pointed to a supposed high frequency of the XYY condition among mental patients and prison inmates. Subsequent studies, however, found that only a small number of individuals with this syndrome were institutionalized (about 4 percent), whereas the remaining 96 percent of XYY males had normal behavior patterns indistinguishable from those of the rest of the male population. The presence of an extra Y chromosome does not predispose a male to social pathology, but it has remained a karyotypic curiosity much misunderstood for a long time (Witkin et al., 1976). Another example is the birth of a female with only a single X chromosome (once in every 3,500 female births). She will have a series of anatomical defects known as *Turner syndrome* (45, XO) and the diploid number will be 45 instead of the normal 46. Such individuals have poorly expressed secondary sexual characteristics and tend to be shorter than normal.

The major significance of the X and Y chromosomes, in addition to sex determination, is the influence that the genes that are carried on these chromosomes have on the development of secondary sexual characteristics of form, final adult size, and growth rate and pattern during adolescence. Similar to the distinguishing influences seen in embryonic development, the sex chromosome differences continue to influence child growth. Females reach puberty and pass through their adolescent growth spurt an average of two years earlier than males, and during this growth period they acquire the secondary sexual characteristics that so distinguish male and female. Bodily proportions depart from the childhood form as the pelvic girdle grows more rapidly than the pectoral region (across the shoulders). But linear growth ceases sooner than in the male, resulting in a lower average height. Head and face growth also proceed more slowly, and females retain more of a childlike shape in these two regions. Males in most populations are significantly larger in body size and differ in bodily proportions. They differ also in body hair distribution and density from females, especially in facial hair. These differences, and more (detailed later in Chapter 5), are the result of certain genetic differences in the sex chromosomes, especially the Y-linked genes. Many of the male–female differences are the result of X and Y size contrasts.

The sizes of the twenty-two autosomes plus the sex chromosomes (the human karyotypes) are diagramed in Figure 2-5. Since the autosomes exist as pairs, each homologous chromosome being identical to its mate in size and shape, only a single member of each pair is shown—the haploid number. There is a comparable region on each member of the pair. For example, certain parts of each chromosome will take up a chemical stain, and these darkly stained areas of one chromosome will have a comparable location on the other one of the homologous pair. These stained regions are grouped on each chromosome into a *p* and a *q* region, the short and long arms above and below the centromere where the chromatids are joined. The X chromosomes also pair up in a female, and homologous regions exist. A different situation exists in the male, however. There is little homology between the X and Y chromosomes.

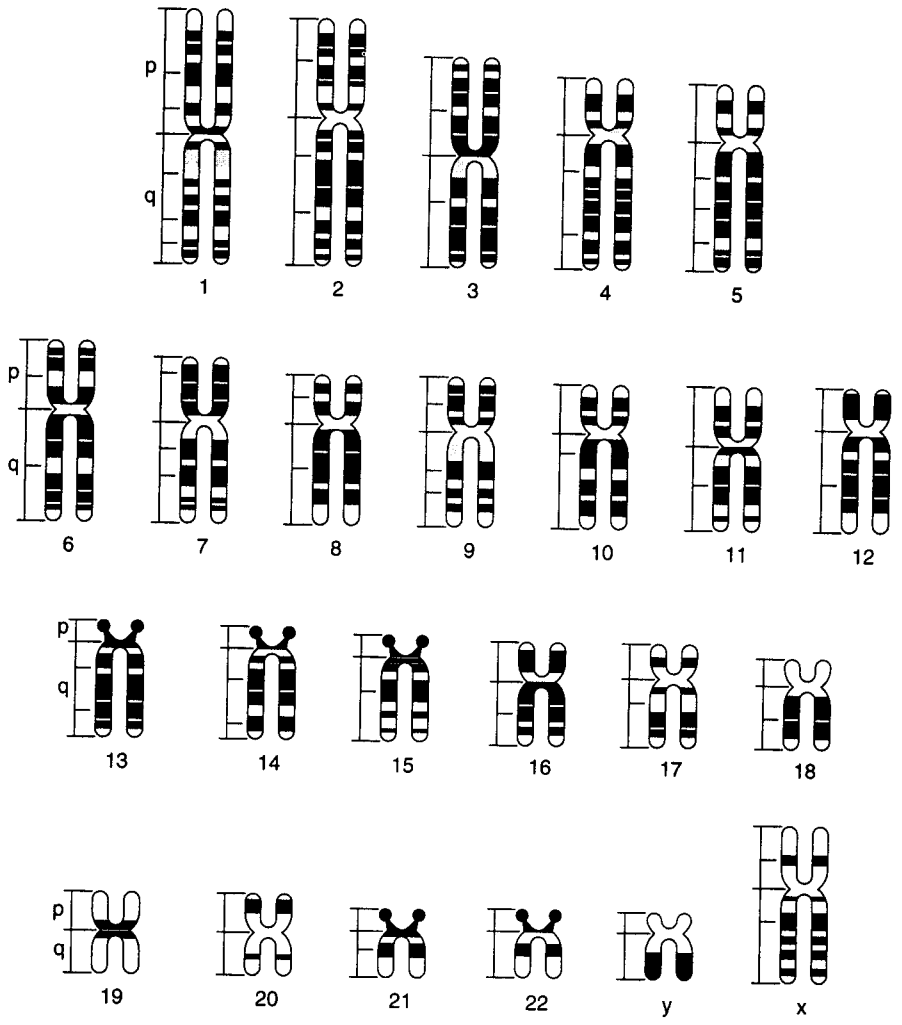


FIGURE 2-5 Human Chromosomes in Mitosis.

This illustration is drawn from a photomicrograph of human chromosomes in mitosis. The twenty-two autosome pairs are grouped according to size (karyotype) and the sex chromosomes are placed separately after pair twenty-two. The dark bands illustrate the locations stained by specific chemicals. The short and long arms of each chromosome are designated by *p* and *q* to assist in locating particular sites.

The Y differs considerably in size and structure from the X. The Y is shorter, so that except for a very small region at the tip of the short arm (*p* region) there is no corresponding region on the X (see Figure 2-6). Any genes

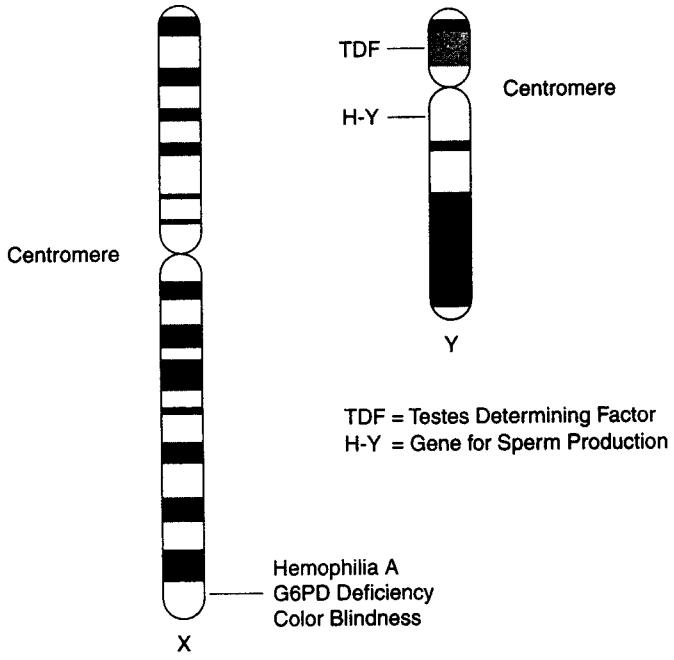


FIGURE 2-6 X and Y Chromosomes and Nonhomologous Region.

appearing on the nonhomologous region of the X chromosome will not be paired up in male cells, since the male is hemizygous (having only a single X chromosome). Therefore, a certain number of genes on the nonhomologous region of the X will express a gene product without any influence from a dominant allele that would be present in the female. This leads to the existence of certain recessively determined traits that occur more frequently in the male than in the female—for example, *hemophilia A* (the most common form of an inherited defect in blood clotting), and two genes for color sensitivity of the cones of the retina (the protan for the synthesis of a red-sensitive pigment and deutran for a green pigment). A recessive allele of either of these closely linked genes will result in defective vision for color in the red–green part of the spectrum that occurs in about 8 percent of males of western European ancestry and is commonly called *color blindness.* There is a third X-linked gene that in recessive form causes a deficiency of an important enzyme of carbohydrate metabolism, *G6PD,* whose function is detailed in Chapter 4. All three of these genes—hemophilia A, colorblindness, and G6PD—are positioned in nearby loci in a region at the lower end of the long arm of the X chromosome (Figure 2-6).

Little is known about Y-linked genes, although major advances have been made in mapping the Y chromosome in the last few years. The *p* region carries genetic loci for maleness, as described above, and some exchange during meiosis may be made of this psydohomologous region with the X. But few

other traits have been traced to genetic loci on this short chromosome. A gene for male pattern baldness is one possibility, and a peculiar hairy growth over the outer edges of the ears, hairy pinna, are two traits that have been frequently mentioned. More recently, studies of regions of the Y chromosome (in some sexually dysfunctional males) with the aid of restriction enzymes (DNA probes—see below) have identified certain regions with particular influences on skeletal and dental growth, testes development, and spermatogenesis (see Vogel and Motulsky, 1986, and McKusick, 1994). These and other studies are beginning to define the role of the Y-linked genetic loci in the development of male secondary sexual characteristics.

THE GENE, DNA, AND THE "CODE OF LIFE"

Knowledge of the gene as a unit of inheritance underwent a slow but steady advance over the first half of this century after the rediscovery of Mendel's experiments. The earliest and perhaps the most important of these advances was an understanding of the intergenerational transmission of traits; parents do not pass on characteristics but transmit the "information base" needed for their development. The association of such information with the darkly staining material in a cell's nucleus was made early in the century, about the time the term "gene" was used to designate the hereditary unit of information. This was followed by a recognition of chromosome pairs, where the genes were thought to occupy a position or *locus* on each. The further advancements made in studies of cell structures and how they divide clearly demonstrated the basic mechanisms of particulate inheritance.

Understanding of the actual nature of gene structure and its varying functions, and of how genetic mutations occur, had to wait until a model was offered to define the structure of the nucleic acid components within the cell's nucleus. Though nucleic acid had been long suspected to be the hereditary material within the nucleus, the connection could not be made to the proteins that directed cellular processes and growth. These processes were thought to be dependent on the chromosome's involvement in the synthesis of products necessary for the cell's metabolism. The composition and structure of the nucleic acids seemed to be the key to understanding not only the nature of heredity but the functioning of the entire organism as well.

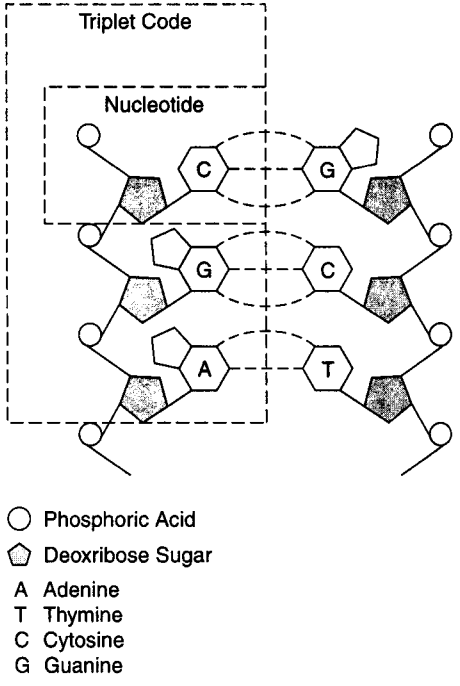
This led to a search for these products, and for many years investigators worked, with some success, to describe the structures of complex molecules, like proteins that were believed to control cellular metabolism or, in some cases, formed the basic components of body tissues. But the structure of the nucleic acids and their relationship to protein molecules escaped definition until mid-century. In 1953, James Watson and Francis Crick offered a model to explain the molecular structure of a compound, *deoxyribonucleic acid* (DNA), whose existence in the nucleus had been known for years. The

model proved to be an accurate description of this complex structure, and their discovery had a momentous impact on biology and was just the type of breakthrough that the field needed to start a new phase of genetics research. The discovery was so important and basic to the understanding of the genetic code of life that Watson and Crick were awarded the Nobel Prize in 1962.[2]

DNA is a long, repetitive, chainlike structure made up of alternating phosphate and sugar (deoxyribose) molecules to which are attached one of four kinds of organic bases (thymine, adenine, cytosine, or guanine). The sugar-phosphate molecules form a basic backbone structure of DNA. The unit composed of sugar, phosphate, and base molecules is called a *nucleotide* (Figure 2-7a), which is joined with the next nucleotide, and this process is repeated over and over until a long chain has been formed. The bases of the nucleotides are attracted to other bases on a complementary DNA strand and the two are held together by a weak hydrogen bond. Each base only attracts one other type; thymine (T) is bonded to adenine (A) and cytosine (C) to guanine (G). The length of the two chains can be diagramed as a ladderlike structure; the long parallel structures are formed by the sugar and phosphate backbones while the connecting rungs are the complementary base pairs (Figure 2-7b). The DNA strands are actually rotated about each other to form the double helix described by Watson and Crick.

FIGURE 2-7A
Nucleotide Structure, the Basic Unit of the DNA Molecule.

A nucleotide is composed of a molecule of phosphate and a deoxyribose sugar to which is attached any one of four types of organic bases: adenine (A), thymine (T), cytosine (C), or guanine (G). This basic unit is attached to an adjacent nucleotide by bonding between phosphate and sugar molecules as shown. Three nucleotides, taken together, provide a particular *triplet code* because of the combination of the three organic bases they contain, and this code specifies a particular amino acid as discussed in the text.



[2]Many researchers in several fields laid the groundwork for molecular genetics, and Watson relates a very interesting and personal account of the events leading to the discovery of the DNA structure. He also describes the fierce competition among scientists to be the first to identify the functioning of this key molecule (see *The Double Helix*, 1980).

The importance of the base-to-base attractions is well illustrated when a cell divides. The DNA strands pull apart and the unbonded bases attract new nucleotides and bond with complementary bases. These nucleotides attach over the length of the original strands, forming two new double helical molecules, as illustrated in Figure 2-7c. This DNA replication at cell division is described as semiconservative; one of the old strands is joined with a newly formed strand so each of the daughter cells will end up with its proper DNA complement. This process occurs repeatedly during cell divisions to produce new tissue and, provided that all replication is correct (no mutations), the daughter cells will be identical to parental cells.

The Gene: Structure and Function

Once the nature of the nucleic acids of the cell nucleus was described, the search was on for an explanation of how they related to cellular functions and division,

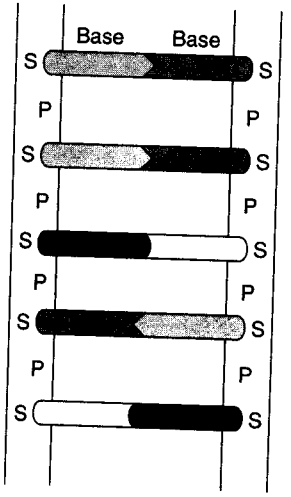


FIGURE 2-7B
Schematic Diagram Depicting the Ladderlike Arrangement of DNA.

The organic bases of opposite DNA strands are attracted and bound together by a weak hydrogen bond that causes DNA to be a double-stranded molecule. Because of their chemical structures, adenine will bind only with thymine and guanine with cytosine. These base-to-base bonds form the "rungs" of the "ladder" while the series of sugars and phosphates are long side pieces to which the rungs are attached.

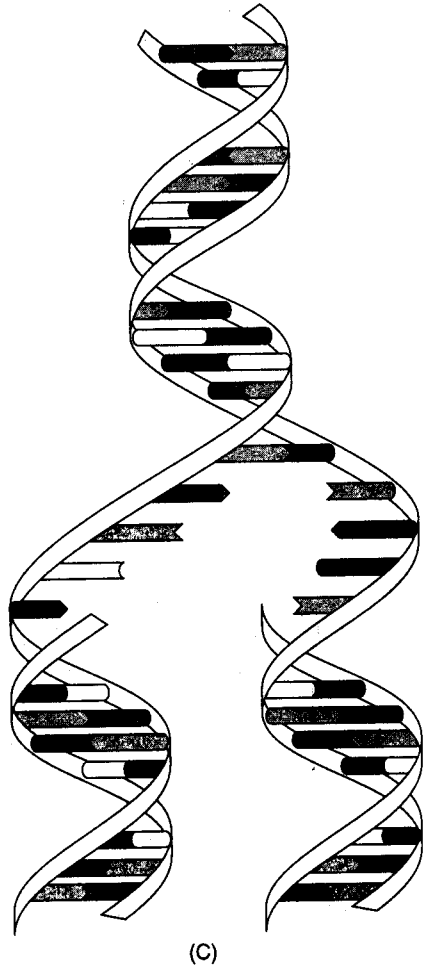


FIGURE 2-7C
Helical Shape of DNA Molecule and Semiconservative Replication.

DNA molecules are rotated so they form a double helical structure; the sides formed by the sugar and phosphates and weakly bound bases are connected as shown in Figures 2-7a and b. When a cell divides the DNA molecules must be duplicated in such a way that the new cells will have the exact quanity and sequence of nucleotides. This is achieved by a process of separation of part of the molecule at a time and bonding the complementary nucleotides to the exposed unpaired bases. When completed, this process has created two daughter helixes, each composed of one of the original DNA strands bound to a newly synthesized one, a process called semiconservative replication.

and what relation these acids had to the transmission of traits throughout generations of cells. The search was directed to protein molecules because of their functions. They provide support (structural proteins) as in the example of collagen, an important protein of skin, bone, and many other tissues in the body; regulate metabolic processes (enzymes), as in digestion, body temperature control, and production of skin pigment (tyrosinase); and influence gene expression through the action of thousands of enzymes in a cell or by hormones.

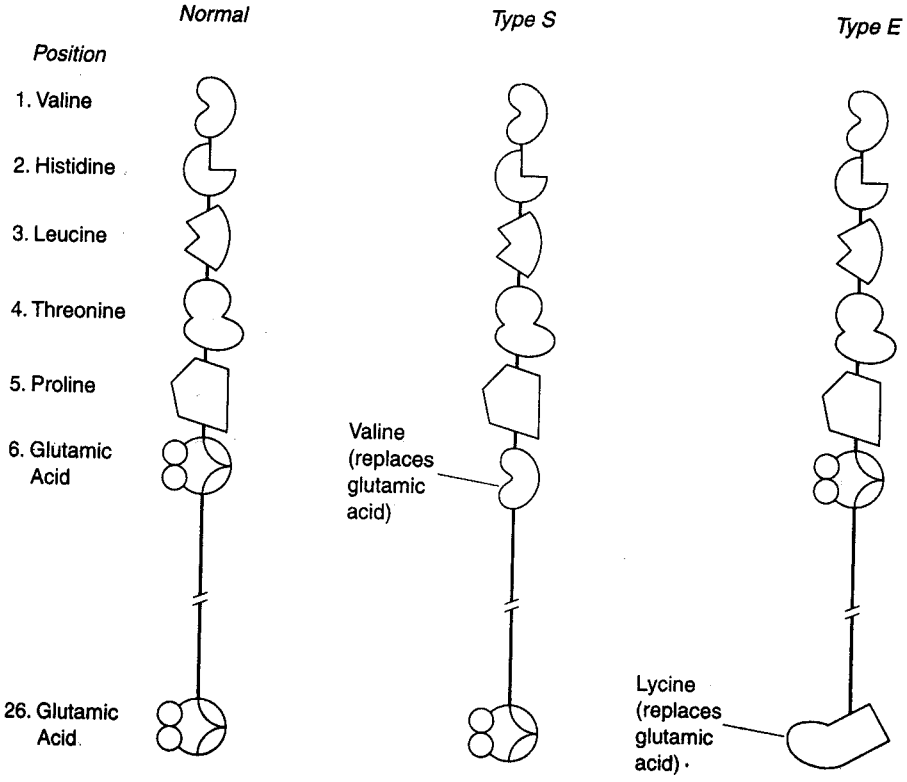
Proteins constitute a class of chemical compounds made up of chains of smaller molecules (amino acids) linked together by peptide bonds.[5] The total

[5]The major organic molecules of living organisms are classified into four categories: carbohydrates (sugars and starches), lipids (fats), proteins, and nucleic acids.

of these linkages is described as a polypeptide chain. The proteins are composed of 20 kinds of amino acids arranged in a linear chain in some combination. The chain may be a few dozen amino acids long, as in human insulin, which has 51 arranged in 2 tightly bonded chains, or a protein may contain a hundred or more amino acids, as in globin of human hemoglobin, with 574 organized into 4 polypeptide chains, 2 alpha and 2 beta. The sequence or linear arrangement of the amino acids is critical and provides the protein molecule with a specific identity and hence its function. In Figure 2-8 the normal sequence is shown for the first 6 positions and for position 26 of the 146 in the beta chain of human hemoglobin. A substitution of amino acid, valine, for glutamic acid changes the identity from hemoglobin A to hemoglobin S (sickle cell type), and under certain conditions its function (oxygen transport) is radically altered. Other substitutions also change hemoglobin type as

FIGURE 2-8 Amino Acid Sequences of Three Types of Beta Globin.

These diagrams depict the amino acids at the first 6 and at the 26th position. Normal, type S, and type E hemoglobin have the same sequence of amino acids at all 146 positions of the beta globin chains with two very important and specific exceptions. Type S has a substitution of valine for glutamic acid at position 6 while type E is the same as the normal for that position but has a replacement of lycine for glutamic acid at position 26.



shown. This importance of amino acid sequence raises the question of how cells and structures that synthesize proteins direct the correct organization of the polypeptide chains.

Because of a pattern of inheritance of different forms of proteins observed in family lineages over the years, protein synthesis was thought to be under genetic control. But the genetic material, though considered to be in the nucleus, was not identified until after mid-century. The major compounds within the nucleus, the nucleic acids, were at first ignored as the genetic code because their chemical composition, analyzed long before the Watson and Crick discovery of their structural arrangement, showed a presence of only four organic bases. Any hypothesis that these bases existed in a regular structural sequence made it difficult to understand how only four units could determine the linear arrangement of twenty types of amino acids into a string of dozens or more. Also, the proportions of the bases varied between DNA and RNA compounds. The major work after 1953 provided the answers and opened up the era of molecular genetics, which has enormously expanded the understanding of cellular function and its inherited basis.

The Genetic Code

Considering DNA as the genetic code, the problem of amino acid identification is solved when the nucleotides are "read" three at a time as a group instead of individually. Since each nucleotide is identical (phosphate and sugar) except for one of the four types of organic base attached, a group of three nucleotides gives the probability of 4^3 or 64 different coded combinations (three positions at which one of four kinds of organic bases may occur). This code then could account for more than twenty amino acids with a number of "codes" left over. After much research, different DNA triplets were shown to code for particular amino acids, and some amino acids could be coded for by any one of several triplets; the DNA code was said to be redundant (see Table 2-1). The problem that remained to be solved was how the DNA code in the nucleus could control synthesis in the cell's cytoplasm where the structures (ribosomes) and raw materials needed for the protein synthesis were located.

RNA. This second nucleic acid compound (ribonucleic acid) proved to be the transporting agent or *messenger* that copied the code and relayed it to the sites of protein synthesis, the *ribosomes.* Figure 2-9 diagrams the basics of this process. It starts in the nucleus when an enzyme, *RNA polymerase,* causes the double strand of DNA to separate along a few triplets. The unbonded bases of one of the strands are temporarily bonded to complementary bases of RNA triplets (codons) that are formed into a chain as the enzyme moves along the DNA. Once a transcription has been made, the mRNA chain segment separates from the DNA. This process is repeated until a terminating triplet is reached (ATT, ATC, or ACT in Table 2-1). At this point a single-

TABLE 2-1 Genetic Code[a]

AMINO ACID	DNA (TRIPLET)	mRNA (CODON)
Alanine (ala)	CGA, CGG, CGT, CGC	GCU, GCC, GCA, GCG
Arginine (arg)[b]	GCA, GCG, GCT, GCC, TCT, TCC	CGU, CGC, CGA, CGG, AGA, AGG
Asparagine (asn)	TTA, TTG	AAU, AAC
Aspartic acid (asp)	CTA, CTG	GAU, GAC
Cysteine (cys)	ACA, ACG	UGU, UGC
Glutamic acid (glu)	CTT, CTC	GAA, GAG
Glutamine (gln)	GTT, GTC	CAA, CAG
Glycine (gly)	CCA, CCG, CCT, CCC	GGU, GGC, GGA, GGG
Histidine (his)[b]	GTA, GTG	CAU, CAC
Isoleucine (ile)[c]	TAA, TAG, TAT	AUU, AUC, AUA
Leucine (leu)[c]	AAC, GAA, GAG, GAT, GAC, AAT	UUG, CUU, CUC, CUA, CUG, UUA
Lysine (lys)[c]	TTT, TTC	AAA, AAG
Methionine (met)[c]	TAC	AUG
Phenylalanine (phe)[c]	AAA, AAG	UUU, UUC
Proline (pro)	GGA, GGG, GGT, GGC	CCU, CCC, CCA, CCG
Serine (ser)	AGA, AGG, AGT, AGC, TCA, TCG	UCU, UCC, UCA, UCG, AGU, AGC
Threonine (thr)[c]	TGA, TGG, TGT, TGC	ACU, ACC, ACA, ACG
Tryptophan (trp)[c]	ACC	UGG
Tyrosine (tyr)	ATA, ATG	UAU, UAC
Valine (val)[c]	CAA, CAG, CAT, CAC	GUU, GUC, GUA, GUG
Terminating triplets	ATT, ATC, ACT	UAA, UAG, UGA

[a]Symbols for bases in nucleic acids: A (adenine); C (cytosine); G (guanine); T (thymine); and U (uracil), used as a substitute for T.
[b]Essential in diet of young child.
[c]One of the eight amino acids that humans cannot synthesize and therefore must obtain from the diet.

stranded mRNA (messenger) has been produced, and the DNA strands are rejoined into the double helix as before.

The mRNA is only a primary messenger, however, because it includes a number of noncoding sequences of the gene, the *introns.* The primary mRNA undergoes a process of maturation that causes a looping of the strand over the intron area, which is then cut and discarded. The remaining ends are joined, linking together those segments of DNA that code for proteins, the *exons.* The finished mature product of mRNA then moves out to the ribosomes in the cytoplasm, the site of protein synthesis. This single chain of triplets (codons) provides an attraction for the bases of short strands of transfer RNA (tRNA) to attach temporarily. Each tRNA or anticodon (because its bases are complementary to mRNA) carries a particular amino acid. When the codon–anticodon bonds are established, the amino acids, carried by the tRNA, form peptide bonds, and the process is repeated until the stop signal is reached. The completed protein molecule is thus produced (Figure 2-10). The tRNA are free to pick up more amino acids and the process begins all over. The triplets (codons) of mRNA are listed with the corresponding DNA triplets and the

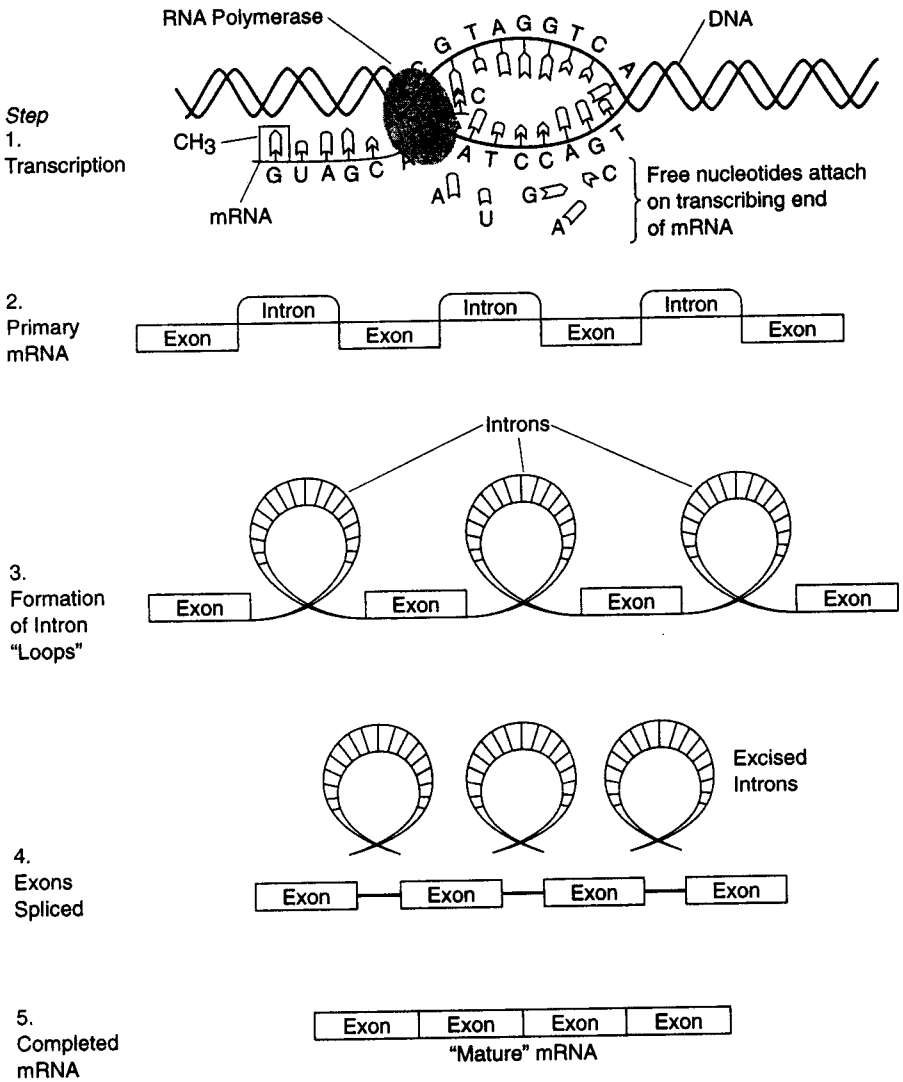


FIGURE 2-9 Transcription of DNA and the Synthesis of Messenger RNA (mRNA).

The five basic steps in the process of synthesizing mRNA to transcribe the DNA code are illustrated in this diagram. *Step One* begins by the separation of a segment of the double helix by the enzyme, RNA polymerase. The exposed bases of one of the DNA strands attract free nucleotides which become bound to one another, forming the single-stranded mRNA, which is a process similar to the formation of complementary daughter strands of DNA during replication. However, the major exception is that the mRNA soon separates, remains single-stranded, and uses the base uracil to bind with adenine instead of thymine. The enzymes move along the DNA and the process is repeated, and when a length of DNA has been "read" the strands rebind to one another.

Step Two is reached when the transcription is completed and the primary mRNA is separated. At this step the mRNA contains a copy of all of the DNA, both exons and introns.

Step Three is a maturing process that causes the introns, noncoding portion, to contract and form loops that result in shortening the mRNA, bringing the exons closer together.

Step Four excises the intron loops and splices the exons together.

Step Five produces the now shortened and completed strand of mature mRNA that contains the code to direct the production of polypeptide chains of amino acids.

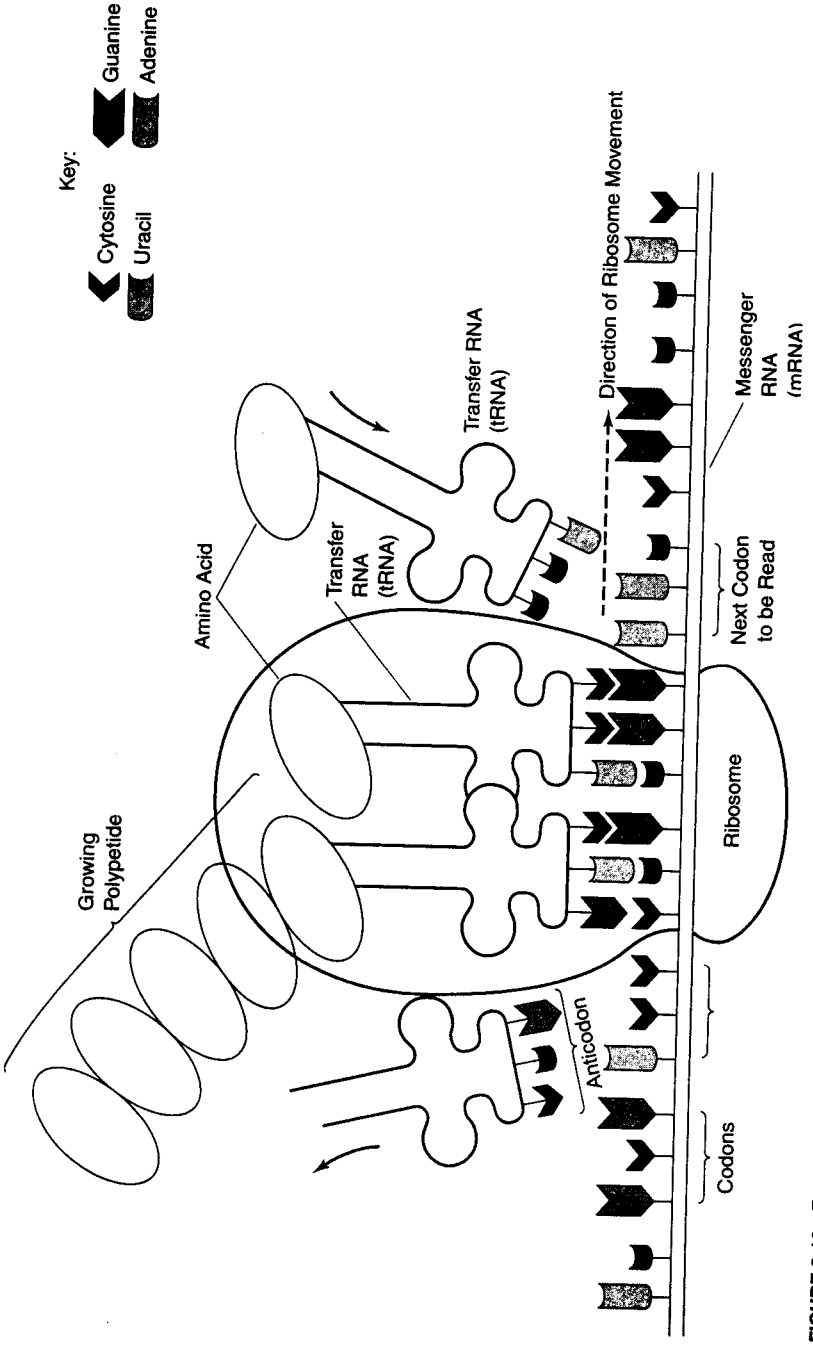


FIGURE 2-10 Formation of Polypeptide Chains.

This drawing illustrates the roles played by messenger and transfer RNA in directing the orderly arrangements of amino acids to form polypeptide chains that make up proteins. The *ribosomes*, small granular bodies in the cell's cytoplasm, move along the mRNA and bring the anticodon of the tRNA into contact with the complementary codon of the mRNA. The amino acids carried by the tRNAs form peptide bonds linking together in a chain until the protein is produced. As the process proceeds, the tRNA is released and free to attach to another amino acid.

amino acids in Table 2-1. This code is universal; that is, the amino acids have the same codons throughout all living organisms, whether amoeba or human, and the universality of the code provides strong supporting evidence for the unity of life.

To summarize, a single-stranded nucleic acid (RNA) is transcribed from one strand of the DNA molecule, the sense or template strand. This mRNA is complementary to all those triplet regions between a start and stop triplet sequence and includes the coding (exon) and noncoding (intron) regions of a gene complex. The introns are excised from the primary mRNA and the exons are joined before it is released to the site of protein synthesis, the ribosomes. The mRNA begins to attract a tRNA that carries one amino acid specified by the codon–anticodon sequence. A series of these are attached along the strand and the adjacent amino acids react to form peptide bonds resulting in a long chain that is the protein. From the diagrams and descriptions it is apparent that the mRNA faithfully replicates from the DNA, a complementary chain of triplets that must be modified by removal of those noncoding portions, the introns, since the DNA carries many more triplet units than are used in the protein synthesis process. The discovery of this excess of nucleotide material leads to a reexamination of chromosome structure and the basic nature of the human genome.

CHROMOSOMES AT THE MOLECULAR LEVEL

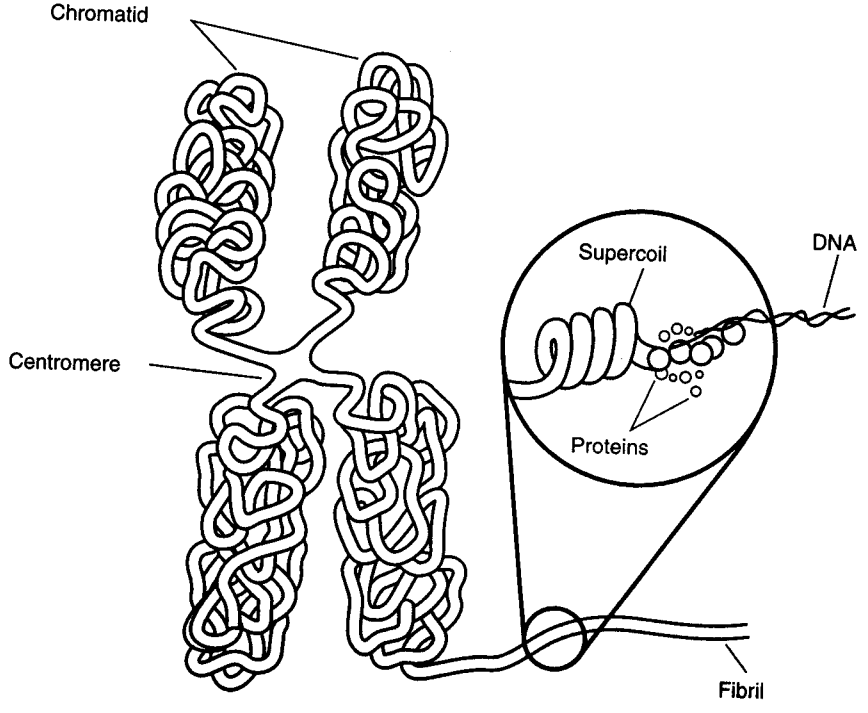
Since the description of chromosomes, their structure, and their changes during cell division was made early in this century, knowledge has reached a point where the finer, molecular components can be defined. We now not only have a sharper focus of what is transpiring at the genetic level but we also have a fairly good idea of how all the genes fit together to form these large structures in the nucleus, the chromosomes.

The human genome contains enough DNA for over one million genes, but the real number of genes is closer to one hundred thousand or even less. This means that a large excess of DNA is replicated and transmitted between generations of cells but is not used in the coding of protein synthesis. This excess, or actually most of the DNA, has other than genetic functions, since DNA fragments may be regulatory, provide start or stop signals, or function simply as spacing devices as in the case of the introns. The scope of this excess may be appreciated by some comparisons. If stretched out to full length, the DNA equivalent in the total haploid genome of a gamete is a molecule one meter long that contains about one billion triplets. But all of the DNA must be contained in a cell nucleus with the dimensions of 10 μm by less than 1 μm. This placement of such a large mass in a restricted space is made possible by compaction due to "super coiling" of the helical structures.

The total DNA of the human genome is divided into chromosomes, and each is a long, continuous chain of DNA coiled and compacted; the longest is 82 μm and the shortest, the Y chromosome, is 2.15 μm. At the beginning stages of mitosis or meiosis (discussed above) a chromosome is composed of two chromatids held together by a centromere somewhere along its length. A visualization of this arrangement is provided by the drawing in Figure 2-11, which compares components of a chromosome at the structural, microstructural, and submicroscopic level. This knowledge of the fine structure of the chromosome, together with knowledge of the nucleotides and triplets of the gene, has contributed to a clearer understanding of genetic loci and has permitted the mapping of portions of the genome. In fact, a major project (the Human Genome Project or HUGO) is now under way to map or identify and label the position of each gene site.

FIGURE 2-11 DNA and Chromosome Structure.

This diagram illustrates, at several microstructural levels, the relationship of DNA strands to the chromosome. The *chromatids* are shown as densely compacted DNA linked together at a point near their centers, the *centromere*. The lower right chromatid shows a section of a fibril that is further magnified to illustrate the supercoiled DNA structures it contains. Further expansion of this segment reveals the particles of proteins and an individual DNA molecule. (Source: Modified and redrawn from Vogel and Motulsky, 1986.)



Regions along the DNA strand of a chromosome have certain noncoding functions or contain codes for particular gene products. The region or locus for certain genetic traits, identified earlier by linkage studies in family lineages described above, has been confirmed and is now clearly established by investigations of chromosome structure at the molecular level. This advance in our knowledge of inheritance has been made possible by the discovery of a class of bacterial enzymes that can cut the giant DNA molecule into shorter strands at specific base pair locations. These enzymes, *endonucleases*, have opened up a whole new frontier for genetic studies.

Endonucleases. These restriction enzymes, so called because they cut the DNA molecule at specific points along the chain, have enabled investigators to separate much of the entire one-meter length of the human genome DNA into many shorter fragments. These fragments are of varying lengths (number of base pairs or Bp) and each has a specific triplet sequence at the point of cleavage; the triplet sequence plus one or two other nucleotides provide a point of attraction for one of these enzymes. Cleaving the entire human genome can produce about 500,000 fragments of from 100 to 10,000 Bp in length. These fragments can be separated by the *Southern blotting* method, a technique that separates the different fragments in an electric field and preserves them on a nitrocellulose filter (Southern, 1975). Basically, the technique takes advantage of the fact that fragments differ not only in size but in electric charge as well, a characteristic that will cause the fragments to move at different rates of speed through an agarose gel plate when an electric current is applied (electrophoresis). Once the fragments are separated, the fluid medium, a buffer solution, is blotted out by squeezing the gel between a weight, blotting paper, and the filter (Figure 2-12). This removes the fluid and leaves the fragments trapped and dispersed on the nitrocellulose filter as they were in the gel.

The hydrogen bonds of these DNA fragments are broken by treatment with an alkali solution separating the double strands. The now single-stranded fragments with specific base sequences can be examined individually by special chemical staining or by use of a radioactive probe. This probe is a short strand of DNA of known base sequence labeled with a radioactive isotope of phosphorus (^{32}P). The DNA probe combines with the complementary DNA strands and forms a double helical structure on the filter. A piece of photographic film is exposed to the radioactive labels over a period of several days. The result is a series of dark bands on the film that identifies the positions of those fragments that had combined with the probes. Because 500,000 fragments are an impossible number to deal with at one time, the chromosomes are first separated and then examined individually. The chromosome DNA may be cut by one or more restriction enzymes[4]

[4]There are hundreds of restriction enzymes (endonucleases) known at this time, with more being added to the list as the HUGO project proceeds.

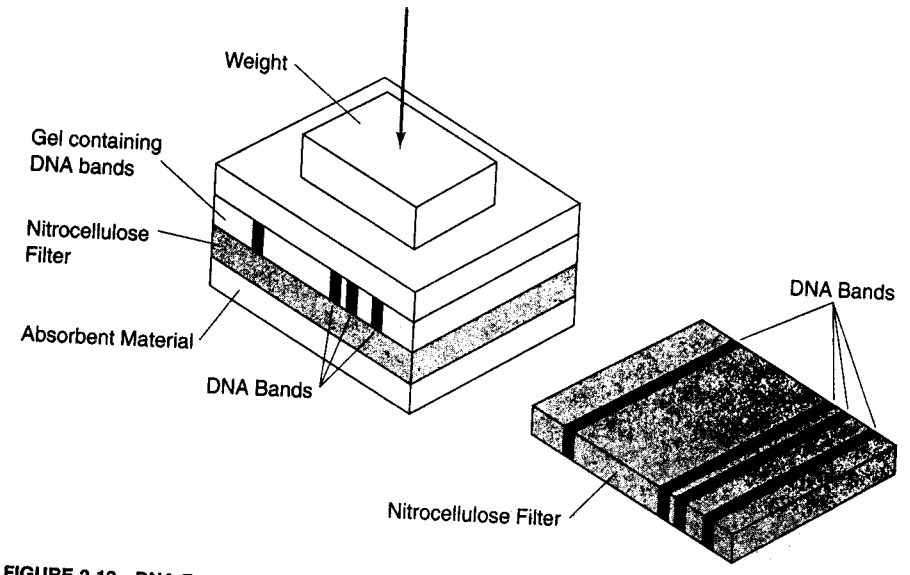


FIGURE 2-12 DNA Fragment Separation by the Southern Blotting Technique.

The weight in the drawing compresses the gel layer and squeezes the buffer fluid onto a nitrocellulose filter. The porous filter allows the fluid to pass through to the absorbent material while retaining the DNA fragments. These fragments are positioned as bands relative to one another as they were on the gel.

and the resulting fragments are then treated by the Southern blotting method.

DNA Probes. A large and ever-growing "library" of probes has made it possible to pinpoint the location of a gene on a particular chromosome. These DNA probes are radioactively labeled single-stranded DNA, which attach to the complementary fragments forming a hybrid double helix on the filter as described above. This allows for identification of specific genes or gene clusters because of the way in which the DNA probes are produced, by a method called reverse transcription. Where mRNA is available for proteins of known amino acid sequence, as in the case of the hemoglobins, insulin, etc. (see Table 2-2), an enzyme, reverse transcriptase, assembles DNA nucleotides complementary to the mRNA chain. This process is, as the name signifies, the reverse of the transcription process that occurs in the cell nucleus to produce the mRNA. The difference is, of course, that the molecular geneticist assembles complementary DNA (cDNA) in the laboratory using mRNA material as a template. The cDNA, when treated, becomes a "probe" to hybridize with the DNA fragments produced upon cleavage of the chromosome DNA by restriction enzymes.

Briefly, a gene product of interest is identified (enzyme, structural protein, etc.), and the mRNA is obtained from a cell or assembled in a test tube,

TABLE 2-2 Examples of Genetic Diseases Detected by Gene Probes

Disease
Achondroplasia
Alpha1-antitrypsin deficiency
Diabetes mellitus (type 1)
Globin gene cluster (alpha)
Globin gene cluster (beta)
Growth hormone deficiency
Hemophilia A
Hemophilia B
HLA genes
Immunoglobulin genes
Lesch-Nyhan syndrome
Osteogenesis imperfecta (type II)
Phenylketonuria
Prealbumin (amyloidosis)
Sickle cell anemia
Thalassemias
Thrombosis III deficiency

Selected from various sources. For a complete listing, see Cooper and Schmidtke (1986).

which is possible when the amino acid sequence of the gene product is known. The mRNA provides a template for the synthesis of cDNA. The cDNA is used as a probe to locate the gene sequence on a chromosome fragment. A similar method, in situ hybridization, also uses a radioactive DNA strand, but adds the probe to chromosomes in their metaphase of division. The probe hybridizes with the intact chromosome DNA at a specific segment. These methods have permitted many highly imaginative genetic studies and, as investigation of human chromosomes expands at an accelerating rate, the number of useful probes and their applications increase rapidly. Some examples of the locations of genes for specific traits are given in Table 2-3. The location is listed by chromosome number, the area (*p* or *q* arms) and the site on the arm. For example, a collagen gene is located on the long arm (*q*) of chromosome 7 at the 22nd site. Development of these processes has taken place over years of intensive biochemical work, and many are highly sophisticated, so much so that no brief description can do them justice. For more details and elaboration of additional investigation to identify or "map" the human genome, the reader is directed to texts like Hartl (1988), Nichols, (1988), or Vogel and Motulsky (1986).