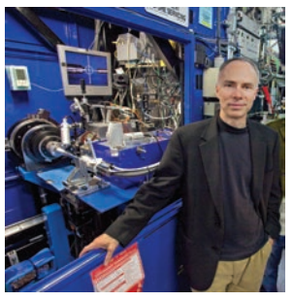
# 2 DNA: The Repository of Biological Information



James Berger
*[Source: Courtesy James Berger.]*

## MOMENT OF DISCOVERY

The first time I had an "Aha!" moment in science was when I was a graduate student. The question that intrigued me was related to the mechanism proposed for topoisomerases, which are essential enzymes that coil or uncoil DNA during DNA synthesis in all cells. Topoisomerase II–type enzymes (called Topo II) pass DNA strands through each other by cutting and rejoining DNA without marking or changing the genome in any way. In textbooks, the enzyme was shown as a sphere that bound to one segment of DNA, cut it, and then split in half to pass a second DNA segment through the split. But what held the DNA ends together during the passage of the DNA duplex through the double-stranded break? There had to be something else going on.

Francis Crick once said that you can't understand how an enzyme works unless you see its structure, and *I wanted to see the structure of Topo II.* I spent the next couple of years trying to crystallize the enzyme with no success, and eventually reached the point where I wondered if my project would ever work, and whether I had what it took be a scientist. I made one last preparation of the enzyme, and after working overnight in the lab, I put the purified enzyme on ice and went home to bed. When I came back the next day, the protein in the tube had turned white, and I was crushed, thinking it had precipitated into a useless aggregate. But when I looked at a sample under the microscope, I saw crystals growing in the tube! At that moment I knew I had a project. I spent the next nine months solving the molecular structure of the enzyme, and I'll never forget the thrill of seeing the structure for the first time.

It was instantly clear how Topo II must work. The enzyme has two jaws, one of which grabs and cleaves the DNA duplex and holds it while the other jaw passes a different segment of DNA through the gap. I experienced the intense joy of discovering this fundamental mechanism of DNA metabolism, and of knowing that at that moment I was the first person in the world to have this understanding of the natural world.

*—James Berger, on his discovery of the structure and mechanism of topoisomerase II*

**Genetics** is the science of heredity and the variation of inherited characteristics. Today, we know that biological information is stored and transmitted from generation to generation by deoxyribonucleic acid, or DNA, but this understanding arose only gradually. DNA was not widely accepted as the chemical of heredity until the 1940s, and its structure was not determined until 1953, when James Watson and Francis Crick introduced the world to the DNA double helix. (The structure of DNA is described in Chapter 6.) Our knowledge of the beautiful double-helical DNA structure has transformed the way that science is performed, to the extent that it is tempting to think of the field of genetics in terms of before and after DNA structure. But genetics has a wonderfully rich and varied history, every bit as exciting in the decades before the double helix as afterward.

The beginnings of modern genetics can be traced to the 1850s, when Gregor Mendel studied the inheritance of traits in the garden pea. He deduced that organisms contain particles of heredity (what we now call genes) that exist in pairs and that the paired particles split up when **gamete cells** (sex cells, the ovum and pollen in peas) are formed; pairs of hereditary particles are reformed on the union of two gametes during fertilization. Mendel was absolutely correct, but decades ahead of his time. His marvelous work went unnoticed for more than 30 years, until well after his death.

In contrast, a contemporary of Mendel's, Charles Darwin, was exceedingly famous in his lifetime. Darwin's theory of evolution started an awakening, one that continues to this day (see Chapter 1). For evolutionary theory to work, there must be diversity among individuals within a species, and variants more suited to the environment are selected and survive to produce offspring. Darwin's evolutionary theory, as wonderful as it is, completely lacks an explanation for how this diversity is produced. In fact, Darwin spent considerable time pondering this problem. He espoused the theory of *pangenesis*, first proposed by the ancient Greeks, in which genetic traits are shaped by life experience and transferred by "pangenes" to gamete cells, via the blood, enabling the traits to be inherited. In principle, the mistaken pangenesis theory is a variation of Jean-Baptiste Lamarck's theory of inheritance of acquired characteristics (see the How We Know section at the end of Chapter 1).

Darwin's theory of evolution became widely known, but Mendel's work fell into obscurity. During the late 1800s, advances in microscopy pushed the optical limits, enabling scientists to visualize subcellular structures. Of particular interest to geneticists were chromosomes, structures found in the nuclei of cells. A rash of intense studies documented chromosome behavior during cell division, fertilization, and the formation of gamete cells. New discoveries revealed that the number of chromosomes in **somatic cells** (all cells in a multicellular organism other than sex cells) is constant for a given species and that the total number of chromosomes is halved to form gametes. When Mendel's work was rediscovered in 1900, his principles of heredity and particles of inheritance fit nicely with the behavior of chromosomes observed under the microscope.

Proof that genes reside on chromosomes soon followed, from a series of wonderful studies on fruit flies started in 1908 by Thomas Hunt Morgan. Central to Morgan's work were mutants, flies displaying physical traits not found in the average fly. The variety of mutant flies accumulated by Morgan's lab during 15 years of study—generations of flies reared in milk bottles—was amazing, including flies with bodies of different shapes and sizes, a variety of wing patterns, legs of different sizes, and a whole spectrum of eye colors. These fly mutants simply appeared spontaneously over generations of growth in Morgan's lab. Here was the answer to the variation required to make Darwin's theory of evolution work. Spontaneous mutants are infrequent, but given the expanse of evolutionary time, sufficient numbers and types of mutants are produced for nature to select and mold new species.

Genes and mutations explain heredity and illuminate evolutionary theory. But what are genes made of, and how is the information within them translated into the physical traits of an organism? Chromosomes were known to consist of both DNA and protein—but which of these is the genetic material? Several elegant and now classic experiments identified DNA as the molecule of heredity and found that DNA contains a code to direct the synthesis of RNA and proteins. The structure of the DNA double helix intuited by Watson and Crick revealed an architecture more beautiful than anyone could have imagined. The DNA molecule consists of two long strands twisted about each other, each chain a series of repeating units called nucleotides.

Watson and Crick immediately realized that the cell must have a mechanism to untwist the two strands in order to duplicate the DNA molecule and pass the genetic information to the next generation. Indeed, as we shall see in Chapter 9, the cell contains a complex arsenal of enzymes devoted to untwisting and altering the topology of DNA. These enzymes, called topoisomerases, are important targets of anticancer drugs, and their mechanisms of action are still actively investigated by James Berger and many other researchers. By extension, it is also critical to have a thorough understanding of the DNA molecule itself, and that is the subject of this and several other chapters in this textbook.

Despite the intrinsic beauty of the newly discovered double helix, the process by which a sequence of nucleotides could code for a sequence of amino acids

in a protein remained a mystery. In a rapid series of advances in the 10 years following Watson and Crick's breakthrough, mRNA, tRNA, and rRNA were discovered and the workings of the directional flow of biological information, DNA→RNA→protein, were understood.

The field of molecular biology developed from these great discoveries. It seeks a detailed explanation of how biological information brings order to living processes. Molecular biology—the subject of this book—is a rapidly evolving scientific pursuit. The fundamental discoveries described in this chapter provided the groundwork for all subsequent studies in the field.

## 2.1 MENDELIAN GENETICS

Gregor Mendel was a monk at the monastery of St. Thomas at Brünn (now Brno in the Czech Republic). Renowned for teaching the sciences, the monastery sent Mendel to the University of Vienna in 1851, to obtain his teaching credentials. Although Mendel failed his final exams, he returned to the monastery and began a 10-year program of experiments that were so well conceived and executed that his results form the cornerstone of modern genetics.

In Mendel's time, plant hybrids were highly desired for their unique ornamental varieties. But the inheritance of colorful hybrid flower patterns was perplexing and unpredictable. The inability to decipher general principles of inheritance was not for lack of trying. Many well-known scientists performed extensive plant-breeding experiments, but no fundamental principles of inheritance could be formulated from these endeavors.

Gregor Mendel, 1822–1884
*[Source: Pictorial Press Ltd/Alamy.]*

Mendel's success where others failed can be attributed to his sound scientific approach to the problem. For his studies Mendel picked the garden pea, *Pisum sativum*, an excellent choice for several reasons. Because the pea was economically important, many varieties were available from seed merchants. The pea plant is also small, so that many plants can be grown in a confined space; and it grows quickly, reaching maturity in one growing season. But perhaps more important than Mendel's choice of experimental organism was his approach to studying it. Others before him, in studying plant breeding, had looked at the plant as a whole, dooming a study of heredity from the start because many genes control the overall appearance of an organism. Mendel focused instead on separate features of the plant, carefully observing isolated characteristics of the seeds, flowers, stem, and seed pods (**Figure 2-1**).

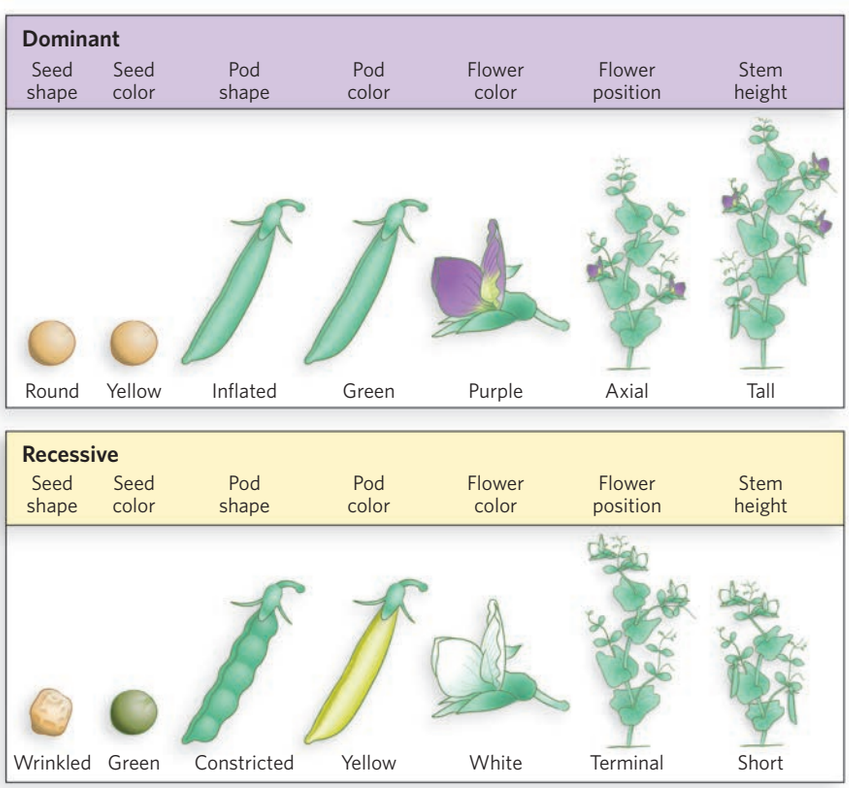



**FIGURE 2-1 Traits of the garden pea examined by Mendel.** Mendel picked seven pairs of traits to study: seed shape, seed color, seed pod shape, seed pod color, flower color, flower position along the stem, and stem height. The dominant and recessive forms of the traits are shown.

## Mendel's First Law: Allele Pairs Segregate during Gamete Formation

Mendel spent the first two years growing different varieties of peas to ensure that each was a true-breeding, or **purebred**, with the offspring produced by crossing two plants of the same variety always having the same appearance as the parents. Then he carefully selected seven different pairs of traits and cross-pollinated plants with contrasting traits. For example, plants with round seeds were crossed with plants having wrinkled seeds. The parental plants are referred to as the **P generation** (*P* for parental). The **hybrid** offspring are called the **$F_1$ generation** (*F* for filial, from the Latin for "son or daughter"; $F_1$ indicating first filial). This first generation produced only round seeds; the wrinkled-seed trait seemed to have disappeared. Mendel observed a similar result in crosses for all seven pairs of traits (**Table 2-1**). He referred to the trait that appears in the $F_1$ generation as the **dominant** trait, and the trait that disappears as the **recessive** trait (see Figure 2-1).

Mendel's finding that one trait is dominant and the other is recessive was novel and completely contrary to the prevailing view that parental traits blend together in the offspring. Other experimenters might have stopped at this new and dramatic discovery, but not Mendel. In his next experiment, he allowed $F_1$ plants to self-pollinate and produce the **$F_2$ generation** (second filial generation). Surprisingly, the $F_2$ generation was a mixture: most plants produced round seeds (the dominant trait), but some had wrinkled seeds. The recessive wrinkled-seed trait that disappeared in the $F_1$ generation reappeared in the $F_2$ generation! In contrast to the view that parental traits blend in the offspring, Mendel did not observe a blending of the traits he studied. For example, there were no partly wrinkled seeds, and seed colors were either yellow or green, not yellow-green.

Unlike other scientists before him, Mendel kept close track of the numbers of offspring with the dominant and recessive traits. He counted 5,474 dominant round seeds and 1,850 recessive wrinkled seeds in the $F_2$ generation, for a ratio of 2.96 dominant to 1 recessive trait. His experiments examining seed color produced a similar result. He observed 6,022 dominant yellow seeds and 2,001 recessive green seeds, for a ratio of 3.01:1. The other pairs of traits also appeared in a 3:1 ratio of dominant to recessive offspring in the $F_2$ generation, as summarized in Table 2-1.

Mendel's interpretation of these results was brilliant. Reappearance of the recessive wrinkled-seed trait in the $F_2$ generation suggested that traits do not really disappear. Mendel therefore proposed that traits are "hereditary particles"—now called **genes**—and that they come in pairs. Organisms that carry two copies of each gene are **diploid**, and the different variants of a given gene are called **alleles**. In other words, a diploid parental plant has two alleles of the seed-shape gene (for example, one for smooth seeds and one for wrinkled seeds). Furthermore, Mendel proposed that one allele could mask the appearance of the other. This explained why traits could disappear but then reappear in future generations. Even though the $F_1$ plant carried one dominant allele (round seed) and one recessive allele (wrinkled seed), only the dominant round-seed allele is evident in the outward appearance, or **phenotype**, of the $F_1$ plant. Mendel also reasoned that each parent contributes only one copy of each gene to the offspring; that is, the gamete cells are **haploid**, having only one allele of each gene. When two $F_1$ generation gametes combine and each carries the recessive allele for seed shape, the resulting $F_2$ plant will produce wrinkled seeds.

We have introduced several genetic terms in the preceding paragraphs. These terms, and others that follow, are part of the scientific language of genetics and are defined in **Table 2-2**.

Mendel was well-trained in mathematics, which helped him make sense of his results. To explain the ratios of phenotypes in the offspring in mathematical terms, he referred to the dominant allele with a capital

**TABLE 2-1**
**Results of Mendel's Single-Factor Crosses**

| *Characteristic Isolated for Study* | *Parental Cross* | *$F_1$ Phenotype (dominant trait)* | *$F_2$ Phenotypes (dominant and recessive)* | *$F_2$ Ratio (dominant:recessive)* |
|---|---|---|---|---|
| Seed shape | Round × wrinkled | All round | 5,474 round and 1,850 wrinkled | 2.96:1 |
| Seed color | Yellow × green | All yellow | 6,022 yellow and 2,001 green | 3.01:1 |
| Pod shape | Inflated × constricted | All inflated | 882 inflated and 299 constricted | 2.95:1 |
| Pod color | Green × yellow | All green | 428 green and 152 yellow | 2.82:1 |
| Flower color | Purple × white | All purple | 705 purple and 224 white | 3.15:1 |
| Flower position | Axial × terminal | All axial | 651 axial and 207 terminal | 3.14:1 |
| Stem height | Tall × short | All tall | 787 tall and 277 short | 2.84:1 |

**TABLE 2-2**

**Commonly Used Terms in Genetics**

| *Term* | *Definition* |
|---|---|
| P generation | Parents used in a cross |
| $F_1$ generation | Progeny resulting from a cross in the P generation |
| $F_2$ generation | Progeny resulting from a cross in the $F_1$ generation (succeeding generations are $F_3$, $F_4$, etc.) |
| Purebred | Individual homozygous for a given trait or set of traits |
| Hybrid | Progeny resulting from a cross of parents with different genotypes |
| Gene | Section of DNA encoding a protein or functional RNA |
| Allele | Variant of the gene encoding a trait (e.g., seed color: yellow or green) |
| Phenotype | Outward appearance of an organism |
| Genotype | Alleles contained in an organism |
| Homozygous | Having two identical copies of an allele for a gene |
| Heterozygous | Having two different alleles for a gene |
| Dominant allele | Allele expressed in the phenotype of a heterozygous organism |
| Recessive allele | Allele masked in the phenotype of a heterozygous organism |

letter (e.g., *R* for round) and to the recessive allele with the lowercase version of the same letter (*r* for wrinkled). A lowercase *w* might seem more fitting for the wrinkled-seed allele, but using different letters would make it harder to keep track of allele pairs of the same gene. A purebred round-seed plant has two *R* alleles, *RR*, and a purebred wrinkled-seed plant has two recessive alleles, *rr*. *RR* plants exhibit the dominant *R* trait (round seeds), and *rr* plants exhibit the recessive *r* trait (wrinkled seeds). This double-letter nomenclature, representing the allelic makeup of an organism, is a way of denoting the organism's **genotype**.

In a cross of purebred parents, round (*RR*) and wrinkled (*rr*), all $F_1$ progeny receive one allele from each parent and are thus *Rr*. The *R* allele is dominant to the *r* allele, so all $F_1$ *Rr* hybrid plants have round seeds. The $F_1$ plant produces *R* and *r* gametes in equal amounts, and therefore self-pollination of $F_1$ plants produces three different diploid genotypes: *RR*, *Rr*, and *rr*. A convenient way of displaying the genes that come together during a cross such as this is a **Punnett square** analysis (**Figure 2-2**). In this analysis, the gamete genotypes of one parent are written along the top of the Punnett square, and those of the other parent are written along the left side of the square. The various combinations in which the alleles can come together during pollination are entered in the grid. The results yield the three different $F_2$ genotypes in the following ratios: 1 *RR*, 2 *Rr*, and 1 *rr*. These genotypes, together with the concept of dominant and recessive alleles, explain the ratio of phenotypes that Mendel observed: 3 dominant (1 *RR* + 2 *Rr*) to 1 recessive (1 *rr*). We now refer to an organism with identical alleles for a given gene as **homozygous** for that gene (such as *RR* or *rr*). An organism with two different alleles, such as an *Rr* plant, is characterized as **heterozygous**.

To determine whether the $F_2$ plants really were of three genotypes in a 1:2:1 ratio, Mendel analyzed the offspring of self-fertilized $F_2$ plants (the $F_3$ generation). The $F_2$ recessive wrinkled-seed plants (25% of the total $F_2$ plants) all bred true, giving only wrinkled-seed $F_3$ progeny, and thus were homozygous *rr*. The dominant round-seed $F_2$ plants were of two types. One-third (25% of the total) bred true; their offspring always produced round seeds, and therefore these $F_2$ plants were homozygous *RR*. The remaining two-thirds of the round-seed $F_2$ plants (50% of the total) produced $F_3$ plants with round and wrinkled seeds in a 3:1 ratio, and therefore these $F_2$ plants were heterozygous *Rr*. The analysis fit the 1:2:1 ratio for the $F_2$ genotype exactly: 1*RR*:2*Rr*:1*rr* (see Figure 2-2).

In summary, Mendel hypothesized that traits are carried by particulate genes, that somatic cells contain two copies (two alleles) of each gene, and that gamete cells obtain only one allele for each gene during gamete formation. When two gametes fuse at fertilization, allele pairs are restored, producing the diploid genotype of the offspring. The general principle summarizing this proposal is often referred to as Mendel's first law, or the **law of segregation**, which states that equal and independent segregation of alleles occurs during formation of gamete cells.

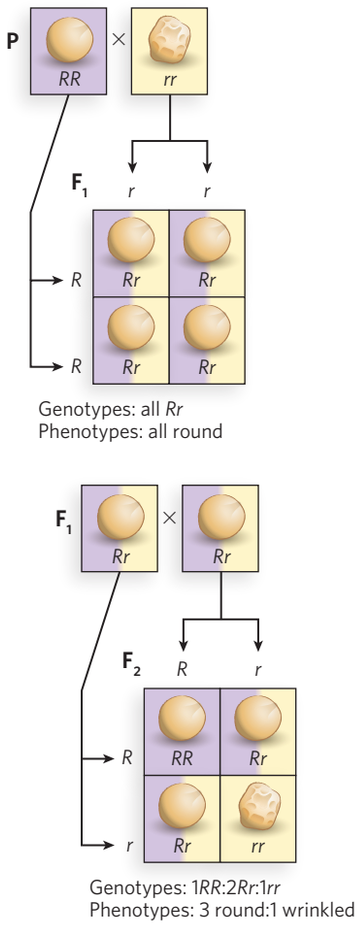



**FIGURE 2-2 An example of Mendel's first law.** Alleles of the same gene segregate independently into gametes. Plants that are homozygous for dominant round-seed shape (*RR*) were cross-pollinated with homozygous recessive wrinkled-seed plants (*rr*) to produce $F_1$ progeny. The Punnett square analysis shows the gametes of each parent along the top and left side of the grid, from which—if the gametes are formed in equal amounts—one can predict the possible progeny and their frequency. Punnett square analysis is a common way of illustrating genetic crosses today, but was developed after Mendel's work. In the $F_1$ generation, the only progeny that can be produced are *Rr* hybrids. $F_1$ plants were self-fertilized, and as the Punnett square analysis for the $F_2$ generation predicts (based on the assumption that the different alleles (*R* and *r*) segregate independently into $F_1$ generation gametes), round seeds and wrinkled seeds were produced in a 3:1 ratio.

## Mendel's Second Law: Different Genes Assort Independently during Gamete Formation

Mendel's results demonstrated that two alleles for one gene separate during gamete formation, but how do alleles for two *different* genes behave? There are two possibilities. Alleles for two different genes could separate during gamete formation, assorting randomly into the gametes. Alternatively, they could remain associated, traveling together into the same gamete cells. These two scenarios have distinct outcomes. For example, if particular alleles for seed shape and seed color stay together during the formation of gamete cells, future offspring will retain both the same seed shape and the same seed color as one parent or the other. But if alleles for the two genes separate during gamete formation, some of the $F_2$ offspring will exhibit new combinations of seed shape and color, distinct from those of either parent.

To test these hypotheses, Mendel's next experiments were two-factor crosses, analyzing the transmission of two different genes in each cross. He began by cross-pollinating purebred plants having round, yellow seeds with plants having wrinkled, green seeds. First let's consider the genotype of the two plants. We already know that the round-seed allele (*R*) is dominant to the wrinkled-seed allele (*r*). The genotype of the purebred plant with dominant yellow seeds is *YY*, and the purebred plant with green seeds is homozygous for the recessive green-seed allele, *yy*. Thus, the genotype of a purebred round, yellow-seed plant is *RRYY*, and the genotype of a plant with wrinkled, green seeds is *rryy*. A cross between these plants yields $F_1$ progeny of genotype *RrYy*, phenotypically round and yellow. If the four alleles for the two genes separate and assort randomly during gamete formation, $F_1$ plants will produce four different gametes (*Ry*, *rY*, *RY*, and *ry*), and all combinations of seed shape and color will be observed in $F_2$ plants (**Figure 2-3**).

Mendel's observations of $F_2$ phenotypes are shown in **Table 2-3**. All possible combinations of traits occurred, and therefore the alleles of the two different genes are not physically connected; instead, they separate during the formation of gamete cells. The analysis of the four possible genotypes, as shown in Figure 2-3, predicts a 9:3:3:1 ratio, close to the observed result. Mendel performed many two-factor crosses analyzing different gene combinations, and the results were always consistent with the random assortment of genes during gamete formation. Mendel's second law, or the **law of independent assortment**, states that different genes assort into gametes independently of one another.

Mendel studied garden peas, but his basic principles hold true for sex-based inheritance in all animals and plants. Indeed, many human genetic diseases that can be traced through a family pedigree follow Mendel's simple rules of inheritance.

## There Are Exceptions to Mendel's Laws

The transmission of dominant and recessive traits documented by Mendel is sometimes referred to as "Mendelian behavior." However, not all genes behave in such an ideal fashion. There are many exceptions to Mendel's principles of heredity, a few of which we review here.

***Incomplete Dominance*** Some alleles of a gene are neither dominant nor recessive. Instead, hybrid progeny display

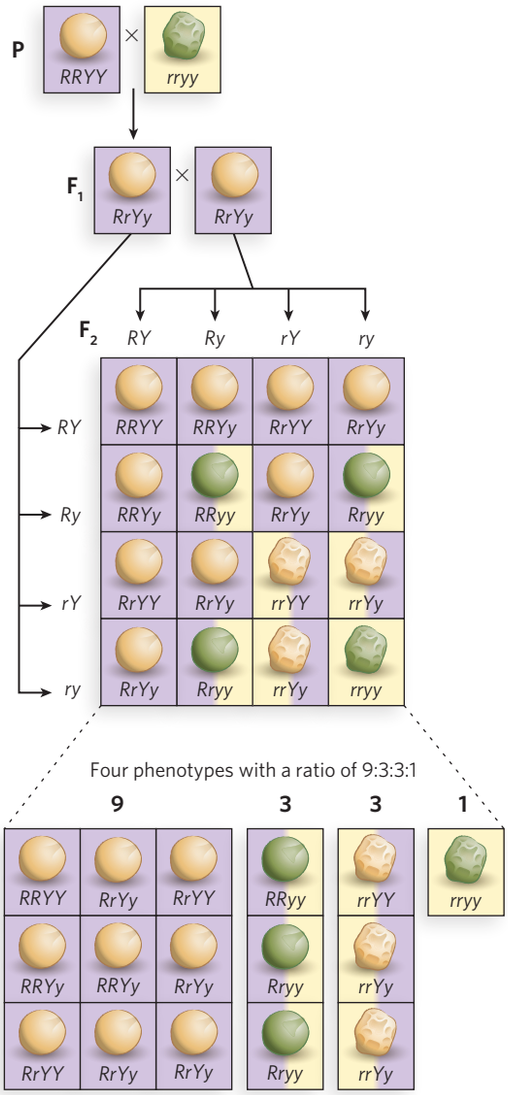



**FIGURE 2-3 An example of Mendel's second law.** Different genes assort independently into gamete cells. The parental cross (round, yellow seeds × wrinkled, green seeds) yields uniform $F_1$ progeny with the dominant phenotype (round, yellow seeds) and genotype *RrYy*. The Punnett square analysis assumes random assortment of the different alleles into the gametes formed by the $F_1$ plant. All possible gamete genotypes are written across the top and left side of the grid. The predicted outcome for independent assortment is seeds of four different phenotypes in a 9:3:3:1 ratio, as illustrated below the Punnett square.

a phenotype intermediate between those of the two parents. This type of non-Mendelian behavior is called **incomplete dominance**. An example of incomplete dominance can be seen in the gene for flower color in four o'clock plants (**Figure 2-4**). Homozygotes are either red (*RR*) or white (*R′R′*; primed capitals are used so as not to confuse this case with recessive alleles), but the $F_1$ heterozygote (*RR′*) is neither red nor white; it is pink. The molecular explanation for the pink heterozygote is the production of sufficient red color from the single *R* allele in the heterozygote to yield a pink coloration.

Interestingly, an example of incomplete dominance can also be found in Mendel's published work. He studied different alleles for the gene controlling the pea plant's flowering time. The $F_1$ progeny had a flowering time that was intermediate between the flowering times of the two parents.

***Codominance*** Recessive alleles often produce nonfunctional proteins, or none at all. However, there are many examples of two alleles of a gene that produce two different functional proteins, neither of which is dominant to the other. This non-Mendelian behavior is known as **codominance**. An example of codominance is human blood type (**Figure 2-5**). The allele for A-type blood, $I^A$, results in a cell surface glycoprotein different from the glycoprotein encoded by the allele for B-type blood, $I^B$. People with A-type blood are homozygous $I^AI^A$, and those with B-type blood are homozygous $I^BI^B$. AB-type individuals are $I^AI^B$ heterozygotes. O-type individuals lack both varieties of surface glycoprotein; they are homozygous for the recessive *i* allele.

***Linked Genes*** The most common non-Mendelian behavior is seen in **linked genes**, in which alleles for two different genes assort together in the gametes, rather than assorting independently. We now know that genes are located on chromosomes, and diploid organisms have two copies of each chromosome, known as **homologous chromosomes**, or homologs. During gamete formation, whole chromosomes, not individual genes, assort into gametes. Genes that are close together on one chromosome are inherited together, contrary to Mendel's second law.

**TABLE 2-3**

**Mendel's Results from a Two-Factor Cross**

| *Characteristic Isolated for Study* | *Parental Cross* | *$F_1$ Phenotype* | *$F_2$ Phenotype* | *$F_2$ Ratio* |
|---|---|---|---|---|
| Seed shape and seed color | Round, yellow × wrinkled, green | All round, yellow | 315 round, yellow | 8.3 |
| | | | 101 wrinkled, yellow | 2.7 |
| | | | 108 round, green | 2.8 |
| | | | 38 wrinkled, green | 1.0 |

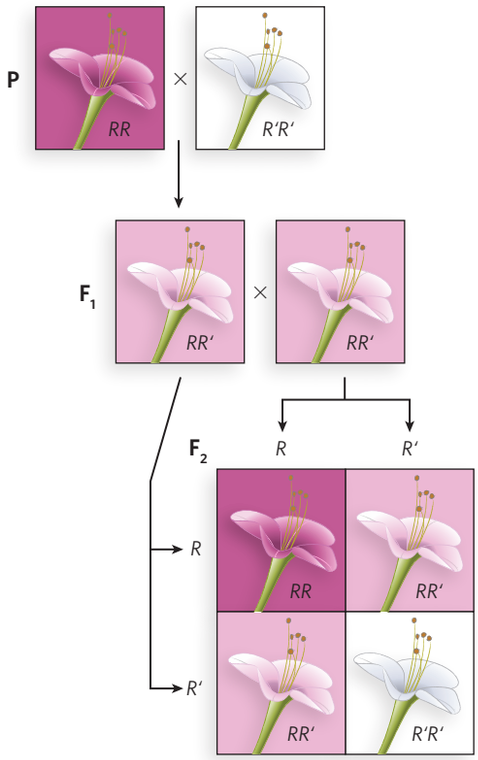



**FIGURE 2-4 Non-Mendelian behavior: incomplete dominance in four o'clock plants.** Cross-pollination of red- and white-flowered plants yields an $F_1$ plant with flowers of intermediate color (pink). Therefore, neither parental allele is completely dominant. The single *R* allele gives rise to sufficient red pigment to produce a pink coloration. Genotypes are given below each flower.

Assortment of linked genes into gamete cells is shown in **Figure 2-6** and is discussed in detail later in the chapter.

Mendel picked traits whose genes assorted independently. However, some of the genes that he studied are on the same chromosome. How could Mendel have observed independent assortment of genes on the same chromosome? As we describe later in the chapter, homologous chromosomes associate together during the cell divisions that lead to gametes. At that time, there is often an exchange of genetic material between the chromosome pair, resulting in some alleles previously found on one chromosome now being found on the other. For the traits Mendel selected for study, genes on the same chromosome are spaced far apart, and this swapping of genetic material occurs frequently between them. Therefore, the genes assort as though they are on different chromosomes.

Many other types of non-Mendelian behavior exist besides those described here. These include traits determined by multiple genes, traits derived from interactions between different genes (epistasis), the inheritance of traits encoded by organelle genes (cytoplasmic inheritance), and traits that depend on whether the gene is inherited from the male or female parent (genomic imprinting).

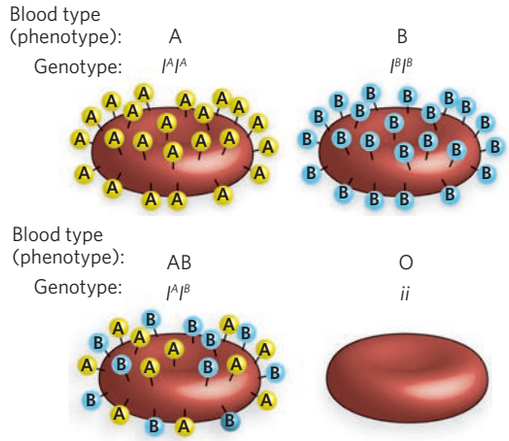



**FIGURE 2-5 Non-Mendelian behavior: codominance in human blood types.** The cell surface glycoprotein antigens on red blood cells (erythrocytes) determine human blood type. Two different alleles encode two variants of the enzyme glycosylase (an enzyme involved in formation of glycoproteins) and produce different cell surface glycoproteins, A (allele $I^A$, yellow circles) and B (allele $I^B$, blue circles). Both alleles are expressed in the heterozygote (AB blood type, genotype $I^AI^B$). Because both alleles produce functional surface glycoproteins, neither allele is dominant to the other. Individuals with O-type blood have two null alleles (*ii*) and thus produce no A or B surface antigens on their red blood cells.

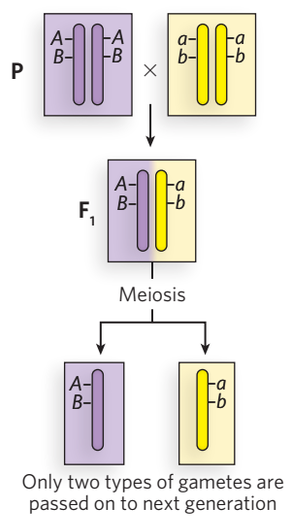



**FIGURE 2-6 Non-Mendelian behavior: linked genes.** Genes *A* and *B* are located on the same chromosome. Shown here is a cross between a homozygous dominant parent and a homozygous recessive parent that results in $F_1$ hybrid progeny, all *AaBb*. Because the alleles for the two genes are close together on the same chromosome (*A*, *B* and *a*, *b*), they cannot separate during the formation of gametes. Each gamete receives one copy of this chromosome, and thus either the *A* and *B* alleles or the *a* and *b* alleles; no gametes containing *A* and *b* or *a* and *B* are formed. The two genes *A* and *B* are linked and do not assort independently during meiosis in the formation of gamete cells.

**SECTION 2.1 SUMMARY**

- Mendel's studies on the garden pea revealed an underlying mathematical pattern in inheritance.
- Mendel postulated that genetic traits are carried by hereditary particles, now called genes. Diploid organisms contain two copies, or alleles, of each gene and produce haploid gametes that contain one allele for each gene.
- Individuals homozygous for a particular gene have two identical alleles for that gene. In heterozygous individuals, the two alleles for a gene are different. The allelic makeup of an organism is its genotype.
- The different alleles for a gene may be dominant or recessive. In a heterozygote, the dominant allele masks the recessive allele in the outward appearance, or phenotype, of the organism.
- Mendel's first law, the law of segregation, states that the two alleles for each gene segregate independently into haploid gamete cells.
- Mendel's second law, the law of independent assortment, states that alleles for different genes assort into gametes randomly. However, we now know that genes reside on chromosomes and that chromosomes, not genes, assort randomly into gametes.
- There are exceptions to Mendel's laws. For example, a gene exhibits incomplete dominance when the phenotype of heterozygous progeny is intermediate between those of the two homozygous parents. Gene alleles exhibit codominance when both produce functional proteins and neither is dominant to the other, as in human blood types. Alleles for two genes close together on the same chromosome are linked and do not assort independently into gametes.

## 2.2 CYTOGENETICS: CHROMOSOME MOVEMENTS DURING MITOSIS AND MEIOSIS

In 1865, Mendel presented his findings on inheritance in two lectures to the Brünn Society for the Study of the Natural Sciences, and they were then published in an obscure journal. Only about 150 copies of this journal were printed, and Mendel's findings lay dormant for decades, to be resurrected only after his death. However, in his lifetime, Mendel was well appreciated at his monastery, was elected abbot, and managed one of the wealthiest cloisters in the land. His claim to fame was an incident in which he refused to pay a new tax imposed on the monastery by the Habsburg Empire. Mendel met the sheriff at the gate and dared him to take the keys from his pocket before he'd pay another pfennig! Of course, this is not what we know Mendel for today.

The years between 1880 and 1900 saw amazing discoveries in **cytology**, the study of cells, which intersected with the rediscovery of Mendel's work. Microscopes had become more advanced, and chromosomes could be stained and visualized in the cell nucleus. Cytologists observed that, unlike other cellular components, chromosomes were meticulously divided between the two new cells during cell division. The diploid nature of somatic cells and the haploid nature of gametes were also discovered around this time. It was in this scientific environment of explosive growth in **cytogenetics** that, in 1900, Mendel's principles of heredity were confirmed experimentally and rediscovered independently by three scientists: Hugo de Vries, Carl Correns, and Erich von Tschermak. The behavior of chromosomes was seen to remarkably mirror the behavior of Mendel's hereditary particles, and the idea that the nucleus, and perhaps the chromosomes themselves, formed the basis of heredity was bandied about.

In this section, we describe the architecture of the cell and the chromosome movements that occur during somatic cell division and gamete formation, setting the stage for the chromosome theory of inheritance (the subject of Section 2.3).

### Cells Contain Chromosomes and Other Internal Structures

Robert Hooke was the first to notice the cellular composition of a biological specimen, during his microscopic examination of cork in 1665 (**Figure 2-7**). In his famous book *Micrographia*, Hooke described a multitude of tiny boxes in the cork sample and coined the word **cell** (Latin *cellula*, "small compartment"). By the early 1800s, it became clear that plants are made up of cells. In 1833, Robert Brown identified the nucleus, the first subcellular structure to be discovered. In 1839, Theodor Schwann realized that animal tissue contains nuclei throughout the cells, and he proposed the **cell theory**, which states that all animals and plants consist of large assemblages of cells.

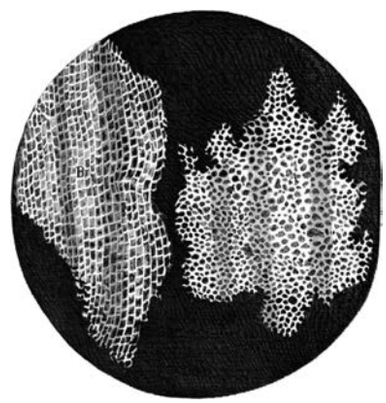


**FIGURE 2-7 Hooke's microscopic examination of cork.** Robert Hooke used a compound microscope to visualize cork cells and catalogued his work with meticulous drawings. This drawing is from his book *Micrographia*. His studies provided the first clue that organisms are cell-based. *[Source: U.S. National Library of Medicine.]*

Microscopic studies of chromosomes within nuclei were first made in plant cells by Karl Wilhelm von Nägeli and Wilhelm Hofmeister between 1842 and 1849. **Chromosomes** were named (from the Greek for "colored body") for their property of taking up large amounts of colored dye. The term was coined by Heinrich von Waldeyer in 1888. **Figure 2-8** shows rapidly dividing cells of an onion, stained to show the chromosomes. Development of the electron microscope in 1931 eventually brought into view the detailed structure of the cell as we know it today. Each cell is bounded by the **cytoplasmic membrane**, encasing the cytoplasm and its variety of subcellular structures called **organelles**. **Figure 2-9** is a schematic depiction of a typical animal cell, a eukaryotic cell; the caption describes each organelle and its function.

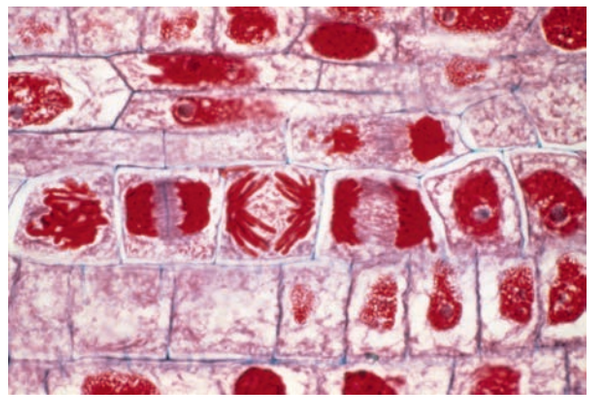


**FIGURE 2-8 Plant cell chromosomes.** A cross section of a rapidly dividing root tip of the onion (*Allium cepa*) shows the chromosomes as darkly stained bodies. Cells in different stages of division (mitosis) are apparent. *[Source: Manfred Kage/Science Source.]*

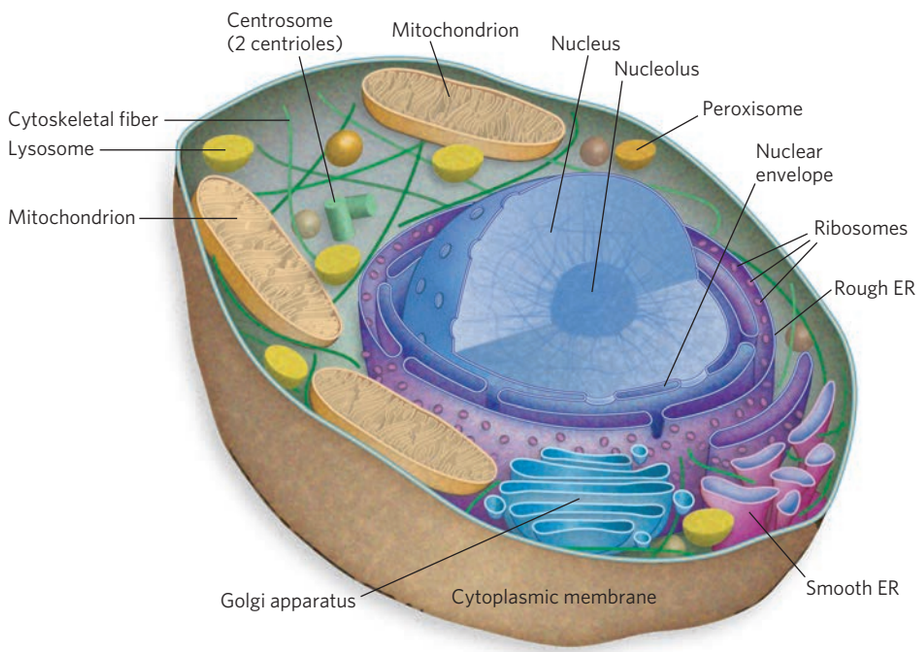



**FIGURE 2-9 Animal cell structure.** Eukaryotic cells are bounded by a cytoplasmic membrane. Chromosomes (not visible as individual structures in a nondividing cell) are located in the nucleus, and the nuclear envelope (nuclear membrane) is a double membrane with large pores through which RNA and proteins move. The nucleolus, a substructure within the nucleus, is the site of rRNA synthesis. All other organelles are in the cytoplasm. The centrosome consists of two centrioles, perpendicular cylinder-shaped protein complexes. Centrosomes organize microtubule spindles that attach to the cytoskeleton. During cell division, the centrosome duplicates and the two centrosomes migrate to opposite poles of the cell. The centrosomes then organize the spindle apparatus, in which microtubules connect chromosomes to the centrosomes for partitioning of the chromosomes into daughter cells. Mitochondria, the energy factories of animal cells, oxidize fuels to produce ATP. Lysosomes, containing degradative enzymes, aid in digestion of intracellular debris and recycle certain components. Peroxisomes help detoxify chemicals and degrade fatty acids. The smooth endoplasmic reticulum (ER) is the site of lipid synthesis and drug detoxification. Ribosomes, composed of both RNA and protein, act as protein-synthesizing factories; many attach to the ER, giving it a rough appearance. The rough ER sorts proteins destined for the cytoplasmic membrane or for other organelles; it is continuous with the outer membrane of the nuclear envelope. The Golgi apparatus, a membranous network, receives proteins from the ER and modifies and directs them to their proper compartments. Cytoskeletal fibers are a network of structural proteins that give shape to the cell and aid in cell movement.

## Mitosis: Cells Evenly Divide Chromosomes between New Cells

For biological information to be faithfully transmitted to daughter cells, it must be duplicated and then each complete information packet correctly partitioned into its own cell. As cells grow, they proceed through four phases, collectively known as the **cell cycle**: $G_1$, S, $G_2$, and M (**Figure 2-10**). In **$G_1$ phase**, cells are diploid, containing two copies of each chromosome. The cellular chromosome content in $G_1$ cells is represented as $2n$, where $n$ is the number of unique chromosomes of that species. $G_1$ is also called the first gap, because it represents a gap in time before S phase.

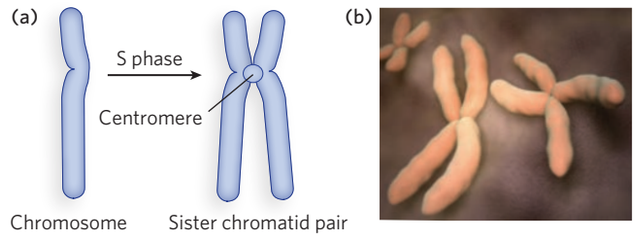



**FIGURE 2-11 Formation of sister chromatid pairs by chromosome duplication.** (a) Chromosomes are duplicated as cells proceed through S phase. Each resulting sister chromatid pair is held together at the centromere. (b) Three sister chromatid pairs, as seen by electron microscopy. *[Source: (b) MedicalRF/ The Medical File/ Peter Arnold Inc.]*

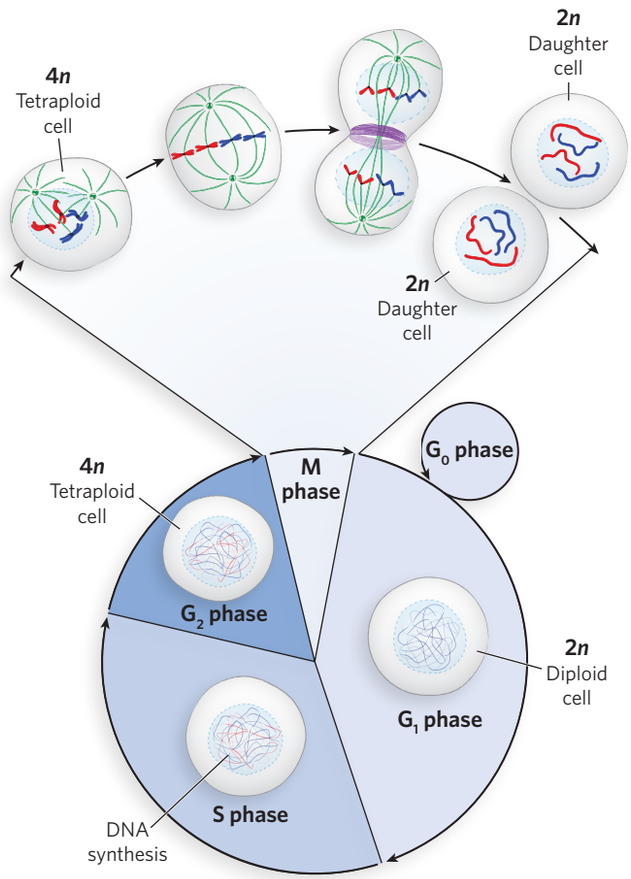



**FIGURE 2-10 The eukaryotic cell cycle.** Cells start in $G_1$ phase and progress to S phase, in which chromosomes are duplicated. In $G_1$, cells are diploid ($2n$, where $n$ is the species' chromosome number). After S phase, cells are tetraploid ($4n$) and enter $G_2$ phase (chromosomes are not visible in interphase; they begin to condense and become visible in the microscope at the beginning of M phase). In M phase, the duplicated chromosomes are equally divided, and the cell splits (by cytokinesis) into two daughter cells, each $2n$. These cells can enter a quiescent phase, $G_0$, which removes them from the cell cycle, or can undergo further division. The duration of each phase varies with species and cell type. A typical human cell in tissue culture has a cell cycle of about 24 hours: $G_1$ phase, 6–12 hours; S phase, 6–8 hours; $G_2$ phase, 3–4 hours; and M phase, 1 hour.

During **S phase** (*S* for synthesis), each chromosome is duplicated, and the two identical chromosomes remain together as a **sister chromatid pair**. The point where the sister chromatids are joined is called the **centromere** (**Figure 2-11**). At the end of S phase, each homologous chromosome exists as a sister chromatid pair, and thus the cell now contains four copies of each chromosome in the form of two sister chromatid pairs (i.e., the cell is $4n$, or tetraploid). The cell next enters **$G_2$ phase**, or the second gap in time, after S phase.

The final phase of the cell cycle is **M phase**, or **mitosis**, in which the duplicated chromosomes separate completely and the cell divides into two daughter cells, each $2n$. The two new cells reenter $G_1$ phase and then either continue through another division or cease to divide, entering a quiescent phase ($G_0$) that may last hours, days, or the lifetime of the cell (see Figure 2-10). Differentiated cells such as hepatocytes (liver cells) or adipocytes (fat cells) have acquired their specialized function and thereafter remain in $G_0$ phase.

Many scientists contributed to the description of events during mitosis, and Walther Flemming figures most prominently among them. By the late 1870s, the quality of microscopes included such developments as the oil immersion lens and the substage condenser. These advances made possible Flemming's detailed observations of dividing cells, published in 1878 and 1882, revealing the stages of mitosis as we know them today. The steps of mitosis are illustrated on the left side of **Figure 2-12** and summarized here.

**Interphase.** Cells not in mitosis are in interphase (which comprises $G_1$, S, and $G_2$). The cell is metabolically active and growing, and the chromosomes are duplicated (in S phase) in preparation for mitosis. The chromosomes are decondensed and not yet visible in the microscope.

**Prophase.** Prophase is the first stage of mitosis, following $G_2$ phase. As cells enter prophase, the sister

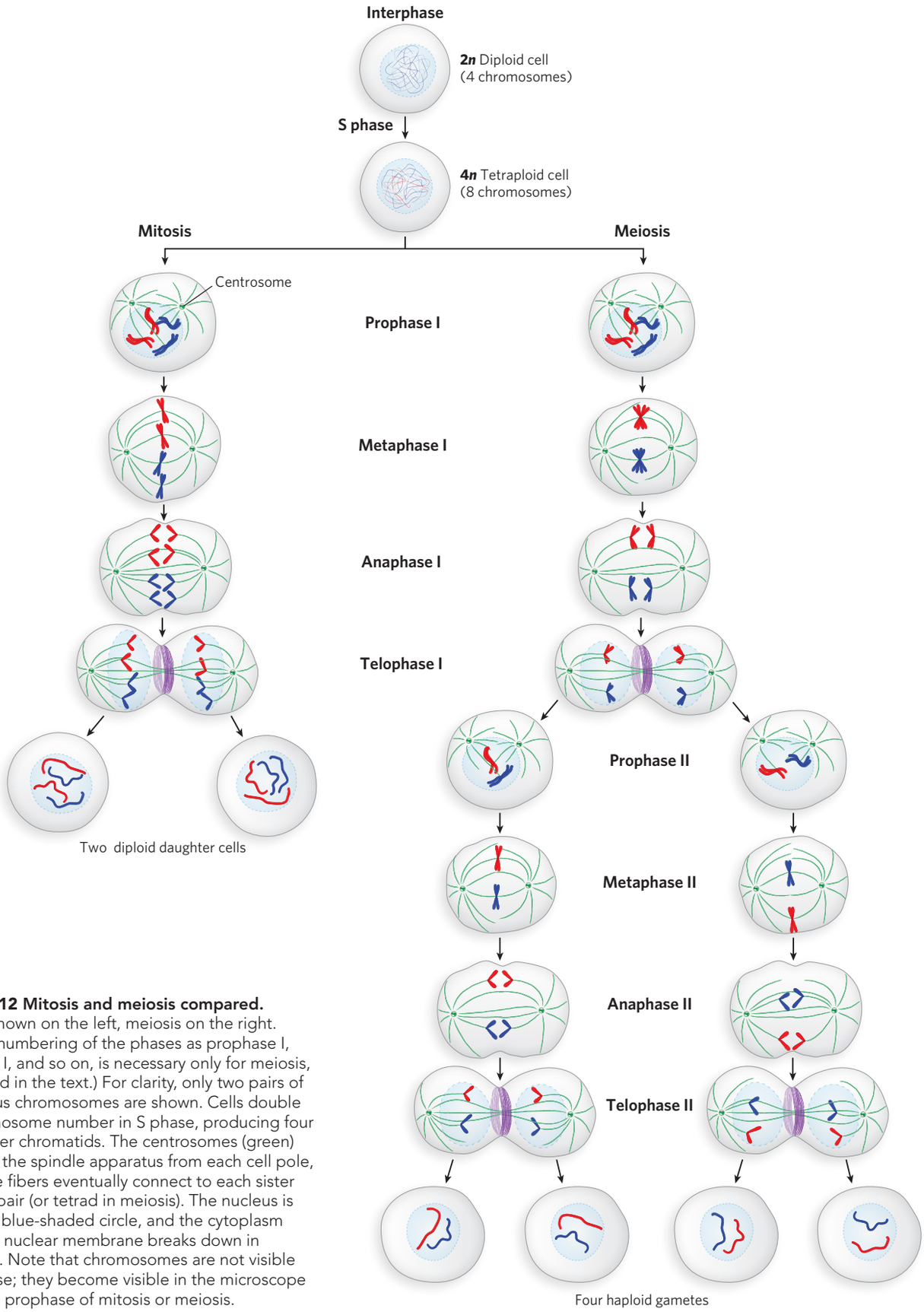



**FIGURE 2-12 Mitosis and meiosis compared.** Mitosis is shown on the left, meiosis on the right. (Note that numbering of the phases as prophase I, metaphase I, and so on, is necessary only for meiosis, as explained in the text.) For clarity, only two pairs of homologous chromosomes are shown. Cells double their chromosome number in S phase, producing four pairs of sister chromatids. The centrosomes (green) give rise to the spindle apparatus from each cell pole, and spindle fibers eventually connect to each sister chromatid pair (or tetrad in meiosis). The nucleus is shown as a blue-shaded circle, and the cytoplasm is gray. The nuclear membrane breaks down in metaphase. Note that chromosomes are not visible in interphase; they become visible in the microscope only during prophase of mitosis or meiosis.

chromatid pairs condense and become visible. Two organelles, called **centrosomes**, move to opposite poles of the cell. There they give rise to the spindle apparatus, an organized structure of protein fibers, which also becomes visible during prophase.

**Metaphase.** In metaphase, the membrane surrounding the nucleus dissolves. The spindle apparatus becomes fully developed and attaches to the centromeres of the sister chromatid pairs, directing them to align in the equatorial plane of the cell, a site also known as the **metaphase plate**.

**Anaphase.** Each sister chromatid pair separates at the centromere, becoming two separate chromosomes. The spindle apparatus moves the separated chromosomes toward opposite poles of the cell.

**Telophase.** The two chromosome sets reach opposite cell poles, nuclear membranes (nuclear envelopes) re-form, and the chromosomes become less distinct as they decondense. Telophase ends with **cytokinesis**, the physical splitting of the cytoplasmic membrane and cell contents to form two daughter cells.

## Meiosis: Chromosome Number Is Halved during Gamete Formation

Studies of fertilization in the late 1800s revealed that gamete cells contain only half the number of chromosomes found in somatic cells, and the union of two gametes reestablishes the diploid chromosome number; mitosis keeps this chromosome number constant during somatic cell division. These findings nicely explained how chromosome number is established and maintained in an organism, but they posed a new riddle: how are haploid gamete cells formed? The answer came in the 1880s from studies by Edouard van Beneden, Oskar Hertwig, and Theodor Boveri. Their studies of the ovary cells of a parasitic worm, *Ascaris*, revealed that the haploid female gamete, the egg (ovum), is formed by two consecutive cell divisions, in a process known as **meiosis** (see Figure 2-12, right). Later studies revealed that male gametes are also formed by meiotic cell divisions.

The most commonly studied organisms (such as *Ascaris*, sea urchin, or salamander larvae) had small, similar-looking chromosomes, and thus it was unclear whether chromosomes had unique identities. As far as anyone could tell, the cell simply divided an amorphous pool of chromosomes into equal parts to form daughter cells or gametes, rather than teasing apart two exact sets of different chromosomes. The unique nature of chromosomes, and the true precision with which the cell deals with them, was to come from work by Walter Sutton, in studies of the grasshopper. The grasshopper has chromosomes with unique morphologies that allow individual chromosomes to be observed during cell division. Sutton's observations demonstrated that during mitosis, one complete set of chromosomes is partitioned into each daughter cell (see the How We Know section at the end of this chapter). Chromosome segregation during meiosis proved to be just as precise, yielding one haploid set of chromosomes per gamete cell.

***The Process of Meiosis*** Meiosis involves a halving of chromosome number to form haploid ($n$) gametes. One might expect that a diploid cell simply divides once to form two haploid gamete cells, but this is not so. Meiosis involves two successive cell divisions, and four haploid gametes are formed from one diploid cell. The meiotic cell first goes through S phase, just as in mitosis, thereby increasing the number of each chromosome to four copies per cell ($4n$). In sharp contrast to mitosis, however, the homologous chromosomes—each of which is a pair of sister chromatids—find each other in meiosis and physically associate to form a **tetrad**. In the first meiotic cell division, the tetrad splits and the two sister chromatid pairs segregate into two new cells, each $2n$. This differs from mitosis, in which each sister chromatid pair splits at the centromere, resulting in two chromosomes that segregate into the two daughter cells.

Whereas mitosis involves only one cell division, the daughter cells from this first meiotic division divide a second time, but without an intervening S phase (no additional chromosome duplication). This second cell division closely resembles mitosis, except that the cells are diploid ($2n$) going into the second meiotic division (rather than $4n$, as in mitosis), so the second division reduces the diploid chromosome number by half, to form haploid gametes ($n$). In other words, the second meiotic cell division resembles mitosis in that sister chromatids separate, but in meiosis, for each chromosome, there is only one sister chromatid pair to split apart, whereas in mitosis the sister chromatid pairs of both homologous chromosomes are present at the metaphase plate, and each pair splits apart.

The phases of meiosis are summarized (and contrasted with mitosis) here and illustrated in Figure 2-12.

**Interphase.** Chromosomes are duplicated to form sister chromatid pairs; no obvious difference from mitosis.

**Prophase I.** Sister chromatid pairs become visible and the spindle apparatus forms. The difference from mitosis is that two homologous sister chromatid pairs find and associate with each other, forming a tetrad. In mitosis, two homologous sister chromatid pairs remain independent and do not associate with each other.

**Metaphase I.** The nuclear membrane breaks down, and the spindle apparatus moves the four homologous chromosomes to the metaphase plate as a tetrad, rather

than moving two homologous but independent sister chromatid pairs as in mitosis.

**Anaphase I.** Centromeres stay intact, and sister chromatids do not separate. Instead, the tetrad splits and the two sets of sister chromatid pairs move to opposite poles. By contrast, in mitosis, sister chromatids split at the centromere and individual chromosomes move apart.

**Telophase I.** Telophase occurs as in mitosis. The nuclear membrane re-forms and the cell divides.

The second meiotic cell division is a lot like mitosis, but there is no S phase between divisions and the cell is diploid going into the second cell division.

**Prophase II.** As in mitosis, sister chromatid pairs are visible, but there are half as many as in mitosis because the homologous sister chromatid pair is no longer present (it is in the other daughter cell formed from the first division).

**Metaphase II.** As in mitosis, the nuclear membrane breaks down and sister chromatid pairs align in the equatorial plane.

**Anaphase II.** As in mitosis, the centromere splits and the two separated chromosomes move to opposite poles of the cell.

**Telophase II.** As in mitosis, cytokinesis results in two cells and the two nuclear membranes form. Unlike mitosis, where the resulting daughter cells are diploid ($2n$), the second meiotic division produces daughter cells that are haploid ($n$).

In overview, two cells enter meiosis II, and the end result of meiosis is four $n$ cells from a single $2n$ cell. This is different from mitosis in two ways: the number of daughter cells formed and the $n$ value of each daughter cell.

***Sex Determination*** Cytological studies in many types of cells documented the existence of one or two chromosomes that behaved strangely in meiosis during the formation of male gametes. These were called accessory or X chromosomes; they either pair with a morphologically distinct partner chromosome or do not pair at all. Meiotic divisions therefore produce two types of sperm that differ in the accessory chromosome they contain. In 1905, Edmund B. Wilson and Nettie Stevens identified these accessory chromosomes in insects as the determinants of male and female sex and referred to them as X and Y chromosomes, or **sex chromosomes**. All other chromosomes are called **autosomes**.

Sex can be determined in many different ways, depending on the type of organism. For example, in mammals, a common way is the XY system (**Figure 2-13**). In XY determination, the female is XX and the male is XY; the male gametes are of two varieties, carrying either the X or Y chromosome. In many insects, sex is determined by the XO system, in which females are XX and males have one X and no other sex chromosome. The male gametes contain either an X chromosome or no sex chromosome. In both XY and XO determination, the union of male and female gametes is equally likely to produce male or female offspring. In birds, some insects, and other organisms, the ZW system determines sex. It is like the XY system but in reverse: males have two of the same chromosome (ZZ), whereas females have one copy each of the Z and W chromosomes.

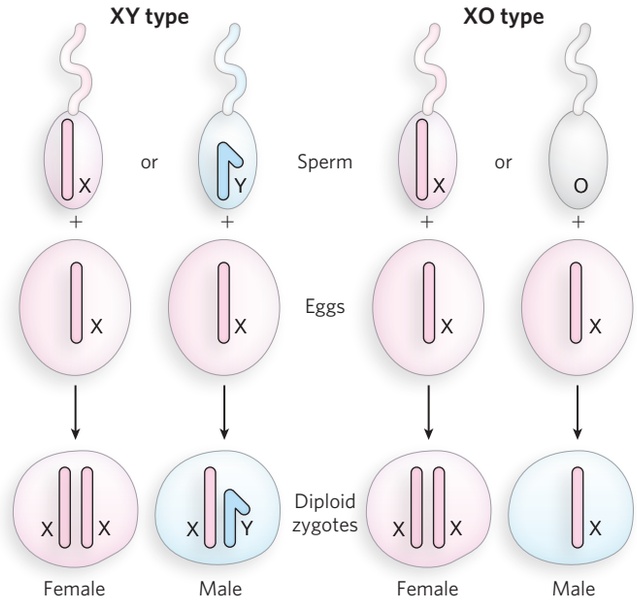



**FIGURE 2-13 The chromosomal basis of sex determination.** In the XY type of sex determination, meiosis produces sperm with either an X or a Y chromosome, in a 50:50 ratio (autosomes are not shown). All eggs have one X. Fertilization results in either an XX (female) or XY (male) zygote. The Y chromosome confers maleness. In XO determination, meiosis produces two types of sperm in a 50:50 ratio, either with or without an X chromosome (X or O). All eggs have one X. Fertilization results in either an XX (female) or an XO (male) zygote. The number of copies of X determines the sex.

Edmund B. Wilson, 1856–1939
*[Source: Courtesy of University Archives, Columbia University in the City of New York.]*

Nettie Stevens, 1861–1912
*[Source: Courtesy of University Archives, Columbia University in the City of New York.]*

**SECTION 2.2 SUMMARY**

- Organisms are composed of cells, which have intricate intracellular structures, including chromosomes located in the nucleus.
- Cells that are not actively dividing contain two complete sets of unique chromosomes in the nucleus; they are diploid, or $2n$ (except gametes, which are haploid, or $n$).
- The cell cycle consists of four stages; $G_1$ phase, S phase (synthesis), $G_2$ phase, and M phase (mitosis). The chromosomes of a diploid cell are duplicated in S phase (during interphase) and then carefully segregated into two daughter cells during mitosis, which proceeds through four stages: prophase, metaphase, anaphase, and telophase. The resulting daughter cells are also diploid.
- Meiosis is a specialized type of cell division that halves the diploid chromosome number ($2n$) to produce haploid gametes ($n$), each containing one complete set of chromosomes.
- Haploid gametes unite during fertilization to reestablish the diploid state of the organism.
- Sex is determined by an accessory chromosome that is paired either with a similar chromosome or with a distinct, differently shaped chromosome, or has no partner at all. These special chromosomes are called sex chromosomes; all other chromosomes are autosomes.

## 2.3 THE CHROMOSOME THEORY OF INHERITANCE

Walter Sutton's studies on chromosomes were performed just as Mendel's work was being rediscovered. Sutton found himself at a remarkable intersection of two fields: cytology and genetics. He made the connection between chromosomes and Mendel's particles of heredity in his classic 1903 paper "The Chromosomes in Heredity" (see the How We Know section at the end of this chapter). He proposed that chromosomes contain Mendel's particles of heredity and that the particles come in pairs: chromosomes exist as homologous pairs in diploid cells. Mendel's particles—gene pairs—separate and assort independently into gamete cells; homologous chromosome pairs also separate and assort into haploid gamete cells.

Sutton's hypothesis that genes are located on chromosomes received much attention and became known as the **chromosome theory of inheritance**. But there was still no proof that genes were actually on chromosomes. This would be left for Thomas Hunt Morgan and his students to establish in their classic studies of fruit flies. Interestingly, Morgan did not initially believe in the chromosome theory of inheritance. But his experiments would inevitably lead him to this conclusion, and his name would become as linked to the chromosome theory as are genes themselves.

### Sex-Linked Genes in the Fruit Fly Reveal That Genes Are on Chromosomes

In 1908, Morgan initiated his studies of the fruit fly, *Drosophila melanogaster* (see the Model Organisms Appendix). In those days, it was essential to keep costs to a minimum, as funding for science was scarce. The fruit fly is small; it could be grown in large numbers and was inexpensive to maintain. Flies also have an array of phenotypic features suitable for genetic studies, and they have just four homologous pairs of chromosomes that can be visualized under the microscope. Most important of all, the generation time of the fly is less than 2 weeks, and each female can lay hundreds of eggs. These features made fruit flies far superior to other model organisms of the day.

Thomas Hunt Morgan, 1866–1945 *[Source: Courtesy of the Archives, California Institute of Technology.]*

Morgan's famous Fly Room at Columbia University was small and cramped, but conducive to science; several of his students became famous for their discoveries in genetics. In 1910, Morgan noticed a male fly with white eyes that spontaneously appeared in a bottle of red-eyed flies. He immediately set up crosses to determine whether the white-eye trait was inheritable. It was, but in an unusual way. **Figure 2-14** shows Morgan's first experiment. Normal flies have red eyes, straight wings, and a gray body, referred to as **wild-type** traits.

**KEY CONVENTION**

The allele that appears with the greatest frequency in a natural population of a species is called the wild-type allele. All other alleles are mutants. Wild-type alleles can be dominant or recessive to a mutant allele.

Morgan crossed a red-eyed female with the mutant white-eyed male. All the $F_1$ hybrid progeny had red eyes, the wild-type phenotype. This told Morgan that the allele for red eyes is dominant to the allele for white eyes. His next cross, an $F_1$ female with an $F_1$ male, produced some $F_2$ progeny with white eyes—the expected result for a typical recessive allele. But surprisingly, all the white-eyed flies were male. All the $F_2$ females had red eyes, and about half the $F_2$ males had red eyes. It seemed that the trait for white eyes was somehow connected to sex. Morgan performed a variety of additional crosses and found, again to his surprise, that the white-eye trait mirrors the segregation behavior of the X chromosome

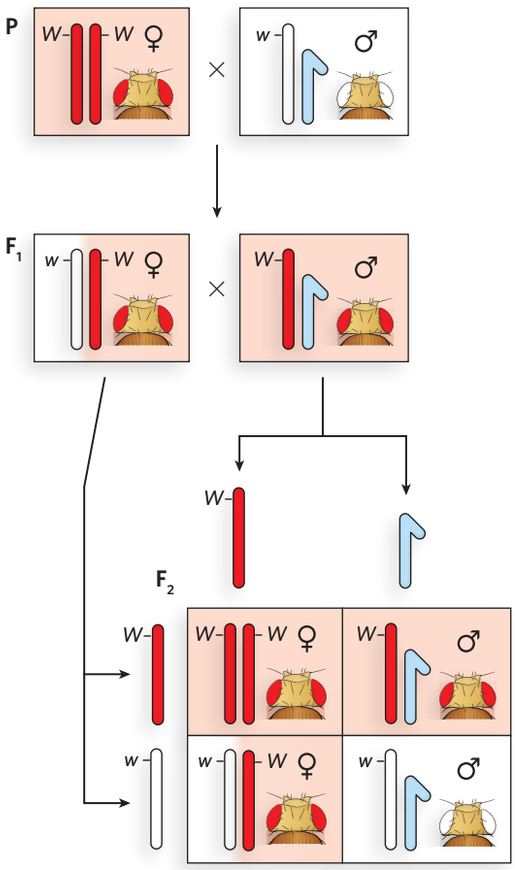



**FIGURE 2-14 X-linkage of the white-eye allele.** Results of a cross between a red-eyed female and a white-eyed male fly. $F_1$ flies are red-eyed. The white-eye trait reappears in the $F_2$ generation. Only males in the $F_2$ generation have white eyes. Morgan realized that these results make sense if the white-eye allele is located on the X chromosome. *[Source: Adapted from T. H. Morgan et al.,* Mechanism of Mendelian Heredity, *Henry Holt, 1915.]*

(see Figure 2-14). Morgan's findings, linking a genetic trait to a particular chromosome, were convincing evidence that genes are located on chromosomes.

The alleles for this eye-color gene can be represented as $X^W$ for red-eyed and $X^w$ for white-eyed. In this genetic nomenclature, the *X* represents the X chromosome, and a superscript *W* is used for the dominant red-eye allele and a superscript *w* for the recessive white-eye allele. The letters *R* and *r* (for red and white, respectively), which might be expected from the convention introduced earlier in the chapter, are not used in this case because there are many different mutant alleles that affect eye color. There is only one wild-type color, so the wild-type and different mutant alleles are named according to the different mutant colors.

Further evidence that genes are located on chromosomes came from Calvin Bridges, an associate in Morgan's laboratory. Bridges hypothesized that if genes are located on chromosomes, then some genetic anomalies should also produce visible abnormalities in the chromosomes themselves. Bridges crossed white-eyed female flies ($X^wX^w$) with red-eyed males ($X^WY$) that he had in his fly collection. Most progeny were the expected white-eyed males and red-eyed females. However, Bridges noticed a few rare (<0.1%) white-eyed females and red-eyed males, which he called "primary exceptionals." Bridges made the unusual prediction that if genes are truly on chromosomes, then primary exceptional flies will have an abnormal chromosome number. He reasoned that the primary exceptional phenotype might be explained by defective meiosis in the female parent, in which X chromosomes did not separate, so producing an egg with two X chromosomes and an egg with no X chromosome (**Figure 2-15a**). Thus, exceptional white-eyed females, which must have two $X^w$ chromosomes, received them from the abnormal $X^wX^w$ egg, plus a *Y* chromosome from the sperm, for a genotype of $X^wX^wY$ (note that an $X^W$ sperm would bring in a dominant red-eye gene) (**Figure 2-15b**). By similar reasoning, the exceptional red-eyed male originated from fertilization of the abnormal egg having no X chromosome by a sperm containing a single $X^W$ chromosome, for a genotype of $X^WO$. Note that although flies (like mammals) have X and Y chromosomes, sex in *D.melanogaster* is determined by the number of copies of the X chromosome, not by the presence or absence of the Y chromosome. Thus, an XXY fly is female and an XO fly is male.

When Bridges examined the chromosomes of primary exceptional flies, the results followed his predictions precisely (**Figure 2-15c**). His study was an impressive demonstration that genes are located on chromosomes, because, to explain the genetic results, he had hypothesized highly unusual outcomes that could be verified by examining the chromosomes directly. This abnormal assortment of chromosomes during meiosis is called **nondisjunction**.

## Linked Genes Do Not Segregate Independently

Chromosomes, not individual genes, segregate into gamete cells, so one might expect two different genes on the same chromosome to stay together during meiosis and thus to be inherited together (i.e., they would not obey Mendel's second law). Take, for example, two genes, *A* and *B*, on the same chromosome. A cross of *AABB* and *aabb* parents will produce the *AaBb* $F_1$ hybrid, but particular combinations of alleles (*A, B* and *a, b*) are linked on the same chromosome. Therefore, the $F_1$ hybrid can produce only two types of gametes, *AB* and *ab*, rather than all four possible gametes produced if the genes separated and assorted randomly—*AB*, *Ab*, *aB*, *ab*.

To determine the genotype of an $F_1$ hybrid experimentally, it is crossed with a strain that is homozygous recessive (*aabb*), and the progeny reveal both the recessive and dominant alleles of the $F_1$ gametes. Such a cross is

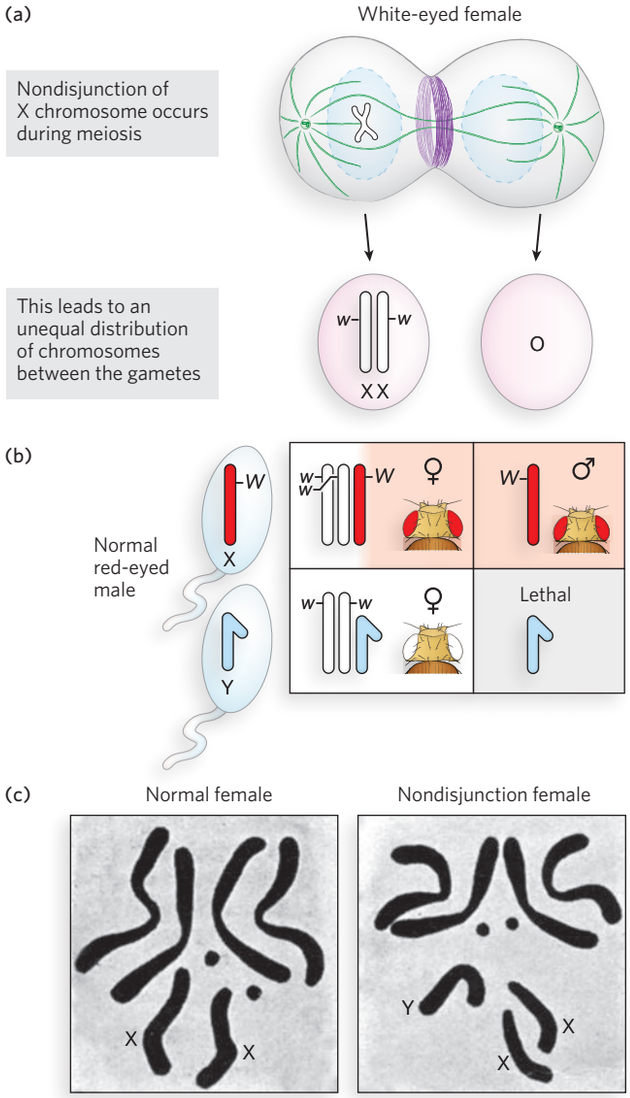



**FIGURE 2-15 Nondisjunction.** A rare occurrence during meiosis, nondisjunction produces gametes that have either one extra chromosome or one fewer chromosome. (a) Nondisjunction of the X chromosome is shown here; the white-eye allele is on the X chromosome. (b) Fertilization with normal sperm produces adult flies with odd numbers of chromosomes, as illustrated in the Punnett square. (c) Bridges predicted the occurrence of nondisjunction to explain rare progeny phenotypes, and cytologic examination of chromosomes in the rare progeny confirmed the predicted extra Y chromosome in a white-eyed female. *[Source: (c) C. Bridges, Genetics 1:107–163, 1916.]*

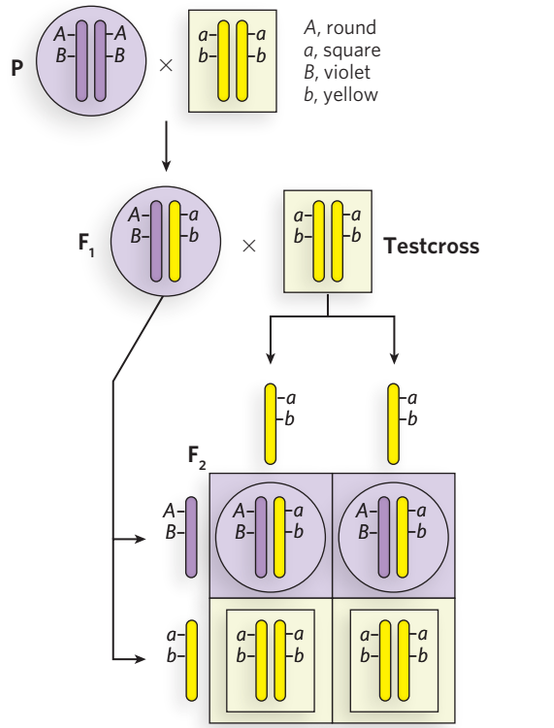



**FIGURE 2-16 The inheritance of linked genes.** Linked genes segregate together because they are on the same chromosome—that is, they are part of the same DNA molecule. In this hypothetical example, the dominant and recessive genes and their phenotypes are: *A*, round; *a*, square; *B*, violet; *b*, yellow. The cross between homozygous dominant and homozygous recessive parents produces $F_1$ *AaBb* progeny with linked alleles *A, B* and *a, b*. A testcross with a double-recessive homozygous individual (*aabb*) reveals the genotypes of the gametes produced by the $F_1$ progeny. The Punnett square shows the expected results for completely linked genes. The $F_1$ generation can produce only *AB* and *ab* gametes, and thus only two types of $F_2$ progeny are observed; they have the same phenotype as the original P generation.

known as a **testcross**. If the two genes separate in the gametes of the $F_1$ hybrid, the $F_2$ generation will exhibit all four possible phenotypes. If the two genes are linked, the $F_1$ hybrid will produce only two types of gametes (*AB* and *ab*) and the $F_2$ generation will display only the two parental phenotypes (**Figure 2-16**).

An example of linked genes in *Drosophila* is illustrated in **Figure 2-17**, for a body-color gene with alleles *b* (black body) and *B* (gray body), and a wing-shape gene with alleles *v* (vestigial wings) and *V* (long wings). Consider the parental cross *BBvv* (gray body, vestigial wings) × *bbVV* (black body, long wings). All $F_1$ progeny (*BbVv*) have a gray body and long wings. To determine whether the two genes are linked, a testcross is performed between an $F_1$ fly and a double-recessive *bbvv* fly. The $F_2$ progeny are mainly of two types and exhibit the same characteristics as the P generation (gray body, vestigial wings; black body, long wings). Thus, the two genes are linked. Had the genes assorted completely independently, mixed phenotypes would have been observed in the $F_2$ generation (black body, vestigial wings; gray body, long wings) in amounts equal to the parental phenotypes.

The results of the experiment, however, do not show complete gene linkage. There are some $F_2$ generation flies with mixed phenotypes, indicating that linked genes sometimes unlink. How can this happen—how do linked genes become unlinked?

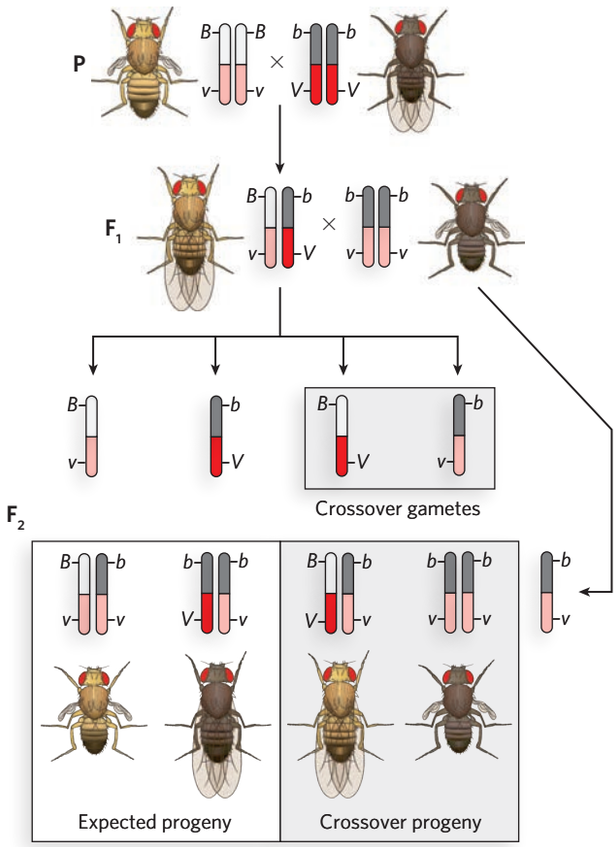



**FIGURE 2-17 Unlinking genes by crossing over.** Linked genes can become unlinked by chromosome recombination, or crossing over, during meiosis. The chromosomes containing the linked genes are illustrated in diploid cells and in gametes. To analyze gametes produced by $F_1$ flies, the $F_1$ hybrid is crossed with a double-recessive fly of genotype *bbvv*. In the $F_1$ gametes, *B* and *v* are linked, and *b* and *V* are linked, so all $F_2$ progeny are expected to contain these same two combinations. The double-recessive fly always contributes a *bv* gamete. But, in fact, four types of $F_2$ progeny are observed: two are the expected phenotypes; the other two contain *b*, *v* and *B*, *V*, resulting from gametes in which the linked alleles were unlinked by recombination during meiosis. The two crossover phenotypes are produced at equal frequency (17% each). *[Source: Adapted from T. H. Morgan et al.,* Mechanism of Mendelian Heredity, *Henry Holt, 1915.]*

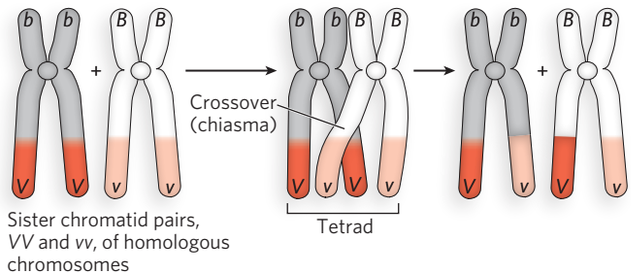



**FIGURE 2-18 Crossing over in the tetrad.** Two homologous chromosomes are duplicated in S phase to produce two sister chromatid pairs (left), one homozygous *VV* and the other homozygous *vv*. The sister chromatid pairs are homologous to each other, and they pair to form a tetrad in prophase I of meiosis, before the first cell division (middle). Recombination—as evidenced by the exchange of *V* and *v* alleles—occurs at crossovers, or chiasmata, the sites where chromosomes intertwine, resulting in genetic exchange between the two chromosomes of the sister chromatid pairs (right).

## Recombination Unlinks Alleles

Morgan noticed that linked genes do not always stay linked, but instead show a low, though reproducible, frequency of separating. Take, for example, the cross of flies with linked genes shown in Figure 2-17. Linked genes should give only parental phenotypes in the $F_2$ progeny, yet a low frequency of mixed-phenotype $F_2$ progeny was observed. These **recombinant** flies could be produced only if the linked genes were unlinked and separated during gamete formation. Both possible types of mixed-phenotype recombinant flies were produced (black body, vestigial wings; gray body, long wings) and appeared with equal frequency: each was 17% of the total $F_2$ population.

To explain how linked genes become unlinked, and why they produce equal amounts of the two mixed phenotypes, Morgan hypothesized that one of the linked alleles on one chromosome (e.g., the long-wing allele) trades places with the homologous allele (vestigial-wing allele) on the homologous chromosome (**Figure 2-18**). In other words, genes hop from one homologous chromosome to the other and do so in a reciprocal fashion. This reciprocal exchange of alleles between chromosomes is called **recombination**, or **crossing over**.

The idea that chromosomes exchange genetic material had been suggested earlier, in cytological studies by F. A. Janssens in 1909. Janssens noticed that during meiosis, the four chromosomes of the tetrad coil around one another and form cross-shaped junctions, which he called **chiasmata** (see Figure 2-18). He proposed that, as the mechanical forces pull the sister chromatid pairs apart during the first division of meiosis, the intertwined chromosomes break at the same place and then rejoin, but with the opposite chromosome. The first experimental proof that genetic crossing over is mediated by physical recombination between two chromosomes came from a study of corn by Barbara McClintock and Harriet Creighton (see the How We Know section at the end of this chapter).

We now know that recombination events are mediated by specialized proteins that catalyze DNA breakage and rejoining within homologous chromosomes of the tetrad. Crossing over is a frequent event during meiosis (**Figure 2-19**), occurring at least once in each tetrad. It is thought that meiotic recombination was selected for during evolution because it helps generate diversity within a species. Homologous recombination is discussed in detail in Chapter 13.

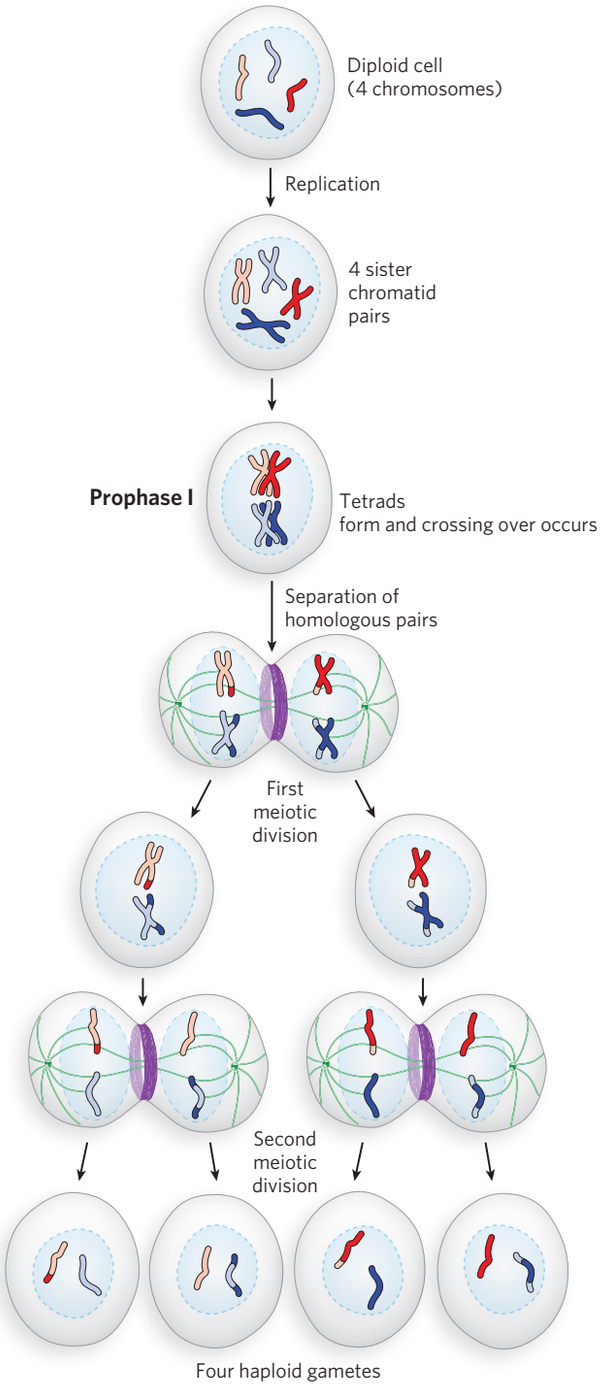



**FIGURE 2-19 Recombination during meiosis.** The chromosomes of a diploid cell (four chromosomes, two homologous pairs, are shown here) replicate, and each pair is held together at the centromere, forming four sister chromatid pairs. In prophase I, at the start of the first meiotic division, the two homologous sets of sister chromatid pairs align to form tetrads. Crossovers occur within the tetrads. In the first meiotic division, homologous pairs of chromosomes segregate into daughter cells. Each sister chromatid pair then lines up in preparation for the second meiotic division, which produces four haploid gamete cells. Each gamete has two chromosomes, half the number of the diploid cell. *[Source: Adapted from D. L. Nelson and M. M. Cox,* Lehninger Principles of Biochemistry, *5th ed., W. H. Freeman, 2008, Fig. 25-31.]*

## Recombination Frequency Can Be Used to Map Genes along Chromosomes

Different pairs of linked genes exhibit different frequencies of crossing over, but the frequency is constant for a given pair of genes. Alfred Sturtevant, a student of Morgan's, rationalized this observation by assuming that the frequency of crossing over corresponds to the distance between the two linked genes. He reasoned that the greater the distance, the more room there is for recombination to occur, thereby allowing linked genes to separate with greater frequency. With this logic, he used the frequency of crossing over to map the relative positions of pairs of linked genes along *Drosophila* chromosomes (**Figure 2-20**). Genetic map units, calculated from the

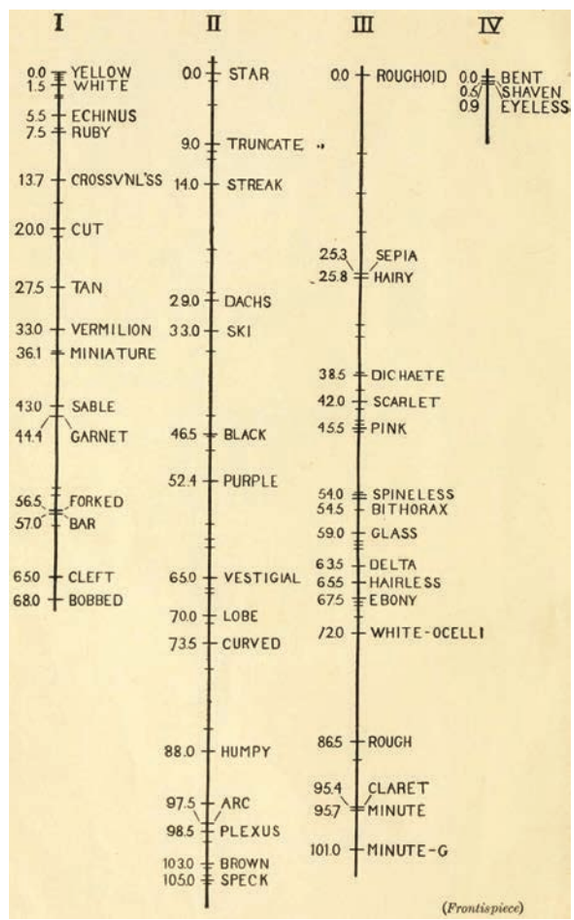



**FIGURE 2-20 Using recombination frequency to create genetic maps.** Sturtevant created genetic maps showing the positions of genes along the four *Drosophila* chromosomes, based on the frequency of crossing over between many pairs of linked genes. Linked genes fall into four groups, corresponding to the four different chromosomes in *Drosophila*, represented here by the vertical lines. Numbers on the left side of each chromosome are genetic map units, in centimorgans. Along the right side are the names of mutant alleles used in the crosses. *[Source: From The Mechanism of Mendelian Heredity by T. H. Morgan, A. H. Sturtevant, H. J. Muller, C. B. Bridges, 1915, 1922 Henry Holt and Company, NY.]*

frequency of crossing over, are called centimorgans (cM) in honor of Thomas Hunt Morgan; however, they do not necessarily reflect accurate physical distances between genes. Some regions of chromosomes tend to promote recombination, giving the impression that genes are farther apart than they really are; conversely, other regions repress crossing over, and genes seem to be closer than they are. The accuracy of genetic map distances is also limited by one crossing-over event interfering with another. However, recombination frequencies do provide useful genetic maps, because the data reveal the linear order of genes along a chromosome and provide a first approximation of the distance between them.

An example of **recombination mapping** is illustrated in **Figure 2-21** for linked genes *A*, *B*, and *C*. Consider the frequency of crossing over of the linked gene pair *A* and *B* in fruit flies (Figure 2-21a). A parent homozygous for dominant alleles is crossed with a double-recessive fly (*AABB* × *aabb*), and the frequency of crossing over (the frequency of production of *Ab* and *aB* gametes by the $F_1$ flies) is determined from the percentage of recombinant $F_2$ progeny (*Aabb* and *aaBb*). This is repeated for the *A*, *C* pair and *B*, *C* pair. The results are shown in Figure 2-21b.

The greater frequency of recombinants for the *A*, *B* pair than for the *A*, *C* pair indicates that genes *A* and *B* are farther apart than genes *A* and *C*. However, gene *C* could be between *A* and *B* or on the opposite side of *A* from *B*. The frequency of crossing over of the *B*, *C* pair resolves the ambiguity: *C* is between *A* and *B*.

The frequency of recombinants for the *A*, *B* pair (26%) is somewhat less than the added frequencies of recombinants for the *B*, *C* pair and *C*, *A* pair (28%). This is because the probability of multiple crossing-over

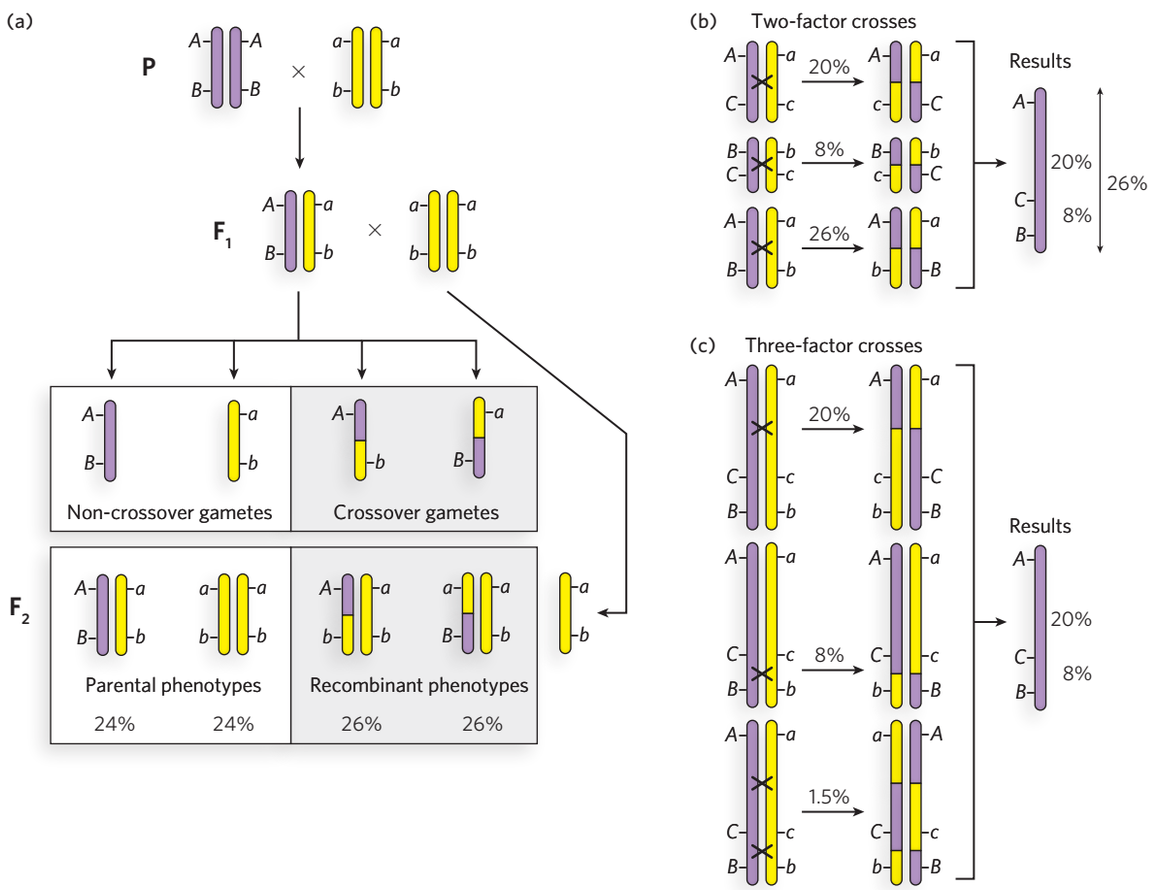



**FIGURE 2-21 Recombination mapping.** (a) The procedure for analyzing a two-factor cross. Diploid cells and gametes are illustrated to show the origin of recombinant $F_2$ progeny. Chromosomes with dominant linked genes are purple; chromosomes with recessive linked genes are yellow. Crossing over results in purple-and-yellow hybrid chromosomes. $F_1$ progeny are crossed with a homozygous double-recessive fly to analyze the genotypes of gametes produced by the $F_1$ hybrid. (b) Analysis of linked genes using three two-factor crosses. Crossing over is indicated by × between two chromosomes. (c) Analysis of linked genes using three three-factor crosses. Two- and three-factor crosses lead to the same conclusion about gene order (*ACB*).

events between linked genes is higher the farther they are apart. For example, a single crossover unlinks the genes, but a second crossover links them again. Therefore, an odd number of crossovers will unlink genes, and an even number of crossovers relinks them, resulting in a maximum frequency of recombination of 50%. The frequency of independent assortment of genes on different chromosomes is also 50%, because there is a 50:50 chance that two chromosomes will segregate together into the same gamete. Because crossing over is a frequent occurrence in meiosis, genes on the same chromosome often assort independently. Therefore, recombination mapping is accurate only for pairs of linked genes that are close together.

Analysis of three genes in a single experiment, known as a three-factor cross, provides a convenient method to identify or confirm their order along the chromosome. To illustrate this, consider a three-factor cross between genes *A*, *B*, and *C* (Figure 2-21c). A fly that is homozygous dominant for three linked genes is crossed with a fly that is double recessive for all three genes. The $F_1$ progeny (*AaBbCc*) are then crossed with a fly that is double recessive for all three genes. Most $F_2$ progeny exhibit the parental phenotypes, but crossing over will produce six possible recombinants: three recombinants containing two dominant traits and three reciprocal recombinants containing one dominant trait. If the gene order is *ACB*, generation of the *AcB* and *aCb* recombinants requires two crossover events—one between *A* and *C*, and another between *C* and *B*. The *aCB* (and *Acb*) or *ACb* (and *acB*) recombinants each require only one crossover. Because a double crossover is much less frequent than a single crossover, the far lower frequency of the double crossover (1.5% in this example, yielding *AcB* and *aCb*) reveals which gene (*C* in this case) is between the other two.

**SECTION 2.3 SUMMARY**

- Direct evidence that genes are located on chromosomes came from intensive studies of the fruit fly, Drosophila melanogaster, by Thomas Hunt Morgan. Segregation of the white-eye mutant allele with the X chromosome suggested that genes are associated with chromosomes.
- Calvin Bridges's correlation of mutant genes with chromosome abnormalities showed definitively that genes are located on chromosomes.
- Linked genes, genes on the same chromosome, violate Mendel's second law and assort together into gametes. However, linked genes must be close together on the chromosome to stay linked. The farther apart they are, the more likely they are to be separated by recombination during meiosis.
- Recombination frequency can be used to map the relative positions of genes along a chromosome.

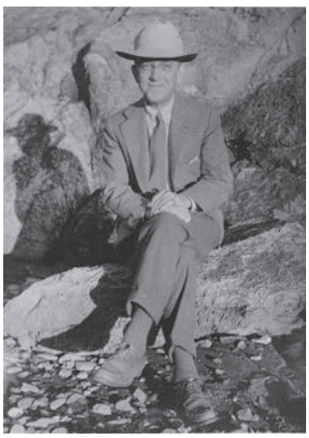
## 2.4 FOUNDATIONS OF MOLECULAR GENETICS

The union of genetics and cytology in the early 1900s was an exceedingly productive time. Heredity was based in genes, which were located on chromosomes. But what are genes made of? To some scientists, genes were almost unreal, a mental construct to explain real phenomena. We now know, of course, that genes are made of DNA. In fact, DNA was discovered decades before its significance was understood. The recognition of DNA as the genetic material, and the solution of its chemical and three-dimensional structure, brought genetics out of the realm of imagination and into the realm of chemistry. These discoveries sparked the fusion of chemistry and genetics to give us an entirely new scientific discipline: **molecular genetics** or, more generally, **molecular biology**. In this section, we outline some discoveries that led to our current understanding of DNA as the repository of biological information. We also describe how the information in DNA is translated into functional RNAs and proteins, and how this knowledge furthers our understanding of human health and disease.

### DNA Is the Chemical of Heredity

**Deoxyribonucleic acid (DNA)**, as we have noted, was identified long before its importance was recognized. The history begins with Friedrich Miescher, who carried out the first systematic chemical studies of cell nuclei in 1868. Miescher obtained white blood cells from pus that he collected from discarded surgical bandages. He carefully isolated the nuclei and then ruptured the nuclear membranes, releasing an acidic, phosphorus-containing substance that he called *nuclein*. Nuclein, a **nucleic acid**, was a new type of chemical polymer, different from all others previously identified. Around the turn of the century, Albrecht Kossel investigated the chemical structure of nucleic acids—both DNA and a similar molecule called **ribonucleic acid (RNA)**—and found that they contain nitrogenous bases, or nitrogen-containing basic compounds. Kossel identified five types of nitrogenous bases: adenine (A), guanine (G), cytosine (C), thymine (T), and uracil (U) (described in Chapters 3 and 6). By 1910, other investigators had determined that nitrogenous bases were but one component of a larger unit called a **nucleotide**, which consists of a phosphate group, a pentose sugar, and a nitrogenous base. DNA is composed of nucleotides that contain the bases A, G, C, or T, and RNA is composed of nucleotides with the bases A, G, C, or U.

By the 1920s, the chemical basis of heredity was thought to lie in chromosomes, but chromosomes are composed of both DNA and protein. Which one is the



Oswald Avery, 1877–1955
*[Source: Science Source.]*

Frederick Griffith, 1879–1941
*[Source: Science Source.]*

hereditary chemical? DNA was initially ruled out because it was believed to be too simple—just a repeating polymer of four different nucleotides. Surely such a monotonous molecule lacked the complexity to code for the working apparatus of a living cell! Attention turned to the other biopolymer, protein, for a chemical explanation of heredity. Only after biochemical studies conducted in the 1940s pointed to DNA as the genetic molecule was attention refocused on the structure and function of this molecule.

DNA was shown to be the chemical of heredity in the 1940s, by Oswald T. Avery and his colleagues at the Rockefeller Institute in New York City. Their starting point was an observation made in 1928 by English microbiologist Frederick Griffith, who studied the pneumonia-causing bacterium *Streptococcus pneumoniae*. This pneumococcus exists as two types, virulent (disease-causing) and nonvirulent. Griffith noticed that virulent bacteria produced smooth colonies when grown on Petri plates, but colonies of a nonvirulent strain appeared rough. The difference in appearance lies in a polysaccharide capsule coat present only on the virulent strains. Griffith found that heat-killed virulent bacteria transformed live nonvirulent bacteria into live virulent bacteria (**Figure 2-22a–d**). The nonvirulent bacteria somehow acquired the smooth-colony trait from the heat-killed bacteria. The results suggested that the genetic material coding for capsules remained intact even after the virulent (smooth-colony) bacteria were killed, and this material could enter another cell and recombine with its genetic material.

Avery and his colleagues reproduced Griffith's results, and they analyzed the heat-killed virulent bacterial extract for the chemical nature of the transforming factor. They selectively removed either DNA, RNA, or protein from the heat-killed virulent bacterial extract by treatment with DNase, RNase, or proteases (enzymes that specifically break down one of these cellular components, leaving the others intact). The DNase-treated extract lost the capacity to transform nonvirulent rough-colony cells into a virulent smooth-colony strain. The researchers then extracted DNA from virulent bacteria, purified it of contaminating proteins and RNA, and

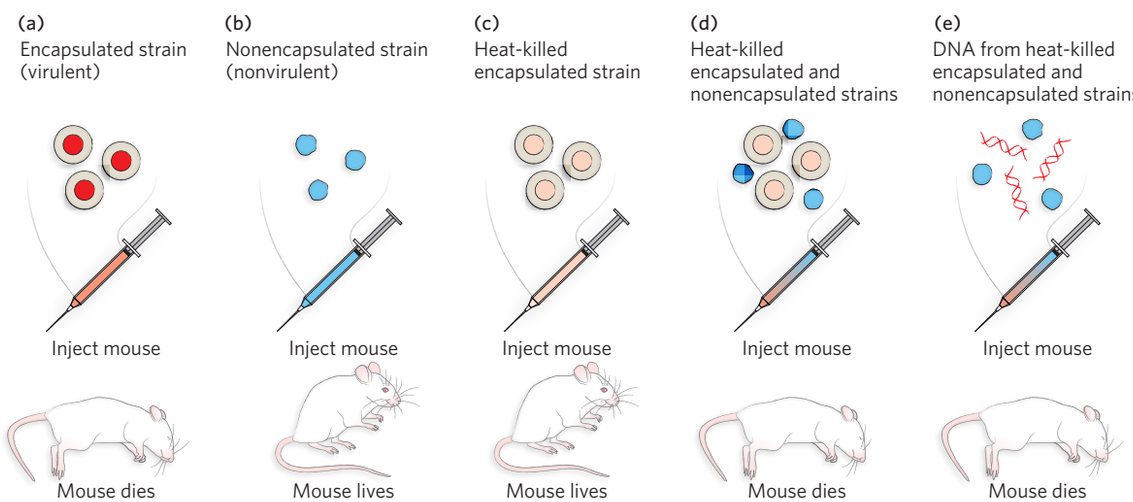



**FIGURE 2-22 Transformation of nonvirulent bacteria to virulent bacteria by DNA.** When injected into mice, (a) the encapsulated strain of pneumococcus (*Streptococcus pneumoniae*), producing smooth colonies, is lethal, whereas (b) the nonencapsulated strain, producing rough colonies, and (c) the heat-killed encapsulated strain are harmless. (d) Griffith's research had shown that adding heat-killed virulent bacteria to a live nonvirulent strain (each harmless to mice on their own) permanently transformed the live strain into lethal, virulent, encapsulated bacteria. (e) Avery and his colleagues extracted the DNA from heat-killed virulent pneumococci, removing RNA and protein as completely as possible, and added this DNA to nonvirulent bacteria, which were permanently transformed into a virulent strain.

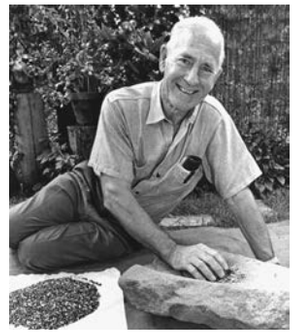
showed that this pure DNA was still capable of transforming nonvirulent bacteria into the virulent strain (**Figure 2-22e**). In 1944, Avery and colleagues reported their surprising conclusion that DNA was the carrier of genetic information. Another classic experiment, by Alfred Hershey and Martha Chase, supported this conclusion that DNA is the chemical of heredity (see the How We Know section at the end of this chapter).

## Genes Encode Polypeptides and Functional RNAs

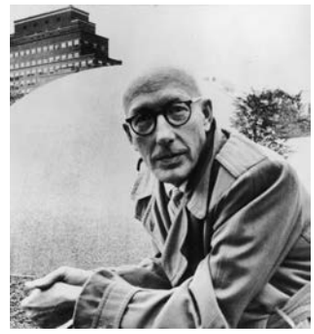
DNA and protein are chemically very different, and it was puzzling how a DNA sequence could code for a protein sequence. Regardless of the details, however, it now became easy to understand that mutations in a gene could lead to altered enzymes. In fact, even before the DNA structure was solved, the relationship between genes, mutations, and enzymes was well understood.

In 1902, the physician Archibald Garrod studied patients with alkaptonuria, a disease of little consequence for the patients, except that they excreted urine that turned black. Mendel's work had recently been rediscovered, and by noticing how alkaptonuria was inherited, Garrod realized that this disorder behaved as a recessive trait. It was already known that the synthesis and breakdown of biomolecules occur in multistep pathways, each step requiring a different **enzyme**—a protein catalyst that facilitates the reaction. Garrod hypothesized that alkaptonuria was caused by a mutation that inactivated a gene required for the production of one enzyme in a metabolic pathway. Without this functional enzyme, the pathway was blocked, resulting in the buildup of an intermediate compound, which was excreted in the urine and turned black. Garrod's reasoning drew the connection between a mutation in a gene and a mutation in an enzyme.

Formal proof that genes encode enzymes came from a series of elegant experiments in the 1940s by George Beadle and Edward L. Tatum. They introduced a new microorganism into the study of genetics: the bread mold, *Neurospora crassa* (see the Model Organisms Appendix). This haploid organism can grow on a simple, defined medium, called minimal medium. Minimal medium contains sugar, nitrogen, inorganic salts, and biotin, and the cell must make all the rest of the biochemicals that it needs to live from these simple starting compounds. Beadle and Tatum irradiated *Neurospora* spores to intentionally produce mutations, then germinated individual spores on a complete medium (i.e., one made with cell extracts that have all the necessary amino acids, nucleotides, and vitamins) to obtain genetically pure colonies and their spores. These spores were then tested for their ability to germinate on the minimal medium.



George Beadle, 1903–1989
*[Source: Courtesy of the Archives, California Institute of Technology.]*

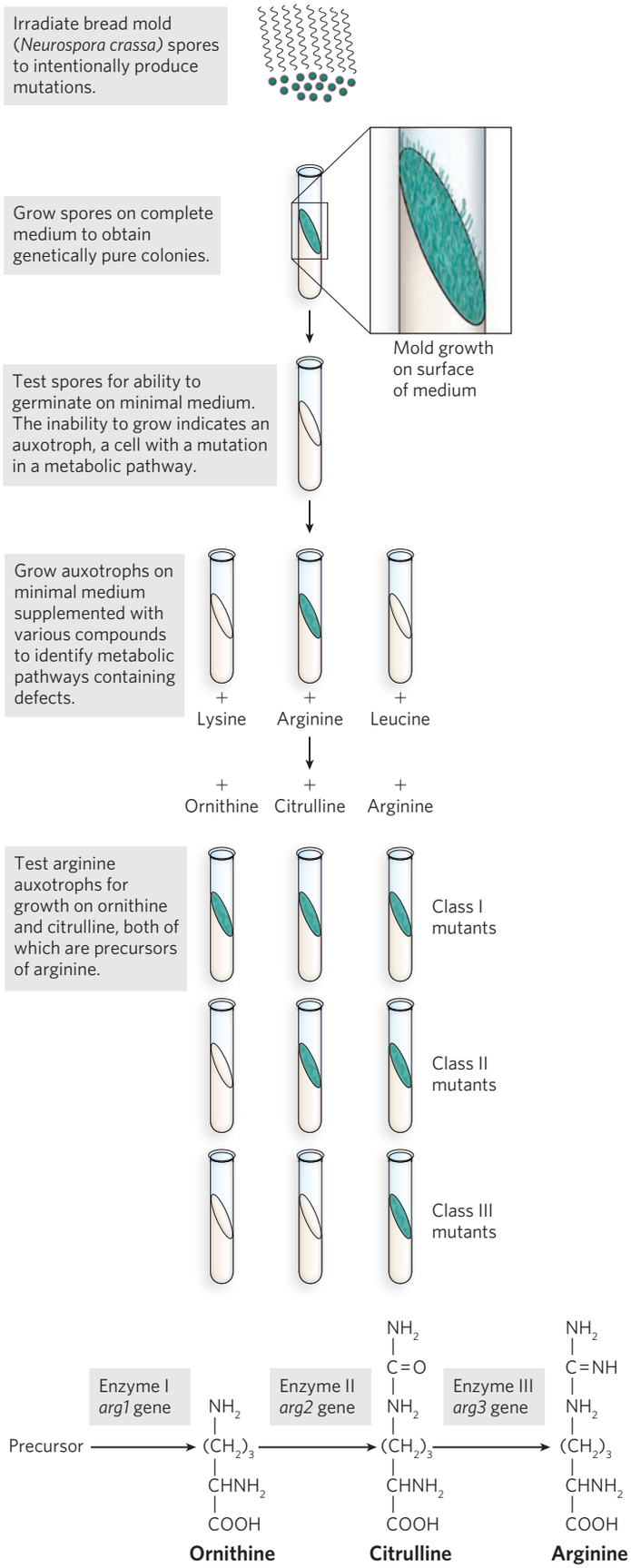


Edward Tatum, 1909–1975
*[Source: American Stock Archives/ Getty Images.]*

An inability to grow on minimal medium indicates a mutation in one of the metabolic pathways required for growth. These mutants are called **auxotrophs**. Spores of different auxotrophs were then analyzed for growth on a range of minimal media supplemented with selected compounds, to identify the defective metabolic pathways and the steps affected. An example of one such study is illustrated in **Figure 2-23**, for auxotrophs of arginine metabolism.

Beadle and Tatum had a collection of *Neurospora* mutants that were auxotrophic for the amino acid arginine. The arginine synthetic pathway was known to include the intermediate compounds ornithine and citrulline, so they tested their arginine auxotrophs for growth on minimal media containing ornithine, citrulline, or arginine. The arginine auxotrophs fell into three classes, depending on which intermediate(s) they required for growth (see Figure 2-23). Beadle and Tatum also mapped the mutant genes and found that mutants in a particular class of auxotrophs mapped to the same chromosomal location. They concluded that each class of mutant was caused by a single defective gene. Their findings also held true for genes in other metabolic pathways.

On the basis of these experiments, Beadle and Tatum proposed the *one gene, one enzyme* hypothesis, which stated that each gene codes for one enzyme. We now know that some enzymes are composed of multiple subunits encoded by different genes; furthermore, not all proteins are enzymes. So, the hypothesis was later revised to *one gene, one polypeptide*. A **polypeptide** is a chain of amino acids, and a functional protein can be composed of a single polypeptide or multiple polypeptide subunits. For a large number of genes, *one gene, one polypeptide* holds true. But as we will see throughout this textbook, even this hypothesis is not entirely accurate. Some genes code for functional RNAs rather than for protein. And through a process called alternative splicing (see Chapter 16), some genes code for more than one type of polypeptide.




## The Central Dogma: Information Flows from DNA to RNA to Protein—Usually

Watson and Crick's determination of DNA structure was a turning point in understanding how information flows in biological systems. Their model of DNA structure, which they reasoned from data collected by other scientists—most notably Rosalind Franklin—consists of two strands of DNA wound about one another in a spiral, double helix. Each strand is composed of a long string of the four nucleotides containing the bases adenine (A), guanine (G), cytosine (C), and thymine (T). The nucleotides in one strand pair with those in the other. Because A pairs only with T, and G pairs only with C, the sequence of each strand contains information about the sequence of the other, and the two strands are said to be **complementary**. The A–T and G–C pairs are referred to as **base pairs**. The detailed structure of DNA and the nucleotide bases, how the nucleotides base-pair in a specific way, and the critical contributions Rosalind Franklin made to Watson and Crick's structure are described in Chapter 6.

The double-helical DNA structure immediately suggested a mechanism for the transmission of genetic information. The essential feature of the model is the complementarity of the two DNA strands. As Watson and Crick realized well before confirmatory data became available, DNA could logically be replicated by separating the two strands and using each as a template to synthesize a new, complementary strand, thereby generating two new DNA duplexes that are identical to each other and to the original double-stranded DNA.

With discovery of the DNA structure, genetics could now be described in chemical terms. Both DNA and proteins are linear polymers, so the sequence of nucleotides in DNA must somehow be converted to a sequence of amino acids. But DNA is located in the nucleus, whereas proteins are synthesized in the cytoplasm. Therefore, there must be an intermediary molecule to shuttle information between

**FIGURE 2-23 "One gene, one polypeptide" analysis of a *Neurospora crassa* auxotroph.** Beadle and Tatum identified mutant *Neurospora* that were unable to synthesize the amino acid arginine. To investigate the metabolic pathway of arginine synthesis, they analyzed arginine auxotrophs for growth on minimal medium plus ornithine or citrulline, both precursors of arginine (or on minimal medium plus arginine, to be sure that the mutant grew when supplied with arginine). They found that class I mutants grow when supplied with any of the three compounds, so these mutants lack an enzyme that is upstream of these three compounds (i.e., an enzyme catalyzing an earlier reaction) in the synthetic pathway. Class II mutants do not grow on ornithine, and thus lack an enzyme downstream of this intermediate but upstream of citrulline. Class III mutants grow only on arginine and therefore lack an enzyme involved in the conversion of citrulline to arginine.

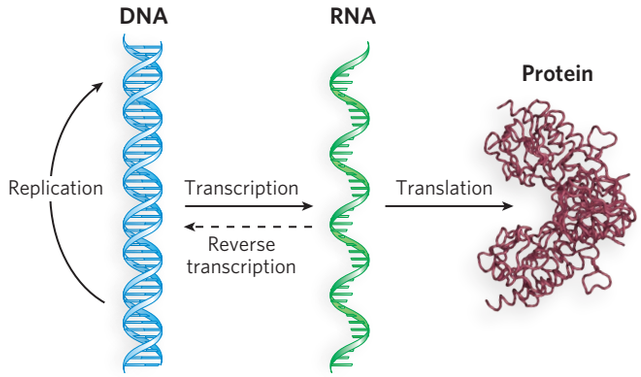



**FIGURE 2-24 Crick's central dogma of information flow: DNA→RNA→protein.** The information to replicate DNA is inherent in its structure (curved arrow). Information flows from DNA to RNA by transcription. Information flows from RNA to protein by translation. In some instances, information can also flow backward, from RNA to DNA (reverse transcription), and some viruses encode enzymes that produce RNA from RNA (not shown in the figure). Other types of information flow exist that do not fall within Crick's central dogma.

the two locations. RNA was believed to play a role in this, and the similarities between DNA and RNA made it a simple matter to understand how an RNA molecule could be made from a DNA template. Crick proposed that biological information flows in the direction DNA→RNA→protein and that DNA acts as a template for its own synthesis (DNA→DNA) (**Figure 2-24**). Crick's proposal is known as the **central dogma** of information flow.

Over the years, it has become obvious that the nice and tidy linear flow of information in Crick's central dogma is not really all that simple after all. Several different paths of information flow are now known to exist. Among these different pathways are the ability of some enzymes to synthesize DNA from RNA (RNA→DNA) and the ability of some viruses to use RNA as a template to make more RNA (RNA→RNA). But the most profound change in what we know about information flow is the finding that the cell makes a huge amount of RNA that is not translated into protein, and it is not just tRNA and rRNA (whose functions we describe below). Indeed, much of the mammalian genome is transcribed into RNA that does not code for protein. We discuss this topic briefly at the end of this section.

RNA was widely expected to be the molecule that mediates the transfer of information from DNA in the nucleus to the site of protein synthesis in the cytoplasm. However, no one imagined that three different types of RNA would be required for the process.

***Ribosomal RNA*** In the early 1950s, Paul Zamecnik and his colleagues identified the site of protein synthesis as particles in the cytoplasm called **ribosomes**. Ribosomes are large structures composed of both protein and RNA. The RNA component is called **ribosomal RNA (rRNA)**. In bacteria and eukaryotes, ribosomes consist of a large subunit and a small subunit.

***Messenger RNA*** The combined findings that ribosomes are the site of protein synthesis and that rRNA is the most abundant RNA (>80%) in the cell led most researchers to believe that rRNA was the carrier of information from DNA to protein. However, some features of rRNA are incompatible with its function as an information carrier. For example, rRNA is an integral part of the ribosome, so there would have to be specific ribosomes to make each specific protein. Further, the nucleotide composition of rRNAs from different organisms was relatively constant, whereas the nucleotide composition of chromosomal DNA varied considerably from one organism to the next.

Studies by Sydney Brenner, Jacques Monod, and Matthew Meselson in the early 1960s, using *Escherichia coli* (see the Model Organisms Appendix), suggested that another type of RNA carries the message from DNA to protein. They discovered a class of RNA that targets preexisting ribosomes, and the nucleotide composition of this RNA was more similar to chromosomal DNA than was rRNA. These properties are exactly those expected for a true messenger between DNA and protein. The investigators called this RNA **messenger RNA (mRNA)** and concluded that ribosomes are protein-synthesizing factories that use mRNA as a template to direct construction of the protein sequence.

RNA synthesis is carried out by the enzyme **RNA polymerase**, which synthesizes RNA by reading one strand of the duplex DNA, pairing RNA bases to the bases in the DNA strand, to synthesize a single-stranded RNA molecule that has a sequence directed by the DNA sequence (**Figure 2-25**). This process of making single-stranded RNA copies of a DNA strand is known as **transcription**.

***Transfer RNA*** The discovery of mRNA was a crucial piece of the information puzzle. But a problem remained: how is a sequence of nucleotides in mRNA converted to a sequence of amino acids in protein? Furthermore, DNA and RNA each consist of only four different nucleotides, whereas proteins have 20 different amino acids. Hence, one must assume the existence of a code that uses combinations of nucleotides to specify amino acids. Combinations of two nucleotides yield only 16 permutations ($4^2$). Combinations of three nucleotides yield 64 permutations ($4^3$), more than enough to specify a code for 20 amino acids.

In 1955, Crick hypothesized the existence of an adaptor molecule, perhaps a small RNA, that could read three nucleotides and also carry amino acids. It was not

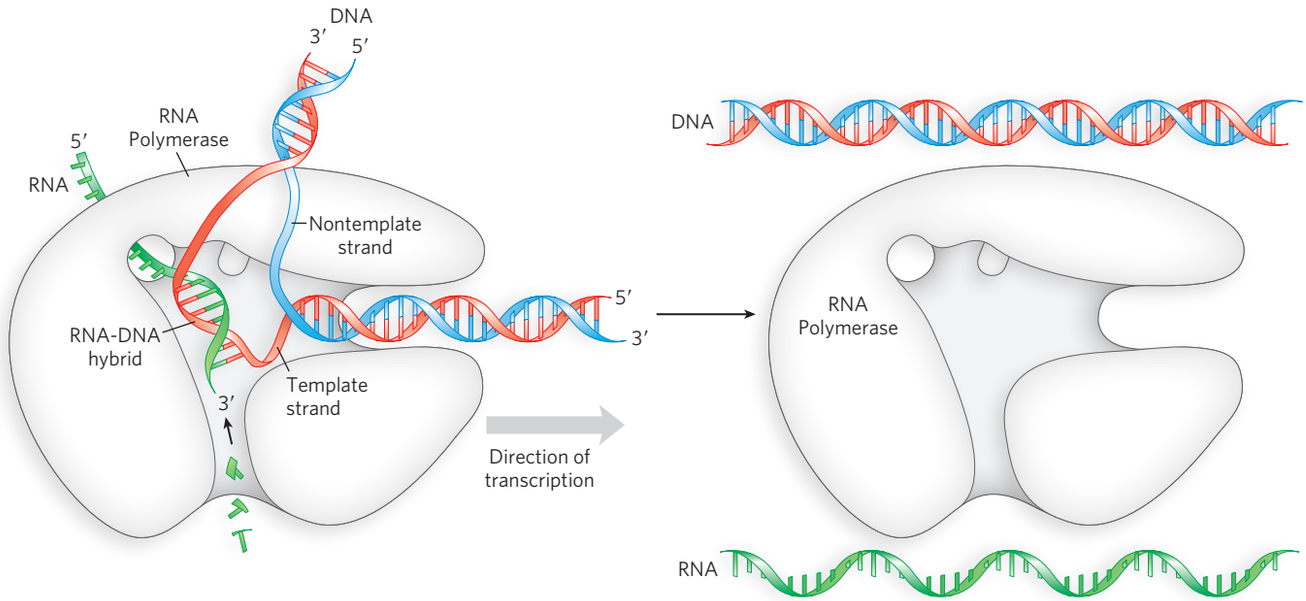



**FIGURE 2-25 The process of transcription.** RNA polymerase opens the DNA duplex and uses one strand as a template for RNA synthesis. The polymerase matches incoming nucleotides to the DNA template strand by base pairing and joins them together to form an RNA chain. As RNA polymerase advances along the template strand, the two DNA strands reassociate behind it to re-form the double helix. When the gene has been completely transcribed, the polymerase dissociates from DNA, releasing the completed RNA transcript. The 3′ and 5′ ends of the DNA are labeled.

long after Crick's adaptor hypothesis (see Chapter 17) that Paul Zamecnik and Mahlon Hoagland discovered a small RNA to which amino acids could attach. This small RNA, later called **transfer RNA (tRNA)**, was the adaptor between nucleic acid and protein.

The discovery of tRNA, combined with the idea of a three-letter code, suggested how the DNA sequence could be converted to an amino acid sequence. Three bases in the tRNA form base pairs with a triplet sequence in the mRNA. When two amino acid–linked tRNAs align side-by-side on the mRNA by base-pairing to adjacent triplets, the amino acids attached to the tRNAs can be joined together. By continuing this process over the length of an mRNA strand, amino acids carried to the mRNA by tRNAs become connected together in a linear order specified by the mRNA sequence. These connections occur as the mRNA-tRNA complexes thread through the ribosome. The overall process of protein synthesis, involving three different types of RNA molecules, is known as **translation** (**Figure 2-26**). (Translation is covered in detail in Chapter 18.)

***Functional RNAs*** All RNAs, whether they code for protein or not, are transcribed from DNA genes. Messenger RNA is needed only transiently, to instruct the synthesis of proteins. But the end products of tRNA and rRNA genes are the RNA molecules themselves. These **functional RNAs** fold into specific three-dimensional shapes and constitute the majority of the RNA in a cell.

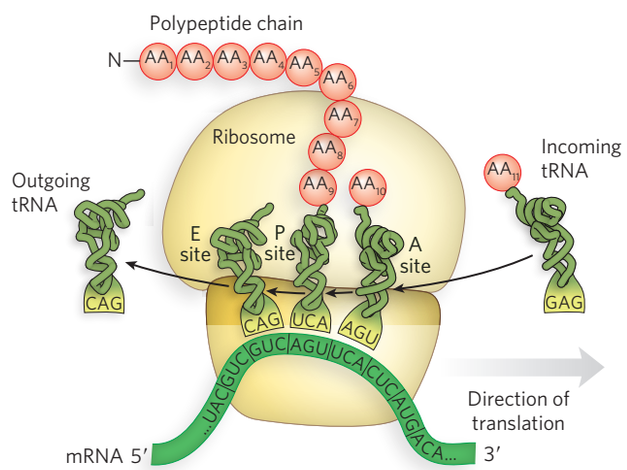



**FIGURE 2-26 The process of translation.** The ribosome, composed of a large and a small "subunit" (each consisting of many proteins and several rRNAs), mediates protein synthesis in cells. It associates with both mRNA and tRNAs as it synthesizes polypeptide chains. The ribosome has three major sites for binding tRNA molecules, the P site, the A site, and the E site. Two tRNAs form base pairs with their respective, adjacent, matching triplets on the mRNA: the tRNA in the P site carries the growing polypeptide chain, and the tRNA in the A site carries an amino acid (AA). The ribosome catalyzes the transfer of the polypeptide attached to the P-site tRNA to the amino acid on the A-site tRNA. The ribosome then shifts relative to mRNA so that the A-site tRNA, now holding the polypeptide, moves into the P site and the tRNA previously at the P site moves to the E site, from which it will depart after the next cycle. The next tRNA carrying an amino acid then binds to the vacated A site to continue extending the polypeptide chain. (N indicates the amino-terminal end of the polypeptide.)

There is an abundance of other functional RNAs besides tRNA and rRNA, some of which have known functions. For example, some small nuclear RNAs (<150 nucleotides) associate with protein to form ribonucleoprotein particles that process the introns from mRNAs (see Chapters 16 and 22). Other types of small RNA, the microRNAs (**miRNA**) and the short interfering RNAs (**siRNA**), have important gene regulatory functions. MicroRNAs anneal to particular mRNAs, usually causing their degradation and thereby effectively turning off, or silencing, the gene (see Chapter 22**)**. Perhaps the most mysterious of the non-protein-coding RNAs are the "long noncoding RNAs." These RNAs (>200 nucleotides) are not translated and have no known function, yet they are more abundant than translated mRNA. There is accumulating evidence that at least some members of this abundant class of RNA may be needed for proper cell function. Identification and understanding of the function of these new RNAs is a fast-paced field, and new types of RNA are almost certain to be discovered in the near future.

There are yet other types of information flow, besides the use of RNAs, that fall outside the classic central dogma. Notable among these is the epigenetic control of gene regulation, based in specific chemical modifications of particular nucleotides and of the proteins that package DNA (in structures called nucleosomes). Combinations of these chemical changes can program the transcriptional control of a cell and can be inherited in cell divisions during an organism's development (see Chapter 10). This *epigenetic inheritance* falls outside the domain of DNA sequence. There are also many other types of protein modifications that transduce the flow of information within the cell and from one cell to another. Suffice it to say that the new dogma is that there is no simple "central dogma." The flow of information is so vital to life and evolution that it takes many forms, some hard to recognize, and scientists have no shortage of work ahead of them to elucidate these important mechanisms.

## Mutations in DNA Give Rise to Phenotypic Change

Most cellular functions are carried out by proteins. The precise sequence of amino acids in each protein molecule and the specific rules governing the timing and quantity of its production are programmed into an organism's DNA. When changes in the DNA sequence occur, cellular function can be altered. Mutations in DNA can be beneficial or harmful to an organism, or can have no effect at all. For example, if the mutation does not change the sequence of a protein or how the protein is regulated, the mutation has no effect and is said to be silent. Evolution depends on mutations that are beneficial, and these usually alter the sequence or regulation of a protein in a way that enhances its function or confers a new, beneficial function that increases the viability of the organism. However, most mutations that change a protein sequence are harmful, because they lead to altered proteins with decreased function or new, detrimental function, and give rise to various diseases. When these DNA mutations occur in germ-line cells (cells that give rise to gametes), the disease can be inherited. There are many examples of inherited diseases, some of which have altered the course of history. One such disease is hemophilia.

Hemophilia afflicted the interrelated royal families of England, Russia, Spain, and Prussia in the 1800s. At the root of this malady is an inability of the blood to clot, resulting in excessive bleeding from even the slightest injury. The disease typically results in death at an early age. Tracing hemophilia through the royal families of Europe indicates that it originated with Queen Victoria (**Figure 2-27a**). It is interesting that none of the current family members are carriers, presumably the result of natural selection against this trait.

Hemophilia is about 10,000 times more common in males than in females. This is because the blood-clotting factor involved in 90% of cases of this disease is factor VIII, encoded by a gene on the X chromosome. A mutant recessive allele of the factor VIII gene is responsible for hemophilia A, the most common form of the disease. Males have only one copy of the X chromosome, and the recessive allele, when inherited, is always expressed. Females have two X chromosomes, and if one X contains a wild-type allele, it masks the expression of the mutant recessive allele. A female with only one copy of the recessive allele is called a carrier, because she is phenotypically normal but may pass on this allele to her offspring (**Figure 2-27b**).

Many other inherited diseases have been mapped to their particular genes. One of the first to be identified was the gene involved in Huntington disease (**Figure 2-28** on p. 51). The gene, *HTT*, is located on chromosome 4. The disease is associated with a region of the *HTT* gene that can have a variable number of repeats of the triplet nucleotide sequence CAG (encoding the amino acid glutamine). The *HTT* gene in healthy individuals has about 27 or fewer of these repeats, but when the number exceeds 36, it is often associated with disease. The likelihood of having Huntington disease increases with the number of trinucleotide repeats in the *HTT* gene. The function of the protein encoded by *HTT* is unknown, but the disease results in the degeneration of neurons in areas of the brain that affect motor coordination, memory, and cognitive function.

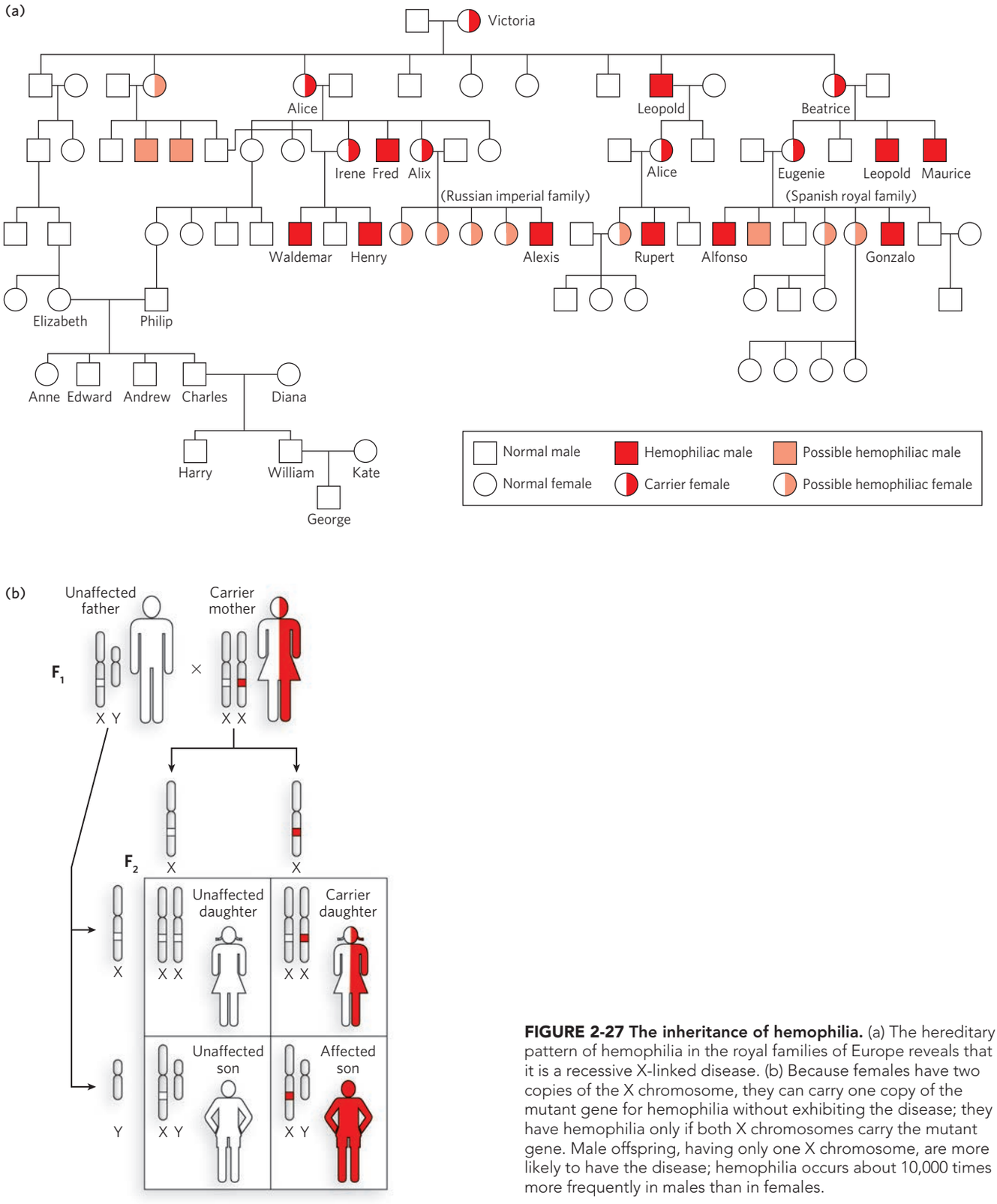
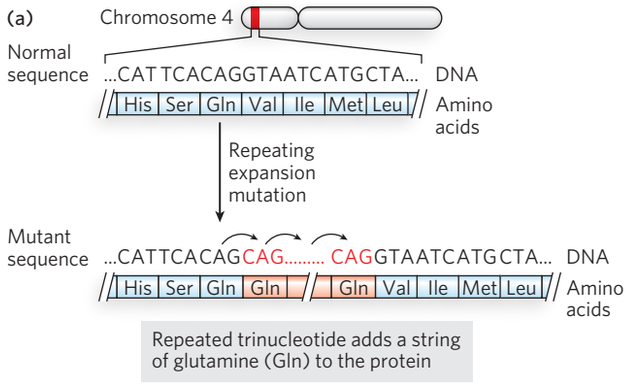



**FIGURE 2-27 The inheritance of hemophilia.** (a) The hereditary pattern of hemophilia in the royal families of Europe reveals that it is a recessive X-linked disease. (b) Because females have two copies of the X chromosome, they can carry one copy of the mutant gene for hemophilia without exhibiting the disease; they have hemophilia only if both X chromosomes carry the mutant gene. Male offspring, having only one X chromosome, are more likely to have the disease; hemophilia occurs about 10,000 times more frequently in males than in females.



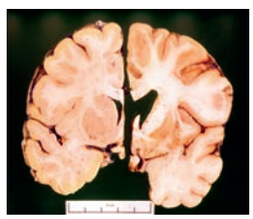
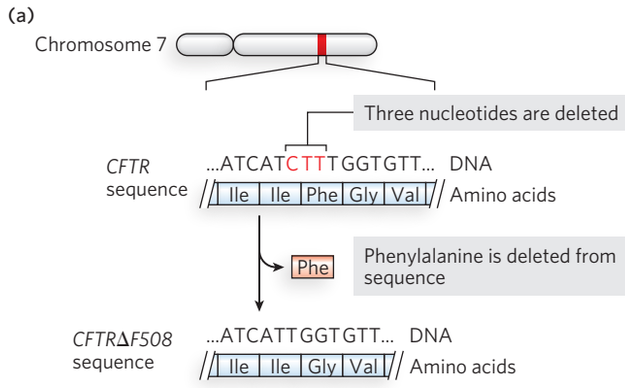



**FIGURE 2-28 Huntington disease.** Huntington disease is an inherited autosomal dominant neurological disease. (a) The gene for Huntington disease (*HTT*) is located on the short arm of chromosome 4, and the disease is associated with CAG repeats (CAG encodes glutamine) in this gene. When the number of CAG repeats increases above 36 copies, disease may occur in midlife. (b) Huntington disease affects the brain by causing degeneration. *[Source: (b) Biophoto Associates/ Science Source.]*



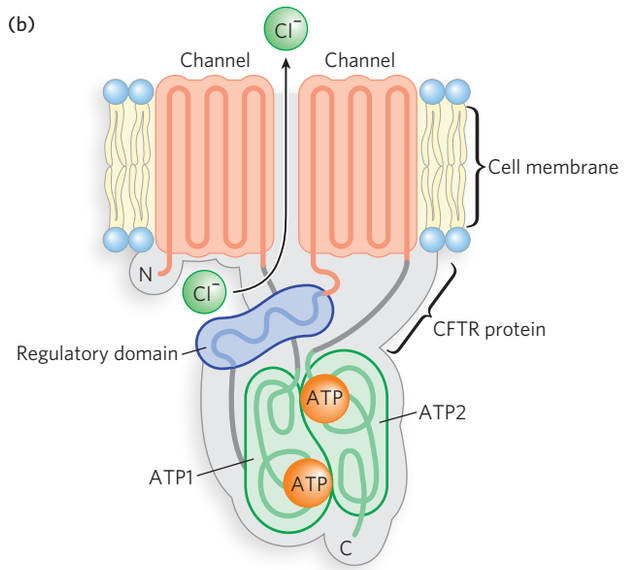




**FIGURE 2-29 Cystic fibrosis.** Cystic fibrosis is caused by a mutation that affects the function of a chloride ion channel. (a) The *CFTR* gene is on chromosome 7. It encodes a channel protein that transports chloride ions. The most common *CFTR* mutation leading to cystic fibrosis is a deletion of three nucleotides that results in the omission of phenylalanine (Phe) at position 508. The isoleucine (Ile) at position 507 remains the same, because both ATC and ATT code for an Ile residue. The omission of $Phe^{508}$ prevents proper protein folding. (b) The chloride ion channel consists of five domains: two domains that form the channel across the cytoplasmic membrane (red), two domains that bind and use ATP as an energy source (ATP1 and ATP2), and a regulatory domain. $Phe^{508}$ is in the ATP1 domain.

The number of triplet repeats in *HTT* can increase during gamete production, resulting in earlier onset and increased severity of the disease over successive generations. This is thought to occur by template slippage (the same segment of DNA replicated more than once) during DNA synthesis due to the repetitive nature of the sequence. Other diseases caused by "triplet expansion" of this type have now been identified. These include Kennedy disease, spinocerebellar ataxia, and Machado-Joseph disease, all caused by an increase in CAG repeats. The CGG repeat is associated with fragile X syndrome, a neurological disorder; expansion of the CTG repeat is associated with myotonic dystrophy, a muscular wasting disease. In these triplet repeat mutations, it is not the sequence of the repeat that matters; rather, it is the disruptive effect of the iterative amino acid they encode within the sequence of the expressed protein that causes the disease.

Cystic fibrosis is another genetic disease that has been identified at the molecular level. The gene (*CFTR*) is on chromosome 7 and encodes a chloride channel protein, the cystic fibrosis transmembrane conductance regulator ($M_r$ 168,173). The protein contains five domains: two domains that span the cytoplasmic membrane for chloride transport; two domains that bind and use ATP, the energy that fuels transport of the chloride ions; and a regulatory domain (**Figure 2-29**). The most common mutation (occurring in about 60% of cases) is *CFTRΔF508*, in which three nucleotides are deleted (denoted by Δ), resulting in the deletion of phenylalanine (denoted by F) at position 508 in the amino acid sequence. This amino acid residue is located in the first of the ATP-binding domains, and its deletion prevents proper folding of the protein. Many other mutations in *CFTR* have also been discovered. *CFTR* mutations are most prevalent in Caucasians from Northern Europe.

**KEY CONVENTION**

The molecular mass of a molecule is commonly expressed in one of two ways, and these are used interchangeably in this book. The first is molecular weight, also called relative molecular mass ($M_r$). The relative molecular mass of a molecule is the ratio of the mass of that molecule to one-twelfth the mass of carbon-12. Because it is a ratio, $M_r$ is dimensionless and has no units. The second common method is molecular mass ($m$), which is the molar mass of the substance divided by Avogadro's number and is expressed in daltons (Da) or kilodaltons (kDa). One dalton is equal to one-twelfth the mass of carbon-12. For example, the molecular mass of a protein that is 1,000 times the mass of carbon-12 can be expressed by $M_r$ = 12,000, $m$ = 12,000 Da, or $m$ = 12 kDa.

The *ΔF508* mutation in *CFTR* is autosomal recessive, and therefore an individual must inherit two copies of the mutant allele to develop cystic fibrosis, one from each parent. Without functional CFTR chloride channels, individuals with cystic fibrosis develop abnormally high sweat and mucus production, and a major complication is the buildup of mucus in the lungs. Patients experience breathing difficulties and often have pneumonia. Individuals with cystic fibrosis have typically had an average life span of about 30 years, but as new treatments are developed, survival is increasing greatly.

Although many mutations are detrimental, other mutations can be beneficial. For example, the protein CCR5 is a coreceptor for HIV, the AIDS virus. There is

**HIGHLIGHT 2-1** **MEDICINE**

## The Molecular Biology of Sickle-Cell Anemia, a Recessive Genetic Disease of Hemoglobin

Genetics, molecular biology, and evolution by natural selection all converge in a striking fashion in sickle-cell anemia, a human hereditary disease. Sickle-cell anemia is a disease of the blood caused by a mutation in the hemoglobin protein. Hemoglobin, the oxygen-carrying protein of red blood cells (erythrocytes), is composed of four subunits, two α chains and two β chains. The sickle-cell mutation occurs in the β chain, and the mutant hemoglobin is called hemoglobin S. Normal hemoglobin is called hemoglobin A. Humans are diploid and thus contain two alleles of the β-chain gene. The two alleles are sometimes slightly different. About 50 genetic variants of hemoglobin are known, usually due to a single amino acid change, and most of these are quite rare. Although the effects on hemoglobin structure and function are often negligible, they can sometimes be extraordinary.

Sickle-cell anemia is a recessive genetic disease in which an individual inherits two copies of the β-chain allele for sickle-cell hemoglobin S (i.e., the sickle-cell allele). A heterozygous individual, with one sickle-cell allele and one normal allele, has nearly normal blood. Two heterozygous parents can potentially have a child who is homozygous for the recessive sickle-cell allele (Figure 1).

The nucleotide sequence of the sickle-cell allele usually contains a thymine (T) in place of an adenine (A), thereby changing one nucleotide triplet from GAG to GTG (Figure 2). This single base change results in hemoglobin S, which contains a hydrophobic (water-fearing) valine residue at one position in the β chain, instead of a hydrophilic (water-loving) glutamic acid residue. This amino acid change causes deoxygenated hemoglobin S molecules to stick together, forming insoluble fibers inside erythrocytes and

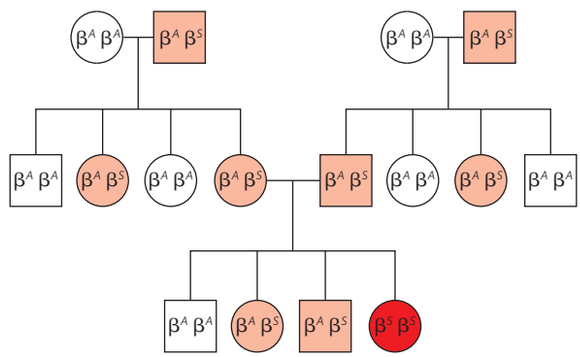



**FIGURE 1** A family pedigree for sickle-cell anemia shows the genotypes for the hemoglobin β chain (circles, females; squares, males). The alleles are $\beta^A$ for wild-type hemoglobin A and $\beta^S$ for sickle-cell hemoglobin S. Heterozygous individuals are shaded in pink, individuals homozygous for $\beta^S$ in red.

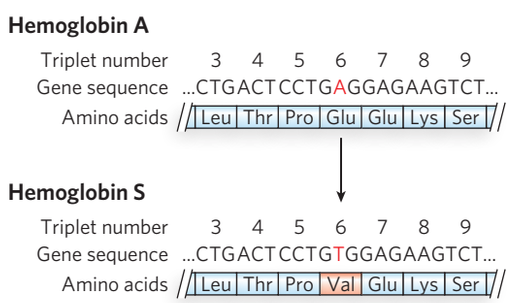



**FIGURE 2** A single nucleotide change in the sickle-cell allele alters the hemoglobin β chain. In hemoglobin A, triplet 6 is GAG, which codes for glutamic acid (Glu). In hemoglobin S, the most common substitution is a T for the A in triplet 6 to form a GTG triplet, which codes for valine (Val).

much speculation about a 32 amino acid deletion mutation of CCR5 (due to the *CCR5Δ32* mutation), which is widely dispersed among people of European descent (an occurrence of 5% to 14% in these groups), although much rarer among Asians and Africans. Researchers speculate that this mutation may have conferred resistance to the bubonic plague or smallpox, thereby becoming enriched in the population, by natural selection, in endemic areas. Although the allele has a negative effect on T-cell (a type of immune cell) function, it seems to provide protection against HIV infection, as well as smallpox.

Another example of a mutation that confers some benefit is the one that causes sickle-cell anemia (**Highlight 2-1**), a mutation of hemoglobin. When the mutation is inherited from both parents, the result is misshapen red blood cells that can get stuck in capillaries and impede blood flow, with possibly fatal results. However, people who are heterozygous for this mutation have enhanced resistance to malaria. Geographic areas where this mutation is prevalent in the population correlate with locations that are plagued by malaria.

Discovery of DNA as the hereditary material, and the understanding of how it is transcribed and translated into RNA and protein, is a most fascinating story in science. Darwin's theory of the origin of species through evolution by natural selection was compelling, but the mechanism that drove the variation on which natural selection could act remained a mystery in his lifetime. Yet, the

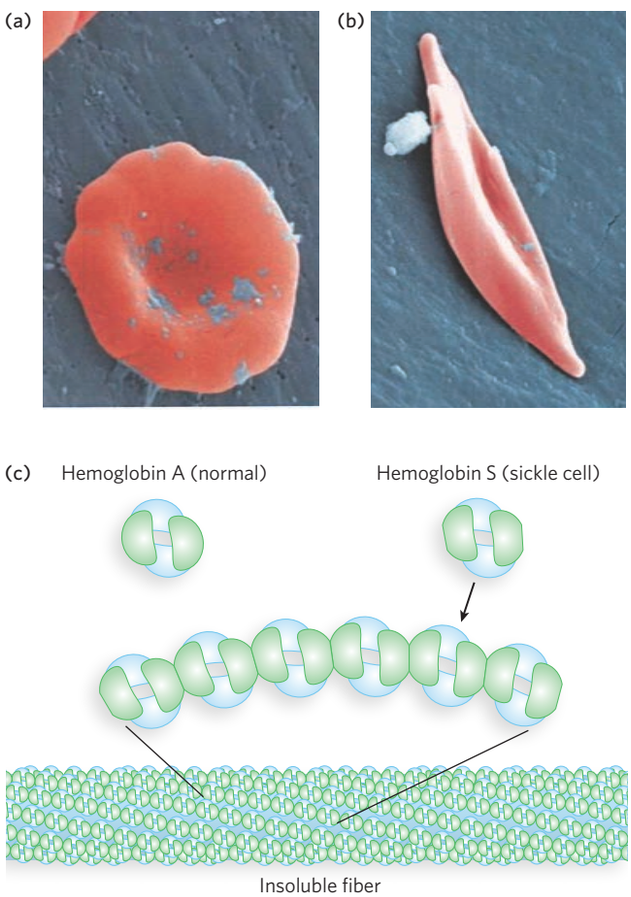



**FIGURE 3** A comparison of (a) a normal, uniform, cup-shaped erythrocyte with (b) a sickle-shaped erythrocyte seen in sickle-cell anemia. (c) The shape change in the hemoglobin molecule, due to the substitution of a Val for a Glu residue in the β chain, allows the molecules to aggregate into insoluble fibers within the erythrocytes. *[Sources: (a) and (b) CDC/ Sickle Cell Foundation of Georgia. Photo by Janice Haney.]*

changing the shape of the cells. The blood of individuals with sickle-cell anemia contains many long, thin, crescent-shaped erythrocytes that look like the blade of a sickle (Figure 3). The sickle shape occurs only in veins, after the blood has become deoxygenated. Sickled cells are fragile and rupture easily, resulting in anemia (from the Greek for "lack of blood").

When capillaries become blocked, the condition is much more serious. Capillary blockage causes pain and interferes with organ function—often the cause of early death. Without medical treatment, people with sickle-cell anemia usually die in childhood. Nevertheless, the sickle-cell allele is surprisingly common in certain parts of Africa. Investigation into the persistence of an allele that is so obviously deleterious in homozygous individuals led to the finding that in heterozygous individuals, the allele confers a small but significant resistance to lethal forms of malaria. Heterozygous individuals experience a milder form of sickle-cell disease called sickle-cell trait; only about 1% of their erythrocytes become sickled on deoxygenation. These individuals can live normal lives by avoiding vigorous exercise and other stresses on the circulatory system. Natural selection has thus resulted in an allele that balances the deleterious effects of the homozygous sickle-cell condition against the resistance to malaria conferred by the heterozygous condition.

key to understanding this mystery had already been discovered by Mendel. Mutations create the natural variation needed for the forces of natural selection to mold new species. It seems almost ludicrous that Mendel and Darwin were alive at the same time and separately uncovered secrets that together explained the diversity of planetary life. Lack of a robust means of communication kept these two vital pieces of information segregated for decades—an improbable situation today, given the rapid pace of global communication. Although most mutations are deleterious, the rare mutation that carries a beneficial change eventually enters the population through natural selection over the expanse of evolutionary time. Natural selection still drives change and the evolution of new species today.

## SECTION 2.4 SUMMARY

- Nucleic acids (DNA and RNA) are composed of repeating units called nucleotides. Each nucleotide contains a phosphate group, a ribose sugar, and a nitrogenous base. Four different bases are found in DNA: adenine, guanine, cytosine, and thymine. RNA also contains adenine (A), guanine (G), and cytosine (C), but uracil (U) instead of thymine (T). Information is encoded by the specificity of pairing of G with C, and of A with T (or U).
- Identification of DNA as the chemical of heredity was determined in experiments using virulent and nonvirulent bacteria. The DNA of virulent bacteria transforms nonvirulent bacteria into a virulent form.
- Even before the DNA structure was solved, studies of mutants drew the connection between genes and enzymes, as in the investigations of defective enzymes in the biosynthetic pathways of auxotrophic mutants of *Neurospora crassa*.
- Information flow in the direction DNA→RNA→ protein is known as the central dogma. RNA is synthesized from a DNA template in the process of transcription. In translation, the RNA sequence is converted to protein. The duplication of DNA is replication. Exceptions to the central dogma exist (RNA→DNA, and RNA→RNA).
- Three types of RNA are required for DNA→ RNA→protein. Ribosomal RNA combines with proteins to form ribosomes, which are factories for protein synthesis. Transfer RNAs are small adaptor RNAs to which amino acids become attached. Messenger RNAs encode proteins and are read by tRNAs in groups of three nucleotides, each of which specifies an amino acid.
- Functional RNAs are RNA sequences that are not translated into protein. Rather, the RNA sequences themselves perform functions in the cell. Both rRNA and tRNA are functional RNAs.
- Mutations are changes in DNA sequence. When a mutation affects the function of a protein or functional RNA, it results in a phenotypic change. Changes in the DNA sequence of germ-line cells underlie inherited human diseases, including hemophilia, Huntington disease, cystic fibrosis, and sickle-cell anemia. Mutations are not always deleterious—sometimes they can be beneficial and, indeed, are vital in creating the diversity needed for the evolution of new species.

# HOW WE KNOW

## Chromosome Pairs Segregate during Gamete Formation in a Way That Mirrors the Mendelian Behavior of Genes

Walter Sutton, 1877–1916
*[Source: University of Kansas Medical Center Archives.]*

*Boveri, T.* 1902. Ueber mehropolige Mitosen als Mittel zur Analyse des Zellkerns. *Verh. Phys. Med. Ges. Wurzburg* 35:67–90.

*Sutton, W. S.* 1902. On the morphology of the chromosome group in *Brachystola magna. Biol. Bull.* 4:24–39.

*Sutton, W. S.* 1903. The chromosomes in heredity. *Biol. Bull.* 4:231–251.

Walter Sutton was only a graduate student when, in 1902, he made observations that led to some of the most profound conclusions in biology. At the time of Sutton's studies at Columbia University, Mendel's laws had just been rediscovered by genetic methods similar to the ones used by Mendel 35 years earlier. But now there were new ways of looking at organisms—namely, observing individual cells under the microscope. Sutton was particularly interested in the process of gamete production, in which one cell undergoes two divisions; in the second division, the chromosome number is halved relative to that of the parent. This process fascinated him. Others who studied these cell divisions used organisms with chromosomes that were too small to allow the observer to discern their individual identity. But Sutton studied the great lubber grasshopper (*Brachystola magna)*, which had large chromosomes with distinctive shapes (Figure 1). This allowed him to see that in meiosis, each chromosome paired with a look-alike partner (a homologous chromosome), and the homologous chromosomes (each a pair of sister chromatids) separated from each other in the first meiotic cell division. During the second cell division, the two sister chromatids of each duplicated chromosome assorted into different daughter cells. On the union of sperm and egg, the homologous pairs of chromosomes were reestablished.

The behavior of chromosomes mimicked the Mendelian behavior of segregation of traits, but on a subcellular level. Sutton hypothesized that paternal and maternal chromosomes exist in pairs and separate into gametes during meiosis, explaining the diploid particles of heredity in Mendel's laws.

Today, a scientist making a groundbreaking discovery of this caliber would have established a solid reputation in science. But in Sutton's day, there were no graduate student stipends or regular sources of scientific funding. So Sutton became a physician and went back to his hometown in Kansas to practice medicine.

Theodor Boveri, a talented German scientist, worked completely independent of Sutton yet reached similar conclusions in the same year as Sutton. Boveri studied the behavior of chromosomes in sea urchin eggs. Although sea urchin chromosomes are small and thus cannot be distinguished by their shape, their number can be observed during fertilization and cell division. Boveri's studies reached the same conclusions as Sutton's, linking chromosomes with the particles of Mendelian inheritance. He also observed that eggs from which the nucleus was removed could be fertilized and then develop into normal—albeit haploid—larvae, and that normal larvae could develop from eggs with only the female set of chromosomes in the nucleus (also haploid). He concluded that each chromosome set, contributed by either parent, had a complete set of instructions for development of the organism. The findings of the two scientists became known as the Sutton-Boveri chromosome theory of inheritance.


**FIGURE 1** Sutton's drawings of chromosomes of the grasshopper *Brachystola magna*, showing their unique shapes and sizes. The pairs of chromosomes are labeled *a* through *k* and *x*. *[Source: W. S. Sutton, Biol. Bull. 4:24–39, 1902.]*

# HOW WE KNOW

## Corn Crosses Uncover the Molecular Mechanism of Crossing Over

*Creighton, H., and B. McClintock.* 1931. A correlation of cytological and genetical crossing-over in *Zea mays*. *Proc. Natl. Acad. Sci. USA* 17:492–497.

Barbara McClintock, 1902–1992 (left); Harriet Creighton, 1909–2004 (right) *[Source: Karl Maramorosch/ Courtesy Cold Spring Harbor Laboratory Archives.]*

Fruit flies have taught us how our body plan is determined. Who would have guessed that fruit flies would teach us so much? Among the many fruitful (pun intended) discoveries by Thomas Hunt Morgan, who developed the fly as a model for genetic study, was the finding that genes cross over between chromosomes. Although researchers presumed that genetic recombination occurred through material exchange between homologous chromosomes, there was no proof that this was indeed the case. Direct proof came in 1931 from a now classic study in corn (maize) by Harriet Creighton and Barbara McClintock.

The insightful experiments of Creighton and McClintock combined genetics and cytologic methods. To visualize crossing over between two homologous chromosomes, one first needs to find two homologous chromosomes that look different—no easy task. Creighton and McClintock searched until they found a plant with an odd-shaped chromosome; in this plant, chromosome 9 had a knob on one end and an extension on the other. Next, they showed that this plant could be crossed with a plant having a normal-shaped chromosome 9 to produce offspring having a homologous chromosome 9 pair that did not look alike. They then mapped two alleles on chromosome 9 to follow recombination genetically. These alleles were seed color—*C* (colored) and *c* (colorless); and seed texture—*Wx* (starchy) and *wx* (waxy). Creighton and McClintock crossed the two plants represented at the top of Figure 2 and looked for colorless, waxy progeny (i.e., progeny that produce colorless, waxy seeds). Genetic crossing over between the misshapen chromosome 9 and its homolog is required to produce a colorless, waxy plant of genotype *ccwxwx*. If genetic crossing over results from physical recombination between the two chromosomes, then the chromosomes of the colorless, waxy progeny should contain chromosome pairs with the misshapen chromosome 9 having only one abnormality—either a knob or an extension at one end (see Figure 2). Indeed, chromosomes of the rare colorless, waxy offspring looked exactly as predicted, confirming that genetic recombination occurs through the physical exchange of material between homologous chromosomes.

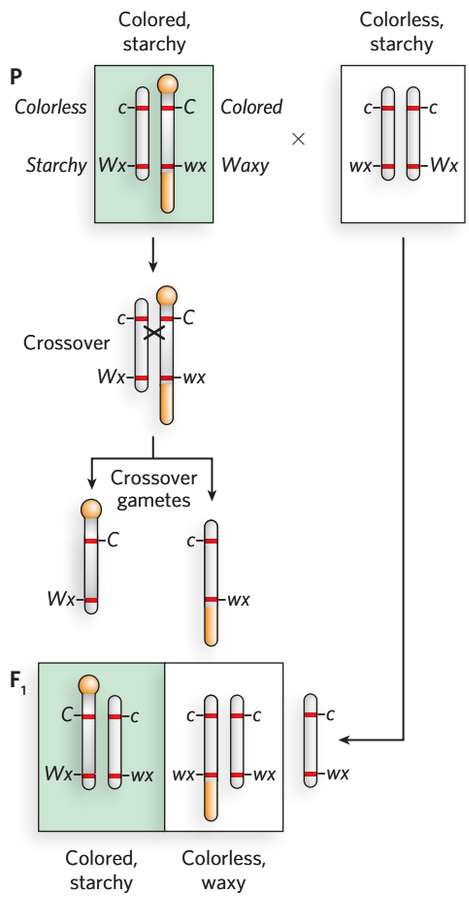



**FIGURE 2** The gametes at the top left represent a corn plant with colored, starchy seeds that is heterozygous for these seed-color and seed-texture genes (*CcWxwx*). One chromosome 9 in the diploid has abnormal extremities. This plant was crossed with a corn plant (top right) having colorless, starchy seeds (*ccWxwx*). Genetic crossing over in the colored, starchy plant produced colorless, waxy progeny of genotype *ccwxwx*. Microscopic examination confirmed that genetic crossing over involves physical recombination of chromosomes: one end of the abnormal chromosome 9 was replaced with a normal end, containing the colorless-seed gene.

## Hershey and Chase Settle the Matter: DNA Is the Genetic Material

*Hershey, A.D. and M. Chase,* 1952. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J. Gen. Physiol.* 36:39–56.

Martha Chase, 1927–2003 (left); Alfred Hershey, 1908–1997 (right) *[Source: Karl Maramorosch/Courtesy Cold Spring Harbor Laboratory Archives.]*

In 1952, Martha Chase and Alfred Hershey performed a now classic experiment, the results of which would convince the world that DNA is the genetic material. They used a bacterial virus, mainly composed of protein and DNA, and set out to determine which of these components carries the hereditary material. Bacteriophage T2, or T2 phage, like other bacterial viruses, consists of a protein coat and a DNA core. Hershey and Chase took advantage of a key chemical difference between these two macromolecules. Using the fact that sulfur is found in proteins but not in DNA, and that phosphorus is found in DNA but not in proteins, they prepared radiolabeled T2 phage using either $^{35}S$ (only protein is radioactively labeled) or $^{32}P$ (only DNA is radioactively labeled). The two T2 phage samples were allowed to attach to their bacterial host, *Escherichia coli*, in two separate flasks (Figure 3). After infection, the bacteria were transferred to a kitchen blender and agitated to strip away any T2 phage material from the outside of the bacterial cell walls. Cells were collected by centrifugation, leaving unattached phage in the supernatant. The results were clear: $^{32}P$-labeled DNA had transferred into the cells, while $^{35}S$-labeled protein remained in the supernatant. Therefore, it is the DNA that carries out the genetic program of the phage. In addition, the progeny phage produced in the infected cells contained $^{32}P$ and no $^{35}S$, further proof that the DNA is the genetic material.

Although earlier experiments by Avery had suggested that DNA was the genetic material, the Hershey and Chase experiment finalized this important conclusion and inspired Watson and Crick in their quest to determine the structure of DNA. Hershey shared the 1969 Nobel Prize in Physiology or Medicine with Max Delbrück and Salvador E. Luria for their discoveries on the replication mechanism of viruses.

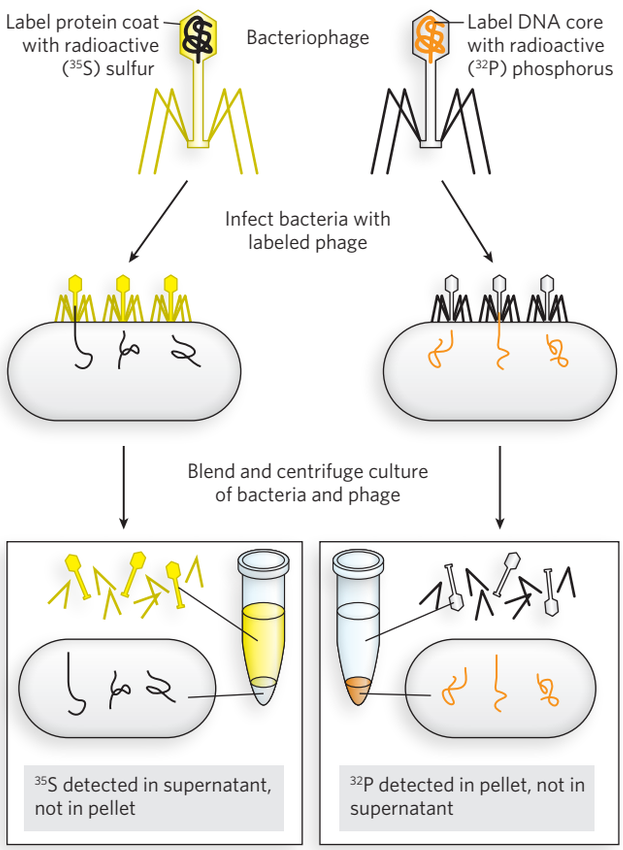



**FIGURE 3** Bacterial cells infected with $^{32}P$-labeled phage contained $^{32}P$ after blender treatment, indicating that viral $^{32}P$-DNA had entered the cells. Cells infected with $^{35}S$-labeled phage had no radioactivity after blender treatment. Progeny virus particles contained $^{32}P$-DNA acquired from the cells infected with $^{32}P$-labeled phage.

## KEY TERMS

gamete cell, 24
somatic cell, 24
gene, 26
diploid, 26
allele, 26
phenotype, 26
haploid, 26
genotype, 27
homozygous, 27
heterozygous, 27
law of segregation, 27
law of independent assortment, 28
chromosome, 32
$G_1$ phase, 33
S phase, 33
centromere, 33
$G_2$ phase, 33
M phase, 33
mitosis, 33
cytokinesis, 35
meiosis, 35
tetrad, 35
crossing over, 40
deoxyribonucleic acid (DNA), 43
ribonucleic acid (RNA), 43
ribosome, 47
RNA polymerase, 47
transcription, 47
translation, 48

## PROBLEMS

**1.** Two purebred pea plants are crossed. One strain has dominant round seeds; the other has recessive wrinkled seeds. (a) What phenotypes would be seen in the $F_1$ generation plants, and in what proportions? (b) What phenotypes would be seen in the $F_2$ generation plants, and in what proportions? (c) If an $F_1$ generation plant is crossed with a plant producing wrinkled seeds, what phenotypes are seen in the progeny, and in what proportions?

**2.** Two pea plants with round seeds are crossed. In the $F_1$ generation, all the plants have round seeds. What can you say about the genotype of the parental plants?

**3.** The $F_1$ plants from the cross in Problem 2 are next crossed at random. There are 129 plants in the $F_2$ generation. The majority, 121 plants, produce round seeds. However, there are 8 plants that produce wrinkled seeds. From this information, what were the genotypes of the original parental plants?

**4.** Purebred white-eyed male fruit flies are crossed with wild-type red-eyed females. If the progeny are crossed with each other repeatedly, which generation will be the first to contain white-eyed female flies?

**5.** Purebred wild-type male flies are crossed with purebred white-eyed female flies. If the progeny are crossed with each other repeatedly, which generation will be the first to contain white-eyed male flies?

**6.** A new species of fruit fly is found on an uncharted island. The flies are brightly colored, with blue and green bodies. After studying these insects for a year or two, researchers find one male with an all-black body. When this male is crossed with wild-type females, all of the male progeny in the $F_1$ generation are black, and all of the female progeny have the blue and green coloring. This same pattern (all black males and colored females) is repeated in the $F_2$, $F_3$, and $F_4$ generations. Explain these observations.

**7.** Two purebred flowering plants are crossed. One has red flowers and small leaves (*RRll*) and the other has white flowers and large leaves (*rrLL*). Using a Punnett square analysis, and assuming that the genes are unlinked, predict the type and frequency of phenotypes in the $F_2$ generation.

**8.** If the $F_1$ plants in Problem 7 had genes for red flower color that exhibited incomplete dominance, the heterozygous *Rr* flower color would be pink. In that instance, what percentage of the $F_1$ plants in Problem 7 would have pink flowers? What percentage of the $F_2$ generation would have pink flowers?

**9.** Two purebred fruit flies are crossed. The male has white eyes and vestigial wings. The female has red eyes and normal wings. All $F_1$ flies have red eyes and vestigial wings. Using a Punnett square analysis, predict the percentage of $F_2$ generation males that will have red eyes and normal wings. Assume that the wing trait is not sex-linked.

**10.** A new and exotic species of fly is found, with green eyes (*G*) and striped wings (*S*). A mutant fly of the same species is found that has orange eyes (*g*) and clear (unstriped) wings (*s*). The mutant is cultured for many generations to obtain a purebred strain with the double-mutant phenotype. A *ggss* female fly is mated with a wild-type *GGSS* male. The $F_1$ progeny all have green eyes and striped wings, as expected. An $F_1$ male is mated with an $F_1$ female. Among the $F_2$ progeny of this cross, only two kinds of flies are observed: 75% with green eyes and striped wings, and 25% with orange eyes and clear wings. Some expected $F_2$ progeny (such as flies with green eyes and clear wings) are absent. Explain this result.

**11.** Both meiosis and mitosis are initiated with a complete replication of the cell's chromosomes in S phase. The replication of each chromosome produces a pair of sister chromatids. During the cell division immediately following replication, how are the chromosomes in the sister chromatid pairs distributed to daughter cells in mitosis and meiosis?

**12.** Two purebred plants, with genotypes *AABBCCDDEEFF* and *aabbccddeeff*, are crossed. In the $F_1$ generation, all individuals are heterozygous for all traits. Geneticists probe the linkage of these various genes by doing a series of crosses, examining two traits in each cross. When all the

crosses are done and the data are tabulated, the researchers find that in the $F_1$ plants, meiosis produces gametes that contain the following combinations of alleles at the indicated frequencies (which correspond to crossover frequencies):

| | |
|---|---|
| $A + b$ | 9% |
| $A + e$ | 13% |
| $A + d$ | 50% |
| $B + c$ | 6% |
| $C + f$ | 50% |
| $C + e$ | 10% |
| $D + f$ | 16% |

With these data, determine how the genes are distributed along the chromosomes. Draw a map, using the crossover frequencies as distances.

**13.** On one chromosome there are three linked genes designated *M*, *N*, and *O*. If crossing over occurs between *M* and *O* 5% of the time, and between *N* and *O* 8% of the time, what are the possible arrangements of the genes on the chromosome?

**14.** In the central dogma developed by Francis Crick and others, three kinds of RNA play important roles: rRNAs, tRNAs, and mRNAs. Describe two features that are characteristic of each type of RNA.

**15.** In the classic Hershey-Chase experiment (see the How We Know section for this chapter), the T2 phage was labeled with either $^{35}S$ or $^{32}P$ before using it to infect a bacterial host. In this experiment, would it have been possible for these researchers to label one batch of T2 phage with both $^{35}S$ and $^{32}P$ and still get a definitive result? Why or why not? What would the results of the experiment be?

**16.** In the bacterium *Escherichia coli*, the amino acid tryptophan is synthesized in a multistep pathway, beginning with an organic precursor called chorismate:

$$\text{Chorismate} \xrightarrow{1} \text{X} \xrightarrow{2} \text{Y} \xrightarrow{3} \text{Tryptophan}$$

The numbers denote steps in the pathway, each catalyzed by an enzyme. There are five bacterial genes that encode polypeptides associated with these enzymes, called *trpA*, *trpB*, *trpC*, and so on. A researcher isolates a series of mutations that eliminate the capacity of the cells to synthesize tryptophan, each mutation affecting one of the five genes. Each mutant cell is tested for its ability to grow on a medium containing either tryptophan, chorismate, or intermediate X or Y. The following results are obtained (+ means growth; − means absence of growth; WT means wild type):

| *Molecule in growth medium* | *trpA* | *trpB* | *trpC* | *trpD* | *trpE* | *WT* |
|---|---|---|---|---|---|---|
| Tryptophan | + | + | + | + | + | + |
| X | − | − | − | − | − | + |
| Y | − | − | − | + | + | + |
| Chorismate | + | + | + | + | + | + |

Which genes encode the enzymes involved in steps 1, 2, and 3 of the tryptophan biosynthetic pathway? Given this information, what can you say about the structure of these enzymes?

**17.** In mitosis and meiosis, all cellular chromosomes are replicated before cell division. In a diploid cell, there are two copies of each autosomal chromosome. For a typical diploid eukaryotic cell and a hypothetical chromosome A, the two copies of the chromosome can be labeled A1 and A2. During replication, each chromosome (including A1 and A2) is converted into two new chromosome copies that are transiently held together. The immediate tethered products of a replicated chromosome are called sister chromatids. The two sets of sister chromatids resulting from replication of chromosomes A1 and A2 can be labeled A1/A1* and A2/A2*. Thus, in all, four chromosomes are held in two pairs of sister chromatids. For both mitosis and meiosis, replication is eventually followed by a cell division (the first of two cell divisions in the case of meiosis). In that cell division, how are the four chromosomes segregated into the daughter cells in mitosis? How are the four chromosomes segregated into daughter cells in the first cell division of meiosis?

**18.** Food for thought: As described in the chapter, substantial segments of the genome of many organisms are transcribed into RNA that does not encode protein and does not correspond to either rRNA or tRNA. Without studying several additional chapters in this book, you may not be familiar with this RNA. However, considering some potential functions for this RNA is a useful prelude to your continued study of molecular biology. Based on the discussion of the RNA world hypothesis in Chapter 1 and other knowledge you may have about RNA, suggest one or two potential functions for this non-protein-coding RNA.

## ADDITIONAL READING

Many of the books and papers listed here are available online, free of charge, from Electronic Scholarly Publishing, a collection of source material on the foundations of classical genetics (www.esp.org/foundations/genetics/classical).

### *General*

**Dolan DNA Learning Center, Cold Spring Harbor Laboratory.** DNA Interactive: Timeline. www.dnai.org/timeline/index.html. This website offers an overview of the major advances in molecular genetics from 1865 to 2000.

**Watson, J.D. 1968.** *The Double Helix: A Personal Account of the Discovery of the Structure of DNA*. New York: Atheneum Publishers. Watson's original account of his adventures; a quick and very interesting read, warts and all.

### *Mendelian Genetics*

**Bateson, W. 1909.** *Mendel's Principles of Heredity*. Cambridge: Cambridge University Press.

**Mendel, G. 1866.** Versuche über Pflanzen-Hybriden. In *Verhandlungen des naturforschenden Vereines* (*Proceedings of the Natural History Society*). Brünn. Fewer than 150 copies were produced; Darwin owned one of them, but the evidence from examining his copy indicates that he did not open it to Mendel's work.

**Mendel Museum, Masaryk University.** Gregor Mendel. www.mendel-museum.com. This original location of Mendel's work is undergoing extensive restoration as a museum.

### *Cytogenetics: Chromosome Movements during Mitosis and Meiosis*

**Hooke, R. 1664.** *Micrographia: Some Physiological Descriptions of Minute Bodies Made by Magnifying Glasses with Observations and Inquiries Thereupon.* London. Available at www.gutenberg.org/etext/15491.

**Sutton, W.S. 1903.** The chromosomes in heredity. *Biol. Bull.* 4:231–251. An outline of the rationale behind Sutton's proposal that chromosomes carry the material of heredity.

### *The Chromosome Theory of Inheritance*

**Creighton, H.B., and B. McClintock. 1931.** A correlation of cytological and genetical crossing-over in *Zea mays*. *Proc. Natl. Acad. Sci. USA* 17:492–497.

**Morgan, T.H., A.H. Sturtevant, H.J. Muller, and C.B. Bridges. 1915.** *Mechanism of Mendelian Heredity*. New York: Henry Holt and Company. This book contains the classic work from Morgan's lab and illustrates the way in which scientific discoveries were published before scientific journals became the norm.

### *Foundations of Molecular Genetics*

**Avery, O.T., C.M. MacLeod, and M. McCarty. 1944.** Studies on the chemical nature of the substance inducing transformation of pneumococcal types. *J. Exp. Med.* 79:137–158.

**Beadle, G.W., and E.L. Tatum. 1941.** Genetic control of biochemical reactions in *Neurospora*. *Proc. Natl. Acad. Sci. USA* 27:499–506.

**Crick, F. 1970.** Central dogma of molecular biology. *Nature* 227:561–563. The classic paper in which Crick proposes the central dogma of information flow in biology.

**Hershey, A.D. and M. Chase. 1952.** Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J. Gen. Physiol.* 36:39–56.