



An albino giraffe with three normally pigmented companions.

Definitions of the Gene

CHAPTER OUTLINE

Evolution of the Concept of the Gene:

Summary

Evolution of the Concept of the Gene:

Function

Evolution of the Concept of the Gene:

Structure

A Genetic Definition of the Gene

Complex Gene-Protein Relationships

Sir Archibald Garrod and Human Inborn Errors of Metabolism

In 1902, just two years after the discovery of Mendel's work, Sir Archibald E. Garrod, a physician at the Hospital for Sick Children in London, published a paper entitled "The Incidence of Alkaptonuria: A Study in Chemical Individuality." Like Mendel's work, the concepts Garrod presented in this paper remained largely unknown by the scientists of the world for 40 years—until the concepts were independently formulated by George W. Beadle and Edward L. Tatum in 1941. Garrod's paper described the results of his studies of individuals with alkaptonuria, an innocuous, but easily detected, disorder. Because of the presence of a chemical called homogentisic acid (formerly called alkapton), the urine of affected individuals turns black when exposed to air. Garrod clearly recognized that alkaptonuria was inherited. In his 1902 paper, he wrote: "There are good reasons for thinking that alkaptonuria is not the manifestation of a disease but is rather the nature of an alternative course of metabolism, harmless and usually congenital and lifelong" (*Lancet* ii:1616).

In addition to alkaptonuria, Garrod studied albinism, cystinuria, pentosuria, and porphyria in humans (See Human Genetics Sidelight: Human Inborn Errors of Metabolism). Garrod summarized the results of these studies in a treatise entitled "Inborn Errors of Metabolism," first presented as the Croonian Lectures to the Royal College of Physicians in London in 1908 and published in book form in 1909. In this book, Garrod clearly articulates the view that the metabolic defects observed in individuals with these disorders are caused by recessive mutant genes. Garrod probably did not develop his amazing insight into metabolic processes totally by himself. His father, Alfred B. Garrod, was also a physician

and was the first to demonstrate the accumulation of the chemical uric acid in patients with gout, a painful inflammation of the joints in the hands and feet. In any case, Archibald E. Garrod was the first scientist to relate defects in genes to blocks in metabolic pathways. In this chapter, we focus on how Garrod's concept evolved into the current understanding of this basic unit of genetic information: the **GENE**.

The gene is to genetics what the atom is to chemistry. Thus throughout this text we have focused our attention on the gene and the alternate forms of a gene or alleles. In the preceding chapters, we have examined the patterns of transmission of independently assorting and linked genes, the chromosomal location of genes, the chemical composition of genes and chromosomes, the mechanism of replication of genes, mutational events in genes, and the mechanisms by which genes exert their effects on the phenotype of the organism. What is this unit of genetic information that we call the gene? As we will see, the concept of a gene is not static; it has evolved through several phases since Wilhelm Johannsen introduced the term in 1909, and it will undoubtedly evolve through additional refinements in the future.

The gene has been defined as the unit of genetic information that controls a specific aspect of the phenotype. Such a description, though accurate, does not provide a precise, unambiguous definition that can be used to identify a gene at the molecular level. At a more fundamental level, the gene has been defined as the unit of genetic information that specifies the synthesis of one polypeptide. However, it is not a very good operational definition. An **operational definition** spells out an operation or experiment that can be carried out to define or delimit something. One gene specifying one polypeptide is a poor operational definition because experimentally relating all the segments of DNA that represent genes with all the polypeptides is not feasible in a complex organism. Moreover, it is preferable to define the gene by using genetic approaches rather than biochemical experiments. Thus, in this chapter, we focus on the **complementation test** as an operational definition of the gene. We will also consider the limitations of the complementation test, along with the unique structural features of selected genes.

EVOLUTION OF THE CONCEPT OF THE GENE: SUMMARY

Before discussing evidence supporting the various concepts, let's summarize the important stages in the evolution of the gene concept. The gene theory of inheritance began with the publication of Mendel's classic paper in 1866 but did not become an accepted part

of scientific knowledge until after the discovery of Mendel's work in 1900. Mendel's gene (not so-named) was the "character" or "constant factor" that controlled one specific phenotypic trait such as flower color in peas. At the time of the discovery of Mendel's work, the English physician Sir Archibald E. Garrod was studying several inherited diseases in humans. Garrod first recognized that homozygosity for recessive mutant alleles can cause defects in the normal processes of metabolism. His concept of the gene is probably stated most accurately as one mutant gene-one metabolic block, which over 30 years later was refined to the one gene-one enzyme concept enunciated by George W. Beadle and Edward L. Tatum. Since many enzymes contain two or more different polypeptides, each encoded by a separate gene, the one gene-one enzyme concept subsequently was modified to one gene-one polypeptide.

Prior to 1940, genes were considered analogous to beads on a string; recombination occurred between, but not within, genes. The gene was both the basic functional unit, which controlled one phenotypic trait, and the elementary structural unit, which could not be subdivided by recombination or mutation. Clarence Oliver's 1940 report that recombination had occurred within the *lozenge* gene of *Drosophila* stimulated both excitement and much debate about its significance. When the debate ended, the nucleotide pair had replaced the gene as the basic unit of structure, the unit of genetic material not subdivisible by mutation or recombination.

In the early 1940s, Edward B. Lewis developed the complementation, or *cis-trans*, test for functional al-

HUMAN GENETICS SIDELIGHT

Human Inborn Errors of Metabolism

As discussed in the text, Garrod's concept of one mutant gene—one metabolic block was based on his studies of a few inherited human disorders. Alkaptonuria, on which much of Garrod's information was based, is described in the text. However, Garrod also studied familial cases of cystinuria, pentosuria, porphyrinuria—inherited disorders characterized by elevated levels of the amino acid cystine, five-carbon sugars, and the iron-binding porphyrin component of hemoglobin, respectively, in urine. The most common disorder that Garrod studied was albinism, an autosomal recessive trait that occurs at a frequency of about 1 in 20,000 newborns in the United States. The red eyes and white skin and hair of albinos result from the absence of the black pigment melanin. The most common forms of albinism are caused by mutations that block the biosynthesis of melanin from the amino acid tyrosine. Of course, the pathway of melanin biosynthesis was unknown in 1909 when Garrod published his book *Inborn Errors of Metabolism*. Nevertheless, he clearly understood that mutations caused specific blocks in metabolic pathways.

Today, over 4000 inherited human disorders have been described, and the number of such hereditary abnormalities is constantly increasing. These disorders range from relatively innocuous disorders like alkaptonuria to those such as Tay-Sachs disease (characterized by rapid degeneration of the central nervous system) that are lethal in early childhood. Sickle-cell anemia and phenylketonuria (PKU) are perhaps the best known of the human inborn errors of metabolism. The severe anemia in individuals homozygous for the sickle-cell mutation is caused by a single amino acid substitution in their β -globin (Chapter 13). Children with PKU lack the enzyme phenylalanine hydroxylase. This enzyme deficiency results in the accumulation of phenylpyruvic acid, which is highly toxic to the central nervous system. If untreated, children with PKU develop severe mental retardation. However, if they are placed on a diet low in phenylalanine, they develop normal mental abilities (Chapter 20). Other inherited human diseases are discussed throughout the book.

In the United States alone, over 120,000 children with inherited defects are born each year. Long-term medical care

of these birth defects is estimated to cost \$10 to \$20 billion. But the cost is even higher in terms of human suffering. For example, consider the tragic degeneration of the central nervous system that occurs in Huntington disease or the loss of body control resulting from progressive muscle degeneration associated with Duchenne muscular dystrophy (both discussed in Chapter 20) or the consequences of the loss of memory that occurs in individuals with Alzheimer's disease (Chapter 26). What are the causes of these inherited diseases? What can humans do, if anything, to keep the frequency of these disorders as low as possible? How many different human inborn errors of metabolism are there? Can the deleterious symptoms of these inherited disorders be eliminated by treatment? Can any of these inherited diseases be cured so that they won't be transmitted from parents to their children?

Human inherited diseases are caused by all of the types of mutations discussed in Chapter 13 and by changes in chromosome structure and number as discussed in Chapter 6. Because mutations are required for evolution, they can never be completely eliminated. However, humans can minimize the frequency of new mutations by avoiding contact with highly mutagenic agents such as irradiation and chemicals that damage DNA. All genes can mutate to nonfunctional states. Therefore, the number of human inborn errors of metabolism is almost certainly equal to the number of genes required to develop and maintain a normal human phenotype. Several inherited disorders such as PKU have proven treatable once their molecular bases are known. Others are currently being treated by somatic cell gene therapy (Chapter 20). In the future, some inherited human diseases may be cured by germ-line gene therapy; but serious ethical issues must be addressed before germ-line gene therapy can be performed on humans. Given the rapidly increasing number of known inherited human diseases and the identification and characterization of the genes responsible for these disorders (Chapter 20), the tools of molecular genetics will play an increasingly important role in the practice of medicine in the future.

lelism in *Drosophila*. This test subsequently was exploited by Seymour Benzer to define experimentally the gene in bacteriophage T4. The complementation test provides an operational definition of the gene. This test allows a geneticist to determine whether two independent mutations occurred in the same gene or in two different genes. Moreover, the gene as defined by the complementation test is a perfect fit to the one gene—one polypeptide concept.

In the 1960s, elegant experiments by Charles Yanofsky, Sydney Brenner, and their collaborators showed that the gene and its polypeptide product were colinear structures, with a direct correlation between the sequence of nucleotide pairs in the gene and the sequence of amino acids in the polypeptide. However, this simple concept of the gene as a continuous sequence of nucleotide pairs specifying a colinear sequence of amino acids in the polypeptide gene prod-

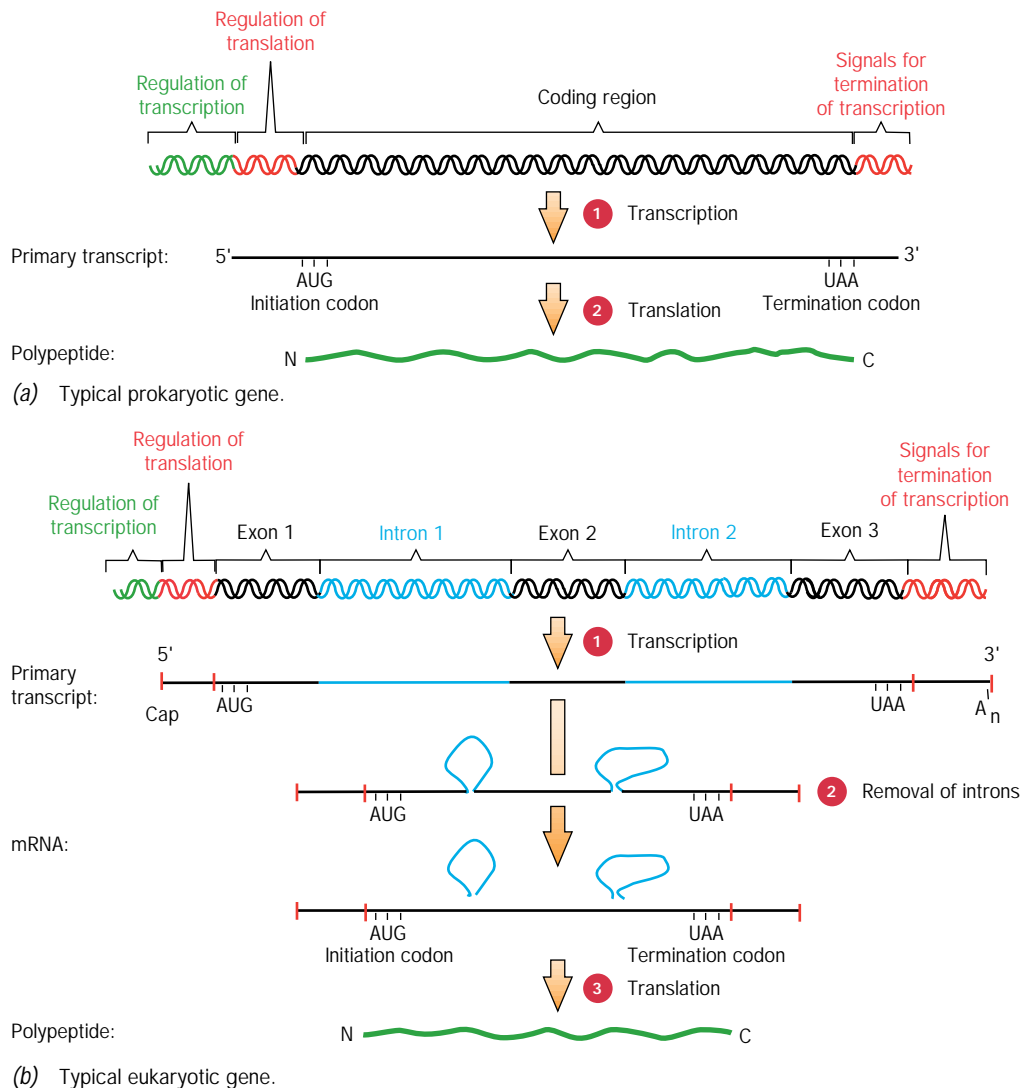


Figure 14.1 Gene structure. Prokaryotic genes usually contain uninterrupted coding sequences (a), whereas the coding sequences of eukaryotic genes are commonly interrupted by noncoding sequences (b).

uct was short-lived. Overlapping genes and genes-within-genes were discovered in the late 1960s, and the coding sequences of eukaryotic genes were shown to be interrupted by noncoding intron sequences in the late 1970s. Moreover, some genes, for example, genes encoding immunoglobulins, were shown to be stored in germ-line chromosomes as short “gene segments,” which are assembled into mature, functional genes during development.

Thus the definition of the gene needs to remain somewhat pliable if it is to encompass all of the different structure/function relationships that occur in different organisms. Here, we define the gene as the unit of genetic information that controls the synthesis of one polypeptide or one structural RNA molecule. As just defined, the gene can be identified operationally by the complementation test. As such, the gene includes the 5' and 3' noncoding regions that are involved in regulating the transcription and translation of the gene and all noncoding sequences or introns

within the gene (Figure 14.1). The structural gene refers to the portion that is transcribed to produce the RNA product. In the case of overlapping genes, this definition requires that some nucleotide-pair sequences be considered components of two or more genes. For those cases where exons are spliced together in various combinations to make related but different proteins, the gene may be defined as a DNA sequence that is a single unit of transcription and encodes a set of closely related polypeptides, sometimes called “protein isoforms.” In germ-line chromosomes, the DNA sequences that encode segments of antibody chains probably should be called “gene segments,” because this genetic information is not organized into units that fit any of the standard definitions of the gene (Chapter 24).

Key Points: The concept of the gene has undergone many refinements since its discovery by Mendel in

1866. Most genes encode one polypeptide and can be operationally defined by the complementation test.

EVOLUTION OF THE CONCEPT OF THE GENE: FUNCTION

In the preceding section, we briefly examined the evolution of the concept of the gene, the basic unit of genetic information. That summary included both functional and structural aspects of the gene. Now, let's take a closer look at the gene as a unit of function.

Mendel: Constant Factors Controlling Phenotypic Traits

The law of combination of different characters, which governs the development of the hybrids, finds therefore its foundation and explanation in the principle enunciated, that the hybrids produce egg cells and pollen cells which in equal numbers represent all constant forms which result from the combinations of the characters brought together in fertilisation. (Mendel, 1866; translation by William Bateson)

Mendel's characters or factors, which are now called genes, controlled specific phenotypic traits such as flower color, seed color, and seed shape. They were the basic units of function, the units of genetic information that governed one specific aspect of the phenotype. This definition of the gene as the basic unit of function is almost universally favored by present-day scientists. There has been no change in the concept of

the gene as the basic unit of function since its discovery by Mendel in 1866. However, the discovery of the chemical nature of the genetic material raised questions about the structure of the gene, and the concept of the molecular structure of this basic unit of function has undergone several changes and refinements since the discovery of Mendel's work in 1900.

If we examine what is known about how genes control phenotypic traits, the need for a more precise definition of the gene will be obvious. The pathway by which a gene exerts its effect on the phenotype of an organism is often very complex (Figure 14.2). Several genes may have similar effects on the same phenotypic trait, making it difficult to sort out the effects of individual genes. All the genes of an organism are located in the same nuclei, and they do not all function independently. The phenotype of an organism is the product of the action of all the genes acting within the restrictions imposed by the environment. Each gene also has an effect on the population to which the organism carrying the gene belongs. Ultimately, each gene has a potential effect, small though it may be, on the cumulative phenotype of the biosphere, for each gene may affect the ability of the organism, or the population, or the species to compete for an ecological niche in the biosphere (Chapter 28).

Garrod: One Mutant Gene—One Metabolic Block

At the time of the discovery of Mendel's work in 1900, Sir Archibald Garrod was studying several congenital metabolic diseases in humans. One of these was the

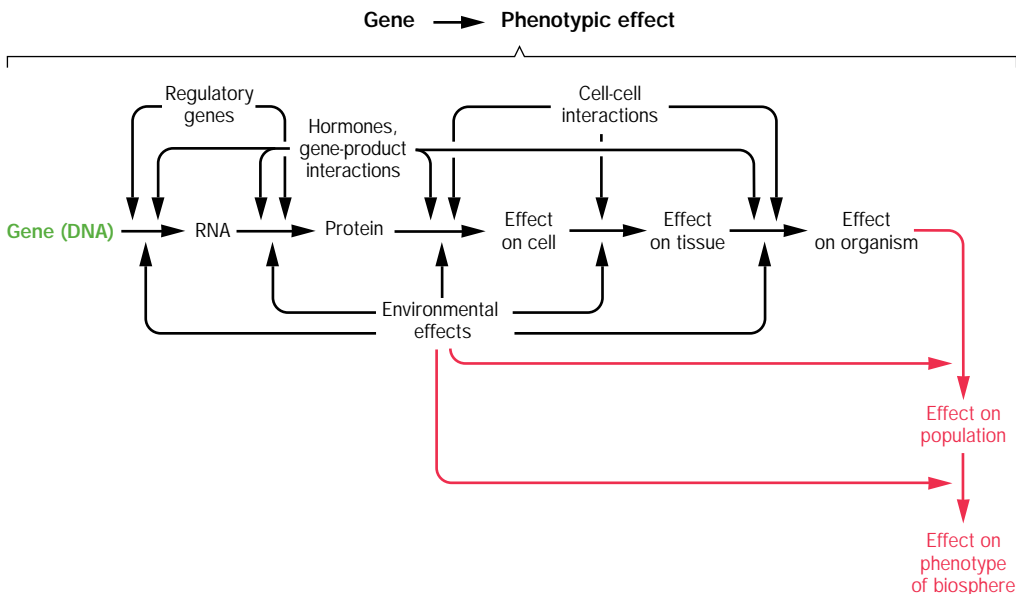


Figure 14.2 The complex pathway by which a gene exerts its effect on the phenotype of an organism, a population, or the biosphere.

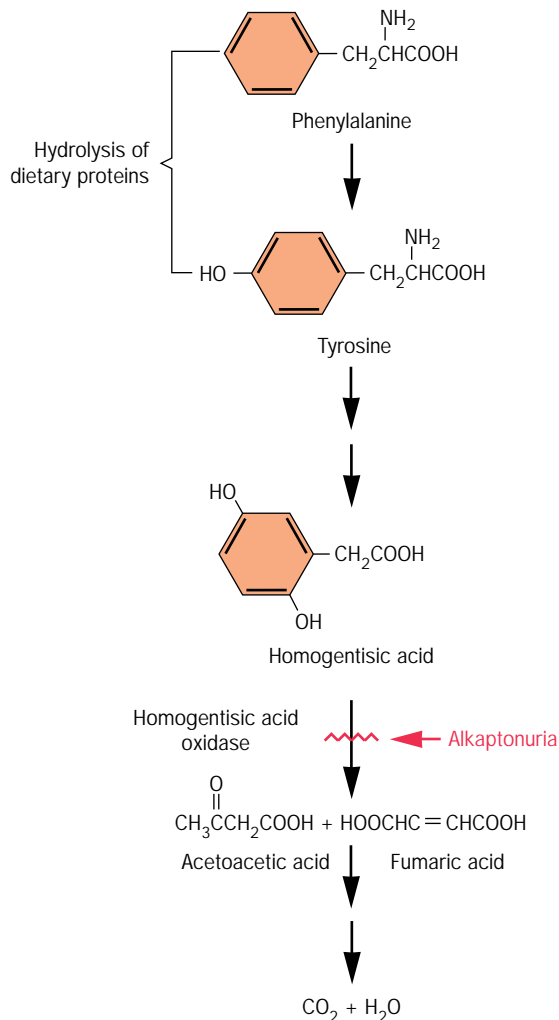


Figure 14.3 Alkaptonuria in humans results from a block in phenylalanine catabolism caused by a mutation in the gene encoding the enzyme homogentisic acid oxidase. When this enzyme is absent or inactive, its substrate, homogentisic acid, accumulates in tissues and in urine.

inherited disease **alkaptonuria**, which is easily detected because of the blackening of the urine upon exposure to air. The substance responsible for this blackening is alkapton (or homogentisic acid), an intermediate in the degradation of the aromatic amino acids tyrosine and phenylalanine (Figure 14.3). Garrod believed that the presence of homogentisic acid in the urine was due to a block in the normal pathway of metabolism of this compound. Moreover, on the basis of the family pedigree studies, Garrod proposed that alkaptonuria was inherited as a single recessive gene. The results of Garrod's studies of alkaptonuria and a few other congenital diseases in humans, such as albinism, were presented in detail in his book, *Inborn Errors of Metabolism*. Although the details of the biochemical pathway affected by the recessive mutations that cause alkaptonuria were not worked out until

many years later, Garrod clearly understood the relationship between genes and metabolism. His concept might be best stated as **one mutant gene—one metabolic block**.

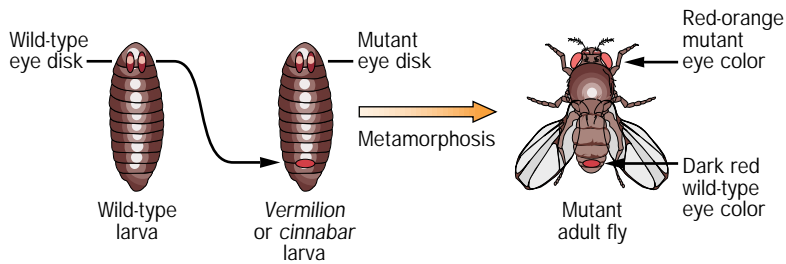
Early Evidence That Enzymes Are Controlled by Genes

In proposing that metabolic reactions were controlled by genes in 1902, Sir Archibald E. Garrod displayed great insight, because there was little direct evidence at the time to support his proposal. Some of the first evidence showing that genes control the enzyme-catalyzed reactions of metabolic pathways was obtained from studies of *Drosophila* eye color mutants. In 1935, George Beadle and Boris Ephrussi suggested that mutations in two genes controlling eye color resulted in blocks at two different steps in the biosynthesis of the brown eye pigment in *Drosophila*. Wild-type fruit flies have dark red eyes resulting from the presence of two eye pigments, one bright red and the other brown. The *vermilion* (*v*) and *cinnabar* (*cn*) mutant flies have bright red eyes owing to the absence of the brown pigment.

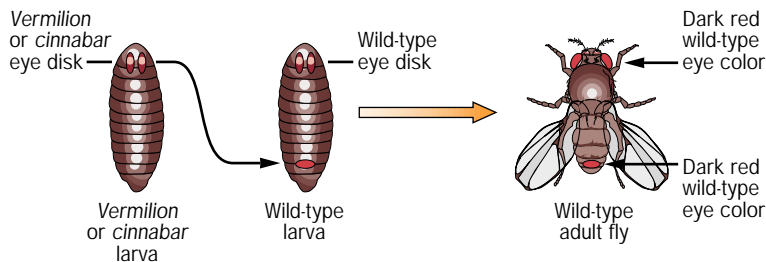
Like other insects, fruit flies undergo metamorphosis; fertilized eggs develop into larvae, then pupae, and finally adult flies (Chapter 2). The larval stages contain groups of embryonic cells called imaginal disks that will develop into the specific adult structures, such as eyes, wings, or legs, of the adult flies. Beadle and Ephrussi discovered that if they surgically transplanted eye imaginal disks from one larva into the abdomen of another larva, the transplanted disk would develop into a third eye, albeit nonfunctional, in the adult fly (Figure 14.4). When they transplanted eye disks from wild-type larvae into *vermilion* or *cinnabar* larvae, the transplanted disks developed into dark red wild-type eyes (Figure 14.4a). In these two cases, the genotype of the transplanted eye disk controlled its own phenotype; the phenotype of the third eye was not influenced by the genotype of the larval host. However, this was not always the case.

When Beadle and Ephrussi transplanted eye disks from *vermilion* or *cinnabar* larvae into the abdomens of wild-type larvae, the disks developed into dark red wild-type eyes (Figure 14.4b). Beadle and Ephrussi reasoned that the wild-type larval hosts must have provided some diffusible substance to the transplanted mutant disks that allowed them to bypass the metabolic block and synthesize the brown eye pigment. In these two cases, the development of the transplanted disks into adult eye structures was influenced by the genotype of the host organism.

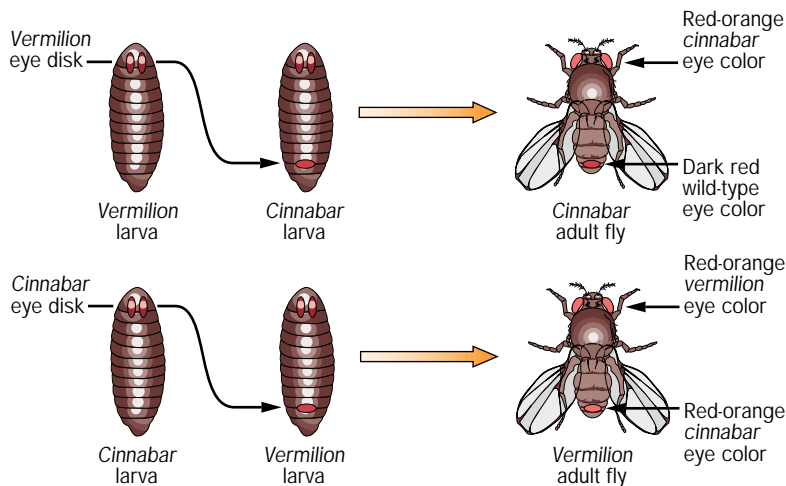
Beadle and Ephrussi next transplanted eye disks from *vermilion* larvae into *cinnabar* hosts and eye disks from *cinnabar* larvae into *vermilion* hosts. The reciproc-



(a) Wild-type disks are transplanted into *vermillion* or *cinnabar* larvae.



(b) *Vermilion* or *cinnabar* disks are transplanted into wild-type larvae.



(c) *Vermilion* disks are transplanted into *cinnabar* larvae and *cinnabar* disks are transplanted into *vermillion* larvae.

Figure 14.4 Transplant experiments of *Drosophila* eye imaginal disks by Beadle and Ephrussi.

cal transplants yielded quite different results! The *vermillion* disks transplanted into *cinnabar* larvae developed into dark red wild-type eyes, whereas the *cinnabar* disks transplanted into *vermillion* larvae developed into bright red *cinnabar* eyes (Figure 14.4c). To explain these results, Beadle and Ephrussi proposed that the *cinnabar* hosts supplied a diffusible substance that the implanted *vermillion* disks converted to the brown eye pigment. They further suggested that the *vermillion* hosts did not produce a metabolite that could be converted to the brown pigment by the *cinnabar* disks.

In fact, the results of Beadle and Ephrussi's reciprocal transplant experiments can be easily explained if

the brown eye pigment is synthesized by a series of enzyme-catalyzed reactions and the v^+ and cn^+ genes control different steps in the biosynthetic pathway (Figure 14.5). Beadle and Ephrussi proposed that (1) a precursor molecule X is converted to a diffusible intermediate substance Y by the product of the v^+ gene, and (2) intermediate Y is converted to the brown eye pigment by the product of the cn^+ gene. If their proposal was correct, the *cinnabar* larval hosts, which were v^+ , would have provided diffusible substance Y to the transplanted *vermillion* disks. Because the *vermillion* disks were cn^+ , they would have converted substance Y to the brown eye pigment and would have developed into dark red wild-type eyes. The *vermillion*

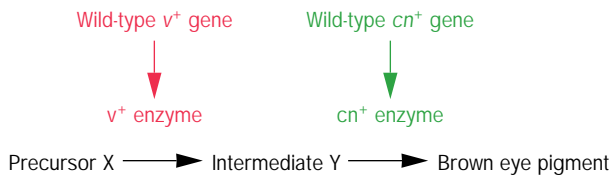


Figure 14.5 Beadle and Ephrussi's interpretation of the results of their reciprocal transplants of *vermillion* and *cinnabar* eye disks in *Drosophila*. They proposed that the v^+ and cn^+ genes control two different steps in the synthesis of the brown eye pigment and that the v^+ gene product acts before the cn^+ gene product.

larval hosts would not have provided anything comparable to substance Y to the transplanted *cinnabar* disks. Since the *cinnabar* disks were v^+ , they would have converted precursor X to intermediate Y. However, in the absence of cn^+ , intermediate Y could not be converted to the brown pigment. As a result, *cinnabar* disks transplanted into *vermillion* larvae developed into bright red mutant eyes.

During the 1940s, Beadle and Ephrussi's proposal was proven correct, albeit with the addition of a few steps to the pathway. The brown eye pigment, xanthommatin, was shown to be synthesized from the amino acid tryptophan (precursor X) by a sequence of enzyme-catalyzed reactions (Figure 14.6). The v^+ gene controls the synthesis of an enzyme called tryptophan pyrrolase, which converts tryptophan to *N*-formylkynurenine. Then, *N*-formylkynurenine is converted to kynurenine (intermediate Y) by the enzyme kynurenine formylase. The cn^+ gene controls the synthesis of the enzyme kynurenine hydroxylase, which converts kynurenine (substance Y) to 3-hydroxylkynurenine. Two additional enzyme-catalyzed reactions then convert 3-hydroxylkynurenine to the brown pigment xanthommatin.

These early studies of *Drosophila* eye color mutants and the pathways of eye pigment biosynthesis provided direct evidence for the genetic control of enzyme-catalyzed metabolic pathways. The results of subsequent investigations demonstrated the genetic control of metabolism in all types of living organisms, from viruses to humans.

Beadle and Tatum: One Gene–One Enzyme

George Beadle's collaboration with Ephrussi in studies of the synthesis of the brown eye pigment in *Drosophila* led him to search for the ideal organism to use in extending this work. The pink bread mold *Neurospora crassa* can grow on medium containing only (1) inorganic salts, (2) a simple sugar, and (3) one vitamin, biotin. *Neurospora* growth medium containing only

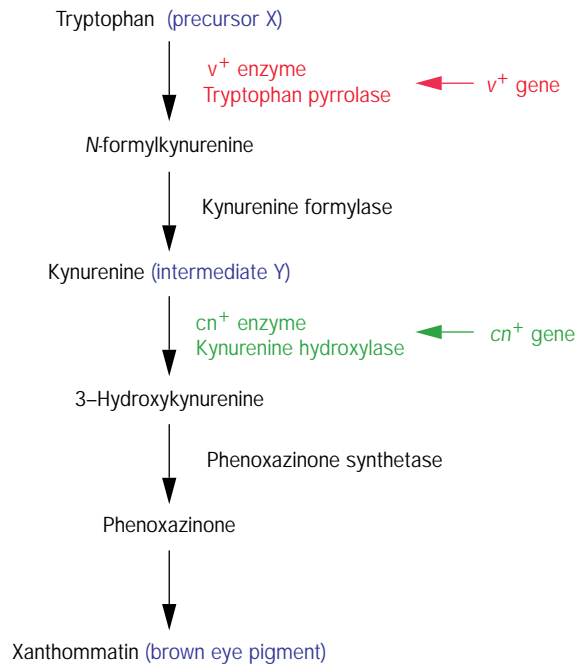


Figure 14.6 Biosynthesis of the brown eye pigment xanthommatin in *Drosophila*. The wild-type *vermillion* and *cinnabar* genes encode enzymes that catalyze two of the reactions in this pathway. Note the consistency of Beadle and Ephrussi's proposed sequence of reactions (Figure 14.5) with the actual biosynthetic pathway.

these components is called “minimal medium.” George Beadle and Edward Tatum reasoned that *Neurospora* must be capable of synthesizing all the other essential metabolites, such as the purines, pyrimidines, amino acids, and other vitamins, *de novo*. Furthermore, they reasoned that the biosynthesis of these growth factors must be under genetic control. If so, mutations in genes whose products are involved in the biosynthesis of essential metabolites would be expected to produce mutant strains with additional growth-factor requirements.

Beadle and Tatum tested this prediction by irradiating asexual spores (conidia) of wild-type *Neurospora* with X rays or ultraviolet light, and screening the clones produced by the mutagenized spores for new growth-factor requirements (Figure 14.7). In order to select strains with a mutation in only one gene, they studied only mutant strains that yielded a 1:1 mutant to wild-type progeny ratio when crossed with wild-type. They identified mutants that grew on medium supplemented with all the amino acids, purines, pyrimidines, and vitamins (called “complete medium”), but could not grow on minimal medium. They analyzed the ability of these mutants to grow on medium supplemented with just amino acids, or just vitamins, and so on (Figure 14.7a). For example, Beadle

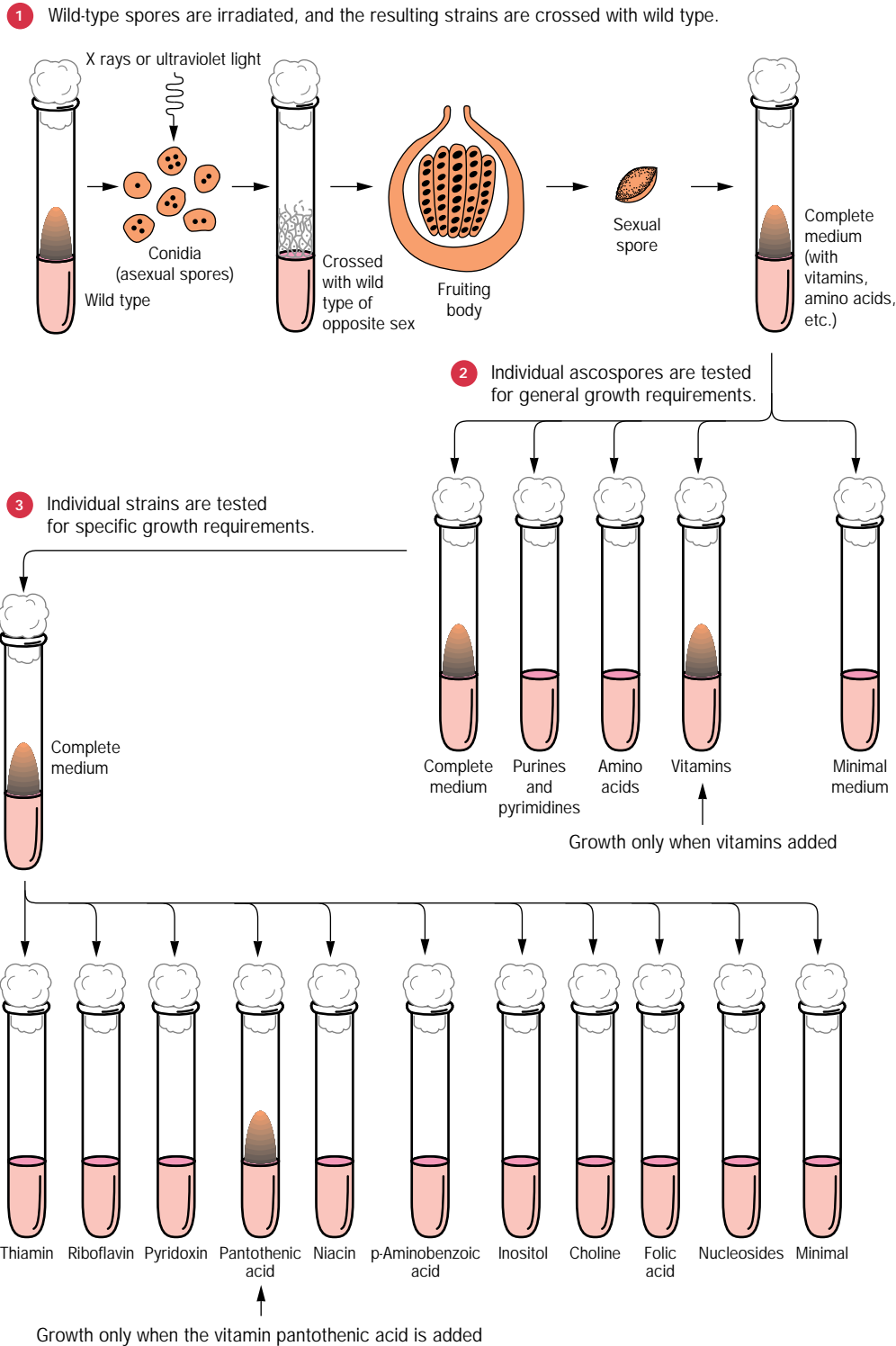


Figure 14.7 Diagram of Beadle and Tatum’s experiment with *Neurospora* that led to the one gene–one enzyme hypothesis.

and Tatum identified mutant strains that grew in the presence of vitamins but could not grow in medium supplemented with amino acids or other growth factors. They next investigated the ability of these vita-

min-requiring strains to grow on media supplemented with each of the vitamins separately (Figure 14.7b).

In this way, Beadle and Tatum demonstrated that each mutation resulted in a requirement for one

growth factor. By correlating their genetic analyses with biochemical studies of the mutant strains, they demonstrated in several cases that one mutation resulted in the loss of one enzyme activity. This work, for which Beadle and Tatum received a Nobel Prize in 1958, was soon verified by similar studies of many other organisms in many laboratories. The **one gene–one enzyme** concept thus became a central tenet of molecular genetics.

Appropriately, in his Noble Prize acceptance speech, Beadle stated:

In this long, roundabout way, we had discovered what Garrod had seen so clearly so many years before. By now we knew of his work and were aware that we had added little if anything new in principle. . . . Thus we were able to demonstrate that what Garrod had shown for a few genes and a few chemical reactions in man was true for many genes and many reactions in Neurospora.

One Gene–One Polypeptide

Subsequent to the work of Beadle and Tatum, many enzymes and structural proteins were shown to be heteromultimeric, that is, to contain two or more different polypeptide chains, with each polypeptide encoded by a separate gene. For example, in *E. coli*, the enzyme tryptophan synthetase is a heterotetramer composed of two α polypeptides encoded by the *trpA* gene and two β polypeptides encoded by the *trpB* gene. Similarly, the hemoglobins, which transport oxygen from our lungs to all other tissues of our bodies, are tetrameric proteins that contain two α -globin chains and two β -globin chains, as well as four oxygen-binding heme groups (see Figure 12.4). In humans, the major form of adult hemoglobin contains two α -globin polypeptides encoded by the *Hb^A _{α}* gene on chromosome 16 and two β -globin polypeptides encoded by the *Hb^A _{β}* gene on chromosome 11. Other enzymes, for example, *E. coli* DNA polymerase III (Chapter 10) and RNA polymerase II (Chapter 11), contain many different polypeptide subunits, each encoded by a separate gene. Thus the one gene–one enzyme concept was modified to **one gene–one polypeptide**.

Key Points: The existence of a basic genetic element, the gene, that controlled a specific phenotypic trait was established by Mendel's work in 1866. Since the discovery of Mendel's results in 1900, the concept of the gene has evolved from the unit that can mutate to cause a specific block in metabolism, to the unit specifying one enzyme, to the sequence of nucleotide pairs in DNA encoding one polypeptide chain.

EVOLUTION OF THE CONCEPT OF THE GENE: STRUCTURE

In the preceding section, we examined the evolution of the concept of the gene as the basic functional component of the genetic material. In this section, we examine the gene from a structural perspective. What is the structure of the gene? Do all genes have the same structure?

The Pre–1940 Beads-on-a-String Concept

Prior to 1940, the genes in a chromosome were considered analogous to beads on a string. Recombination was believed to occur only between the beads or genes, not within genes. The gene was believed to be indivisible. According to this beads-on-a-string concept, the gene was the basic unit of genetic information defined by three criteria: (1) function, (2) recombination, and (3) mutation. More specifically, the gene was

1. *The unit of function*, the unit of genetic material that controlled the inheritance of one “character” or one attribute of phenotype.
2. *The unit of structure*, operationally defined in two ways:
 - a. *By recombination*: as the unit of genetic information not subdivisible by recombination.
 - b. *By mutation*: as the smallest unit of genetic material capable of independent mutation.

Geneticists initially thought that all three criteria defined the same basic unit of inheritance, namely, the gene.

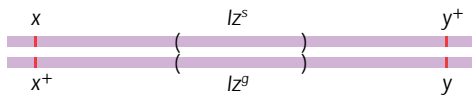
Geneticists now know that these criteria define two different units of inheritance. According to the current molecular concept, the gene is the unit of function, the unit of genetic information controlling the synthesis of one polypeptide chain or, in some cases, one RNA molecule. The unit of structure is simply the structural unit in DNA, the nucleotide pair. Because it clearly does not make sense to call each nucleotide pair a gene, geneticists have focused on the original definition of the gene as the unit of function and have discarded the beads-on-a-string view that the gene is not subdivisible by recombination or mutation. This is clearly appropriate since the emphasis in Mendel's work was on the *Merkmal* (or gene, as it is now called) controlling one phenotypic characteristic.

Discovery of Recombination Within the Gene

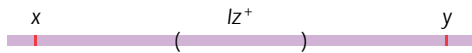
In 1940, Clarence P. Oliver published the first evidence indicating that recombination could occur

within a gene. Oliver was studying mutations at the *lozenge* locus on the X chromosome of *Drosophila melanogaster*. Two mutations, lz^s ("spectacle" eye) and lz^g ("glassy" eye), were thought to be alleles, that is, different forms of the same gene. The data available prior to 1940 indicated that they mapped at the same position on the X chromosome. They had similar effects on the phenotype of the eye, and heterozygous lz^s/lz^g females had lozenge rather than wild-type eyes. However, when lz^s/lz^g females were crossed with either lz^s or lz^g males and large numbers of progeny were examined, wild-type progeny occurred with a frequency of about 0.2 percent.

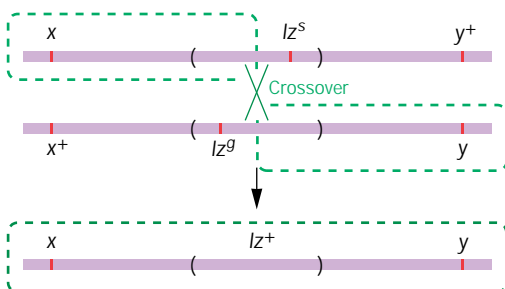
These rare wild-type progeny could be explained by reversion of either the lz^s or the lz^g mutation. But there were two strong arguments against the reversion explanation. (1) The frequency of reversion of lz^s or lz^g to wild-type in hemizygous lozenge males was much lower than 0.2 percent. (2) When the lz^s/lz^g heterozygotes carried genetic markers bracketing the *lozenge* locus, the rare progeny with wild-type eyes always carried an X chromosome with lz^+ that was flanked by recombinant outside markers. Moreover, the same combination of outside markers always occurred, as though the sites of lz^s and lz^g were fixed relative to each other and crossing over was occurring between them. Different sets of outside markers were used and yielded the same result. If the lz^s/lz^g heterozygous female carried X chromosomes of the type



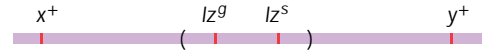
the rare progeny with wild-type eyes all (with one exception) contained an X chromosome with the following composition:



Among progeny of these matings, the reciprocal combination of outside markers (x^+-y^+) never appeared in combination with lz^+ . This result strongly suggested that the lz^s and lz^g mutations were located at distinct sites in the *lozenge* locus, and that the lz^+ chromosome was produced by crossing over between the two sites, as shown in the following diagram.



Definitive evidence for the involvement of recombination required the recovery and identification of the lz^s/lz^g double mutant with the reciprocal combination of outside markers—that is,



Oliver was not able to identify this double mutant because of the inability to distinguish it from the parental single-mutant phenotypes. The identification of both products, the wild-type and double mutant chromosomes, produced by crossing over within the *lozenge* gene, was first accomplished by Melvin M. Green, one of Oliver's students.

The results of these pioneering studies first indicated that the gene was more complex than a bead on a string. They showed that the gene was divisible, containing sites that were separable by crossing over. Oliver's and Green's results were the first step toward the present concept of the gene as a long sequence of nucleotide pairs, capable of mutating and recombining at many different sites along its length.

Recombination Between Adjacent Nucleotide Pairs

The results obtained by Oliver, Lewis, and Green in their studies of *Drosophila* genes all indicated that mutable sites that are separable by recombination can exist within a single gene. Seymour Benzer extended this picture of the gene by demonstrating the existence of 199 distinct sites of mutation that were separable by recombination within the *rIIA* gene of bacteriophage T4 (Chapter 15). Benzer's picture of the gene as a sequence of nucleotide pairs capable of mutating at many distinct sites was soon verified by the results of many researchers investigating gene structure in several different organisms, both prokaryotes and eukaryotes. Given this information about the structure of genes and the known structure of DNA, it followed that the smallest unit of genetic material capable of mutation might be the single nucleotide pair and that recombination might occur between adjacent nucleotide pairs, whether between or within genes. Recombination between adjacent nucleotide pairs of a gene was first demonstrated by Charles Yanofsky in his studies of the *trpA* gene encoding the α polypeptide of tryptophan synthetase in *E. coli*. This enzyme, a tetramer containing two α polypeptides and two β polypeptides, catalyzes the final step in the biosynthesis of the amino acid tryptophan.

Yanofsky and colleagues isolated and characterized a large number of tryptophan auxotrophs with mutations in the *trpA* gene. The wild-type *trpA* gene encodes an α polypeptide that is 268 amino acids long.

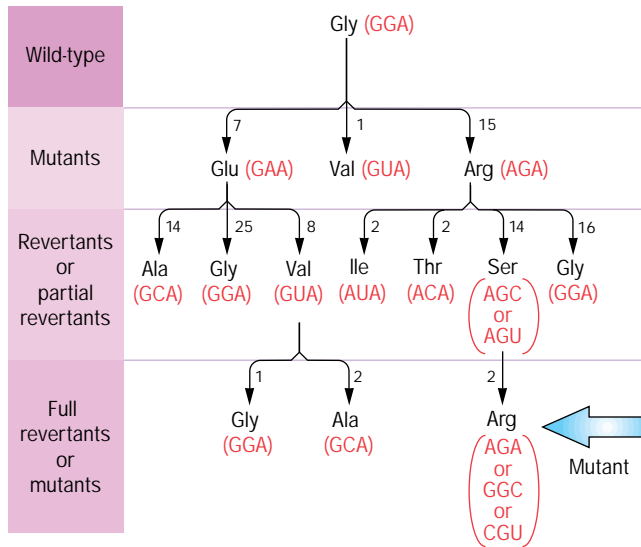


Figure 14.8 Pedigree of amino acid residue 211 (from the NH₂ terminus) of the α polypeptide of tryptophan synthetase of *E. coli*. Amino acid 211 is altered in *trpA23* and *trpA46* mutants (see Figure 14.11). The triplet codons shown in parentheses are the only codons specific to the indicated amino acids that will permit all of the observed amino acid replacements to occur by single base-pair substitutions. The number beside each arrow indicates the number of times that particular substitution was observed. These results indicate that the arginine and glutamic acid codons encoding amino acid 211 of the α polypeptides in *trpA23* and *trpA46* mutant cells are AGA and GAA, respectively. Thus these two mutations alter adjacent nucleotide pairs in the *trpA* gene.

Yanofsky and associates used the laborious techniques of protein sequencing to determine the complete amino acid sequence of the wild-type α polypeptide. They also determined the amino acid substitutions that had occurred in several mutant forms of the tryptophan synthetase α polypeptide. They mapped the

mutations within the *trpA* gene by two- and three-factor crosses, and compared the map positions with the locations of the amino acid substitutions in the mutant polypeptides.

Yanofsky's correlated genetic and biochemical data for the *trpA* gene and the tryptophan synthetase α polypeptide showed that recombination can occur between mutations that alter the same amino acid. Mutations *trpA23* and *trpA46* both result in the substitution of another amino acid (arginine in the case of A23, glutamic acid in the case of A46) for the glycine present at position 211 of the wild-type tryptophan synthetase α polypeptide. However, these two mutations occur at different mutable sites; that is, the A23 and A46 sites are separable by recombination. Yanofsky and colleagues determined the amino acids present at position 211 of the α polypeptide in other mutants as well as in revertants and partial revertants of the *trpA23* and *trpA46* mutants. By using this information and the known codon assignments, they were able to determine which of the glycine, arginine, and glutamic acid codons were present in the *trpA* mRNA at the position encoding amino acid residue 211 of the α polypeptides present in *trp*⁺, *trpA23*, and *trpA46* cells, respectively (Figure 14.8).

Once the specific codons in mRNA are known, the corresponding base-pair sequences in the structural gene from which the mRNA is transcribed are also known. One strand of DNA will be complementary to the mRNA, and the second strand of DNA will be complementary to the first strand. Therefore, Yanofsky's data demonstrated that the mutational events that produced the *trpA23* and *trpA46* alleles were G:C to A:T transitions at adjacent nucleotide pairs. The *trp*⁺ cells produced by recombination between chromosomes carrying mutations A23 and A46 demonstrated that recombination had occurred between adjacent nucleotide pairs in the *trpA* gene as shown in Figure 14.9. Yanofsky's results clearly showed that the

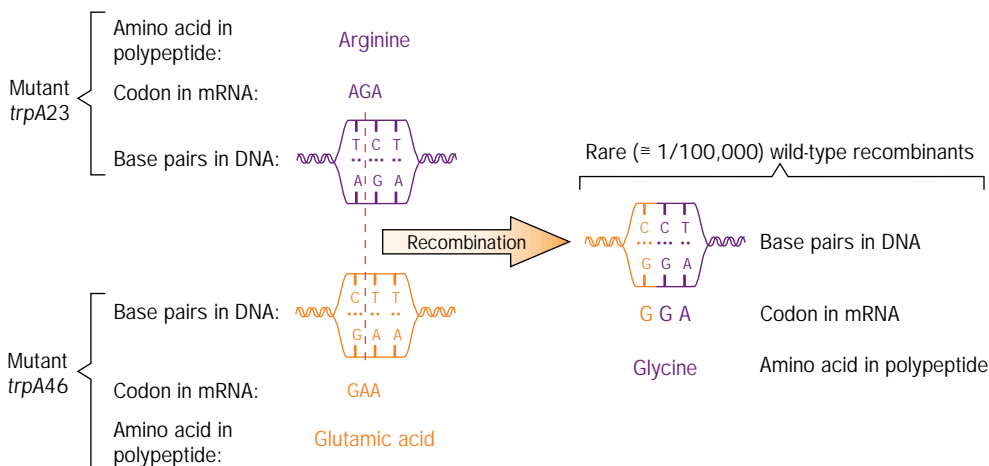


Figure 14.9 Recombination between mutations at adjacent nucleotide pairs in the *trpA* gene of *E. coli*. Mutations A23 and A46 both result in an amino acid substitution at position 211 of the tryptophan synthetase α polypeptide. Wild-type *E. coli* has a glycine residue at this position of the α polypeptide. A23 causes a glycine to arginine substitution; A46 causes a glycine to glutamic acid substitution (see Figure 14.8).

unit of genetic material not divisible by recombination is the single nucleotide pair.

The pre-1940 concept of the gene as (1) the smallest unit of genetic material that could undergo mutation and (2) the unit of genetic information that could not be subdivided by recombination was clearly wrong. Each nucleotide pair of a gene can change or mutate independently, and recombination can occur between adjacent nucleotide pairs. The results of Oliver, Lewis, Green, Benzer, Yanofsky, and many others have compelled geneticists to focus on the gene as the unit of function, the sequence of nucleotide pairs controlling the synthesis and structure of one polypeptide or one RNA molecule.

Colinearity Between the Coding Sequence of a Gene and Its Polypeptide Product

The genetic information is stored in linear sequences of nucleotide pairs in DNA (or nucleotides in RNA, in some cases). Transcription and translation convert this

genetic information into linear sequences of amino acids in polypeptides, which function as the key intermediaries in the genetic control of the phenotype.

It is now known that the nucleotide-pair sequences of the coding regions of the structural genes and the amino acid sequences of the polypeptides that they encode are **colinear**. That is, the first three base pairs of the coding sequence of a gene specify the first amino acid of the polypeptide, the next three base pairs (four to six) specify the second amino acid, and so on, in a colinear fashion (Figure 14.10a). It is also known that the coding regions of most of the genes in higher eukaryotes are interrupted by noncoding introns (Chapter 12). However, the presence of introns in genes does not invalidate the concept of colinearity. The presence of introns in genes simply means that there is no direct correlation in physical distances between the positions of base-pair coding triplets in a gene and the positions of amino acids in the polypeptide specified by that gene (Figure 14.10b).

The first strong evidence for colinearity between a gene and its polypeptide product resulted from stud-

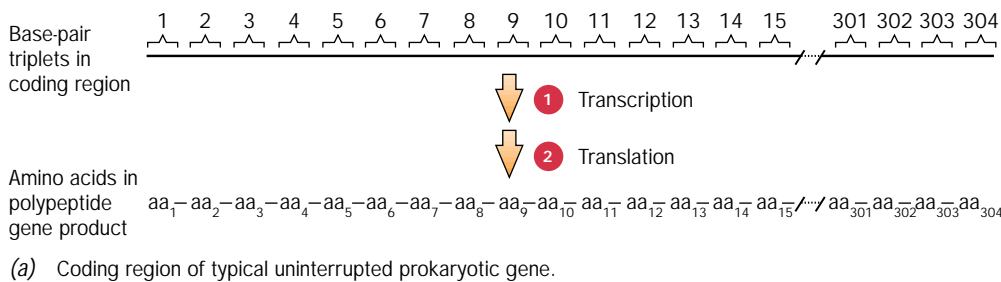
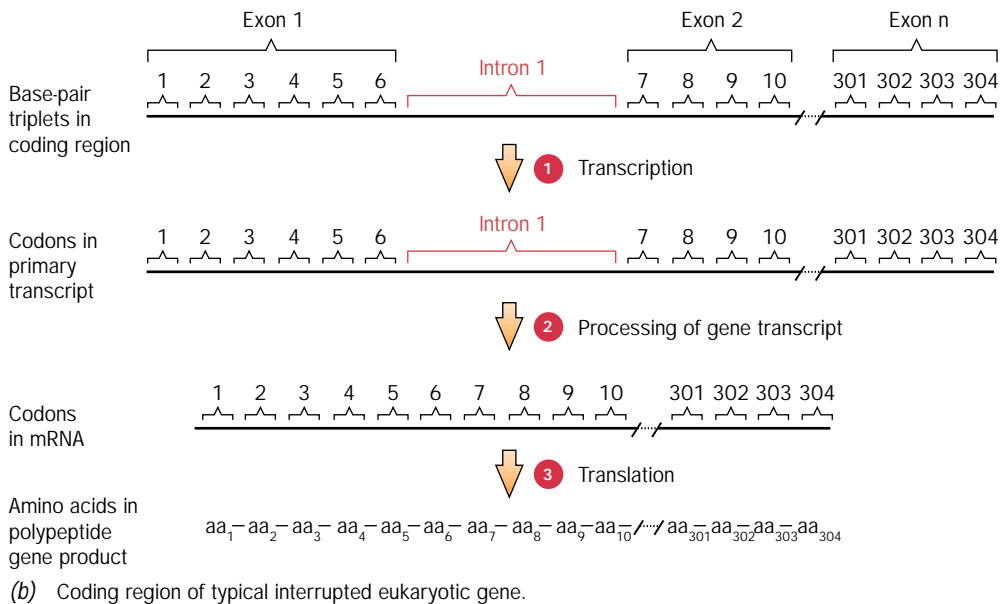


Figure 14.10 Colinearity between the coding regions of genes and their polypeptide products.



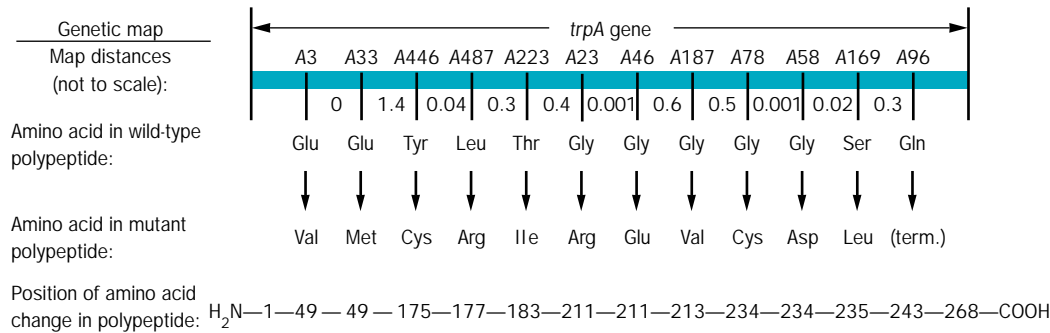


Figure 14.11 Colinearity between the *E. coli trpA* gene and its polypeptide product, the α polypeptide of tryptophan synthetase. The map positions of mutations in the *trpA* gene are shown at the top, and the locations of the amino acid substitutions produced by these mutations are shown below the map.

ies of Charles Yanofsky and colleagues on the *E. coli* gene encoding the α subunit of the enzyme tryptophan synthetase. Tryptophan synthetase catalyzes the final step in the biosynthesis of the amino acid tryptophan. As mentioned earlier, this enzyme contains two α polypeptides encoded by the *trpA* gene and two β polypeptides encoded by the *trpB* gene. Yanofsky and coworkers performed a detailed genetic analysis of mutations in the *trpA* gene and correlated the genetic data with biochemical data on the sequences of the wild-type and mutant tryptophan synthetase α polypeptides. They demonstrated that there was a direct correlation between the map positions of mutations in the *trpA* gene and the positions of the resultant amino acid substitutions in the tryptophan synthetase α polypeptide (Figure 14.11).

About the same time, Sydney Brenner and associates demonstrated a similar colinearity between the positions of mutations in the gene of bacteriophage T4 that encodes the major structural protein of the phage head and the positions in the polypeptide affected by these mutations. Brenner and colleagues studied *amber* (UAG chain-termination) mutations and demonstrated a direct correlation between the length of the polypeptide fragment produced and the position of the mutation within the gene.

In the yeast *S. cerevisiae*, Fred Sherman and colleagues demonstrated an uninterrupted colinear relationship between the map positions of mutations in the *CYC1* gene and the amino acid substitutions in the mutant forms of iso-1-cytochrome c, the polypeptide specified by the *CYC1* gene. More recently, colinearity has been documented for many interrupted genes in eukaryotes. The only difference observed with interrupted genes is that the linear sequence of nucleotide pairs encoding a colinear polypeptide is not one continuous sequence of nucleotide pairs. Instead, noncoding sequences (introns) intervene between coding sequences (exons).

Definitive evidence for colinearity has been provided by direct comparisons of the nucleotide sequences of genes and the amino acid sequences of their polypeptide products. One of the first cases where the amino acid sequence of a polypeptide and the nucleotide sequence of the gene encoding it were both determined experimentally involved the coat protein of bacteriophage MS2 and the gene that encodes it. This small virus has an RNA genome that encodes only four proteins, one being the coat protein that encapsulates the RNA. When the genetic code was used to compare the nucleotide sequence of the coat protein gene with the amino acid sequence of the coat polypeptide, the sequences exhibited perfect colinearity (Figure 14.12). Since then, similar results have established colinearity between many genes and their protein products in organisms ranging from viruses to humans. Thus colinearity between the coding regions of genes and the amino acid sequences of their polypeptide products is a universal or nearly universal feature of gene-protein relationships.

Key Points: The concept of the gene has evolved from a bead on a string, not divisible by recombination or mutation, to a sequence of nucleotide pairs in DNA encoding one polypeptide chain. The unit of genetic material not divisible by recombination or mutation is the single nucleotide pair.

A GENETIC DEFINITION OF THE GENE

With the emergence of the one gene–one polypeptide concept, scientists could define the gene biochemically, but they had no genetic tool to use in determining whether two mutations were in the same or different genes. This deficiency was resolved when Edward Lewis developed the complementation test for func-

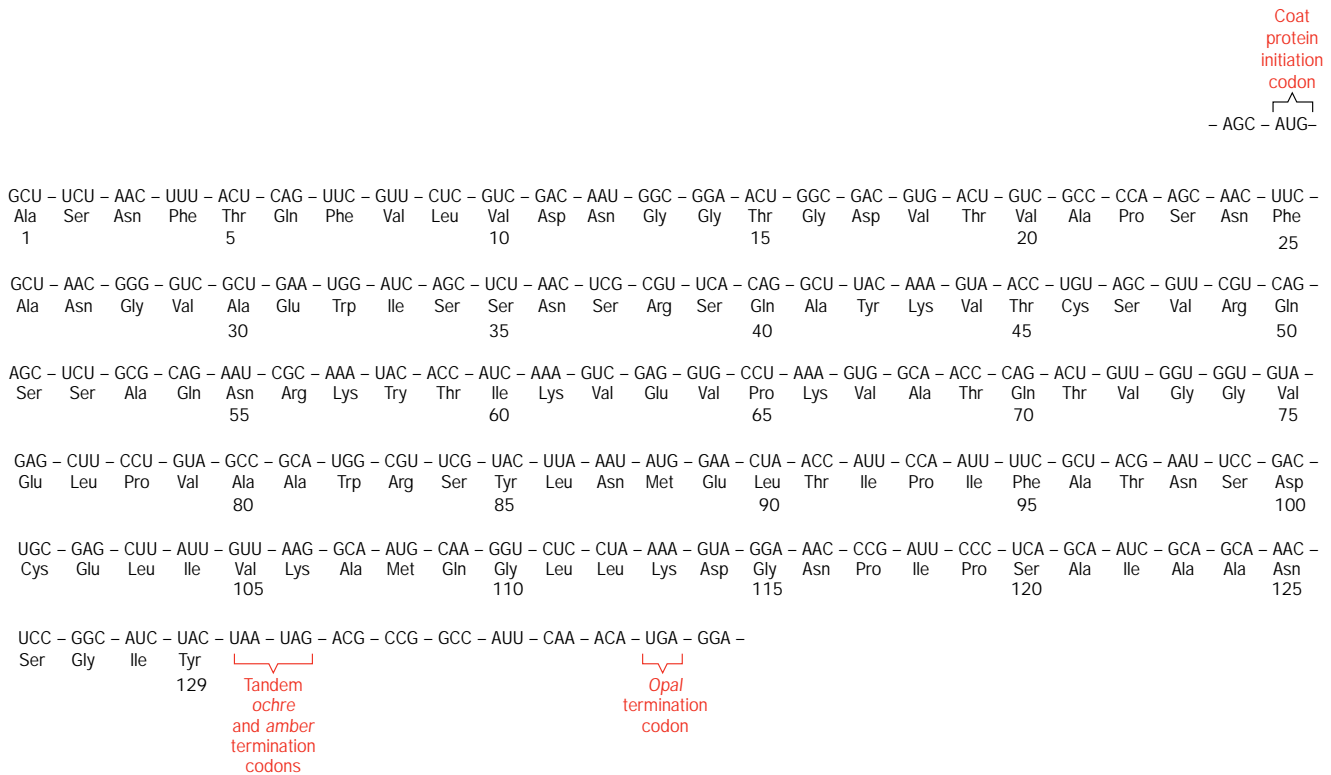


Figure 14.12 Colinearity between the nucleotide sequence of the bacteriophage MS2 coat protein gene and the amino acid sequence of the coat polypeptide that it encodes. Note that the amino acid sequence of this protein is pre-

cisely that predicted from the nucleotide sequence based on the genetic code. In addition, note that all three termination codons are present between the coat gene and the gene downstream from it on the MS2 chromosome.

tional allelism in 1942. Lewis was studying the *Star-asteroid* (small, rough eyes) locus in *Drosophila* and observed that flies of genotype $S^{ast^+}/S^+ ast$ had a more extreme mutant eye phenotype than flies of genotype $S^{ast}/S^+ ast^+$. His results were complicated by the partial dominance of the *Star* (*S*) mutation. Lewis subsequently performed similar experiments with two eye-color mutations, *white* (*w*) and *apricot* (*apr*), and obtained results that were easier to interpret.

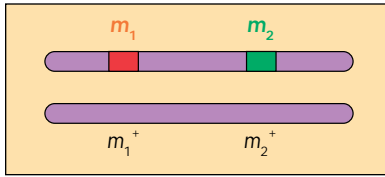
The Complementation Test as an Operational Definition of the Gene

Fruit flies that are homozygous for the X-linked mutations *apr* (now w^a) and *w* have apricot-colored eyes and white eyes, respectively, in contrast to the red eyes of wild-type *Drosophila*.

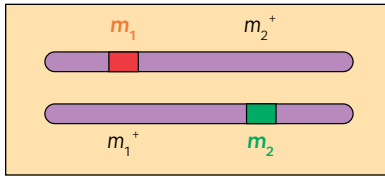
Lewis reported that heterozygous *apr/w* females had light apricot-colored eyes and produced rare red-eyed progeny carrying recombinant apr^+w^+ chromosomes. In addition, Lewis was able to identify progeny flies that carried X chromosomes with the reciprocal *apr w* recombinant genotype. The observed frequency of recombination between the *apr* and *w*

mutations was 0.03 percent. Clearly, the *apr* and *w* mutations were separable by recombination, and *apr* and *w* were not alleles according to the pre-1940 concept of the gene. The two mutations appeared to be in the same unit of function but in two different units of structure. When Lewis produced flies of genotype *apr w/apr^+ w^+*, they had red eyes just like those of wild-type flies. When he constructed flies of genotype *apr w^+/apr^+ w*, they had light apricot-colored eyes. Both genotypes contain the same mutant and wild-type genetic information, but in different arrangements. The presence of different phenotypes in organisms that contain the same genetic markers, but with the markers present in different arrangements, is called a **position effect**, and the type of position effect observed by Lewis is referred to as a **cis-trans position effect**.

Before we analyze Lewis's results in greater detail, we need to define some terms. A double heterozygote, which carries two mutations and their wild-type alleles, that is, m_1 and m_1^+ plus m_2 and m_2^+ , can exist in either of two arrangements (Figure 14.13). When the two mutations are on the same chromosome, the arrangement is called the **coupling** or **cis configuration**; a heterozygote with this genotype is called a **cis heterozygote** (Figure 14.13a). When the two mutations are on



(a) cis heterozygote.



(b) trans heterozygote.

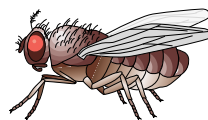
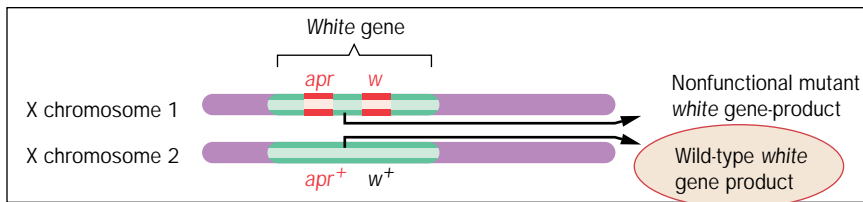
Figure 14.13 The arrangement of genetic markers in *cis* and *trans* heterozygotes.

different chromosomes, the arrangement is called the **repulsion** or **trans configuration**. An organism with this genotype is a **trans heterozygote** (Figure 14.13b).

Recall that *apr w/apr+ w+* *cis* heterozygotes have wild-type red eyes, whereas *apr w+/apr+ w* *trans* het-

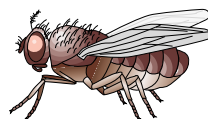
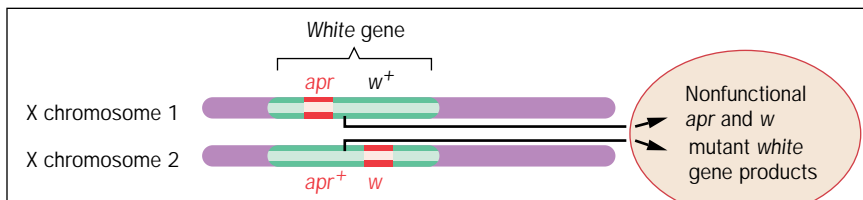
erozygotes have light apricot eyes. This is precisely the result that would be expected if *apr* and *w* are mutations at different sites in the same gene, the unit of genetic information encoding a single polypeptide (Figure 14.14). Thus *apr* and *w* are considered alleles of the same gene, and *apr* is now designated *w^a* in recognition of this relationship. If *apr* and *w* had been mutations in two different units of function, two different genes, both the *cis* and the *trans* heterozygotes should have expressed the wild-type phenotype, namely, red eyes. In the *trans* heterozygote, the *apr+* gene product would be produced by the *apr+* gene on the chromosome carrying the *w* mutation, and the *w+* gene product would be specified by the *w+* gene on the chromosome harboring the *apr* mutation.

Lewis's discovery of the *cis-trans* position effects with *Star-asteroid* and *w^a-w* led to the development of the **complementation test** or **trans test** for functional allelism. The complementation test allows geneticists to determine whether mutations that produce the same or similar phenotypes are in the same gene or in different genes. They must test mutations pairwise by determining the phenotypes of *trans* heterozygotes. That is, they must construct *trans* heterozygotes with each pair of mutations to be analyzed and determine whether these heterozygotes have mutant or wild-type phenotypes.



Therefore, the *cis* heterozygote has wild-type red eyes.

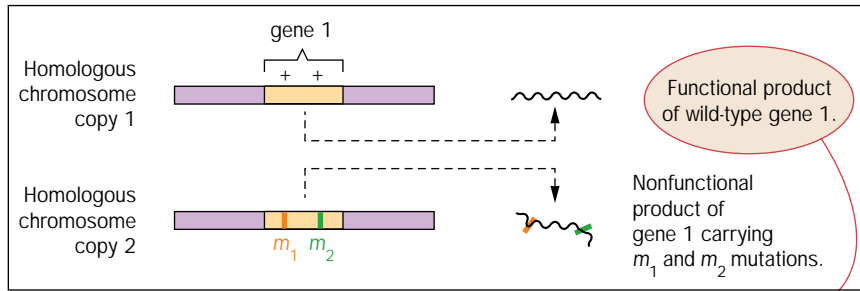
(a) *cis* heterozygote.



Therefore, the *trans* heterozygote has light apricot-colored eyes.

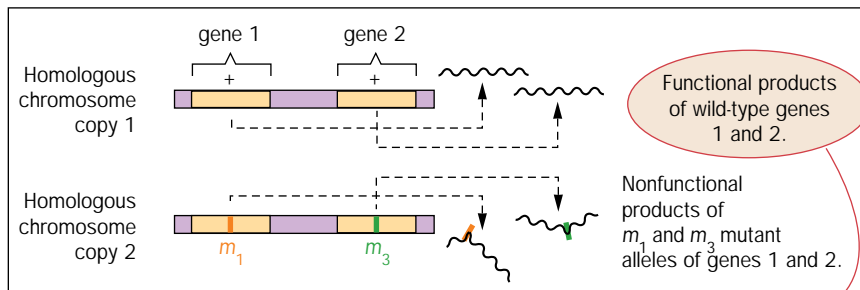
(b) *trans* heterozygote.

Figure 14.14 The *cis-trans* position effect observed by Edward Lewis with the *apr* and *w* mutations of *Drosophila*.



Therefore, the *cis* heterozygote will have the wild-type phenotype.

(a) *cis* heterozygote: mutations in one gene.



Therefore, the *cis* heterozygote will have the wild-type phenotype.

(b) *cis* heterozygote: mutations in two different genes.

Figure 14.15 The *cis* test. The *cis* heterozygote should have the wild-type phenotype whether the mutations are in the same gene (a) or in two different genes (b).

Ideally, the complementation or *trans* test should be done in conjunction with the *cis* test—a control that is often omitted. *Cis* tests are performed by constructing *cis* heterozygotes for each pair of mutations to be analyzed and determining whether they have mutant or wild-type phenotypes. Together, the complementation or *trans* test and the *cis* test are referred to as the *cis-trans* test. Each *cis* heterozygote, which contains one wild-type chromosome, should have the wild-type phenotype whether the mutations are in the same gene or in two different genes (Figure 14.15). Indeed, the *cis* heterozygote must have the wild-type phenotype for the results of the *trans* test to be valid. If the *cis* heterozygote has the mutant phenotype, the *trans* test cannot be used to determine whether the two mutations are in the same gene. Thus the *trans* test cannot be used with dominant mutations. Because *cis* heterozygotes contain one chromosome with wild-type copies of all relevant genes, these genes should specify functional products and a wild-type phenotype.

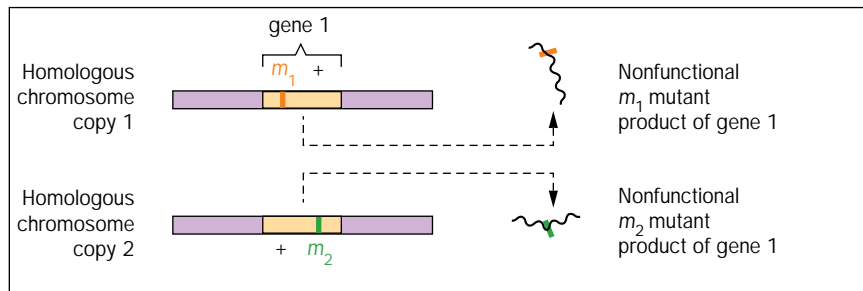
Whether two mutations are in the same gene or two different genes is determined by the results of the complementation or *trans* test. With diploid organisms, the *trans* heterozygote is produced simply by crossing organisms that are homozygous for each of the mutations of interest. With viruses, *trans* heterozygotes are produced by simultaneously infecting host

cells with two different mutants. Regardless of how the two mutations are placed in a common protoplasm in the *trans* configuration, the results of the *trans* or complementation test provide the same information.

1. If the *trans* heterozygote has the mutant phenotype (the phenotype of organisms or cells homozygous for either one of the two mutations), then the two mutations are in the same unit of function, the same gene (Figure 14.16a).
2. If the *trans* heterozygote has the wild-type phenotype, then the two mutations are in two different units of function, two different genes (Figure 14.16b).

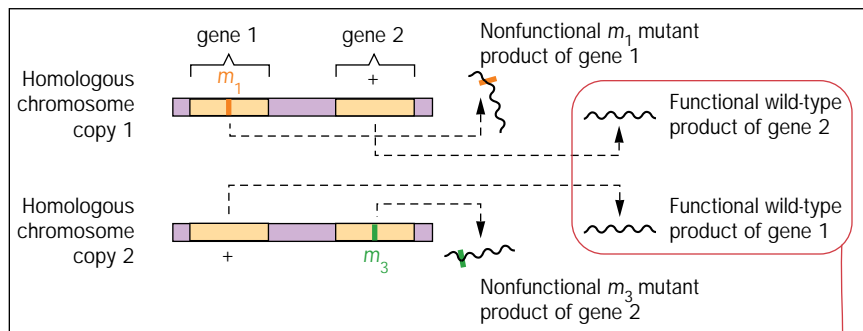
When the two mutations present in a *trans* heterozygote are both in the same gene, as shown for mutations m_1 and m_2 in Figure 14.16a, both chromosomes will carry defective copies of that gene. As a result, the *trans* heterozygote will contain only nonfunctional products of the gene involved and will have a mutant phenotype.

When a *trans* heterozygote has the wild-type phenotype, the two mutations are said to exhibit complementation or to complement each other and are located in different genes. In the example illustrated in Figure



No functional gene 1 product is synthesized in the *trans* heterozygote; therefore, it will have a mutant phenotype.

(a) *trans* heterozygote: mutations in one gene.



Functional products of both genes are synthesized in the *trans* heterozygote; therefore, it will have the wild-type phenotype.

(b) *trans* heterozygote: mutations in two different genes.

Figure 14.16 The *trans* test. The *trans* heterozygote should have (a) the mutant phenotype if the two mutations are in the same gene, and (b) the wild-type phenotype if the mutations are in two different genes.

14.16b, the chromosome carrying mutation m_1 in gene 1 has a wild-type copy of gene 2, which specifies functional gene 2 product, and the chromosome carrying mutation m_3 in gene 2 has a wild-type copy of gene 1, which encodes functional gene 1 product. Thus the *trans* heterozygote shown contains functional products of both genes and has a wild-type phenotype.

Only the complementation test, or the *trans* part of the *cis-trans* test, is included in most genetic analyses. Constructing a chromosome that carries both mutations for the *cis* test is often difficult, especially with eukaryotes. Moreover, the *cis* heterozygotes almost always have wild-type phenotypes if the mutations being analyzed are recessive. Thus, in most instances, the results of complementation or *trans* tests can be interpreted correctly without carrying out the laborious *cis* tests.

Seymour Benzer introduced the term **cistron** to refer to the unit of function operationally defined by the *cis-trans* test. However, today, most geneticists consider the terms gene and cistron to be synonyms. Thus we will use gene, rather than cistron, throughout this text.

The gene is operationally defined as the unit of function by the complementation or *trans* test, which

is used to determine whether mutations are in the same gene or different genes. The complementation test is one of the three basic tools of genetics. The other two genetic tools are recombination (Chapter 7) and mutation (Chapter 13).

The information provided by complementation tests is totally distinct from that obtained from recombination analyses.

The results of complementation tests indicate whether mutations are allelic, whereas the results of recombination analyses indicate whether mutations are linked and, if so, provide estimates of how far apart they are on a chromosome. Nevertheless, students sometimes confuse complementation and recombination. Thus we will contrast complementation and recombination by illustrating both phenomena with the same three mutations in bacteriophage T4. Phage T4 is very similar to phage T2, which we discussed in Chapter 10. Because of the simple structure of the virus and the direct relationship between specific gene products and phenotypes, the phage T4 system provides an excellent mnemonic visualization of the difference between complementation and recombination.

We discussed the *amber* mutations of bacteriophage T4 in Chapter 12 (see Figure 12.23) and the

pathway of morphogenesis of phage T4 in Chapter 13 (see Figure 13.11) *Amber* mutations produce translation-termination triplets within the coding regions of genes. As a result, the products of the mutant genes are truncated polypeptides, which are almost always totally nonfunctional. Therefore, complementation tests performed with *amber* mutations are usually unambiguous. When *amber* mutations occur in essential genes, the mutant phenotype is lethality—that is, no progeny are produced when a restrictive host cell is infected; and the wild-type phenotype is a normal yield (about 300 phage per cell) of progeny phage in each infected restrictive host cell. In Chapter 13, we called such mutations conditional lethals and emphasized their utility in genetic analysis. With conditional lethals, the distinction between the mutant and wild-type phenotypes is maximal: lethality versus normal growth.

Two of the three *amber* mutations that we will consider (*amb17* and *amH32*) are in gene 23, which encodes the major structural protein of the phage head; the other mutation (*amE18*) is in gene 18, which specifies the major structural protein of the phage tail. We can see from Figure 14.17 why complementation occurs between mutations *amb17* (head gene) and *amE18* (tail gene) and why complementation does not occur between mutations *amb17* and *amH32* (both in head gene). Complementation is the result of the interaction of the gene products specified by chromosomes carrying two different mutations when they are present in a common protoplasm. Complementation does not depend on recombination of the two chromosomes or involve any direct interaction between the chromosomes. *Complementation, or the lack of it, is assessed by the phenotype (wild-type or mutant) of each trans heterozygote.*

Figure 14.18 illustrates the occurrence of recombination between the *amber* mutations used to illustrate complementation in Figure 14.17. Although we haven't discussed viral genetics in detail (Chapter 15), we have covered the essential concepts of recombination (Chapter 7). Recombination of phage genes occurs by a process analogous to crossing over in eukaryotes, with linkage distances measured in map units, just as in eukaryotes. Recombination frequencies are measured by infecting permissive host cells with two mutants so that the mutant chromosomes can replicate and participate in crossing over. Then, the progeny are screened for wild-type recombinants by plating them on lawns of restrictive host cells (*E. coli* cells in which only the wild-type phage can grow). In the example shown in Figure 14.18, recombination is observed in both crosses: (1) *amb17* × *amE18*, mutations in two different genes, and (2) *amb17* × *amH32*, mutations in the same gene. The only difference is that more recombinants are produced in cross 1, which involves two *am-*

ber mutations that are relatively far apart on the phage T4 chromosome, than in cross 2, which involves two mutations located near one another in the same gene. *Recombination involves direct interactions between the chromosomes carrying the mutations, the actual breakage of chromosomes, and reunion of parts to produce wild-type and double-mutant chromosomes.*

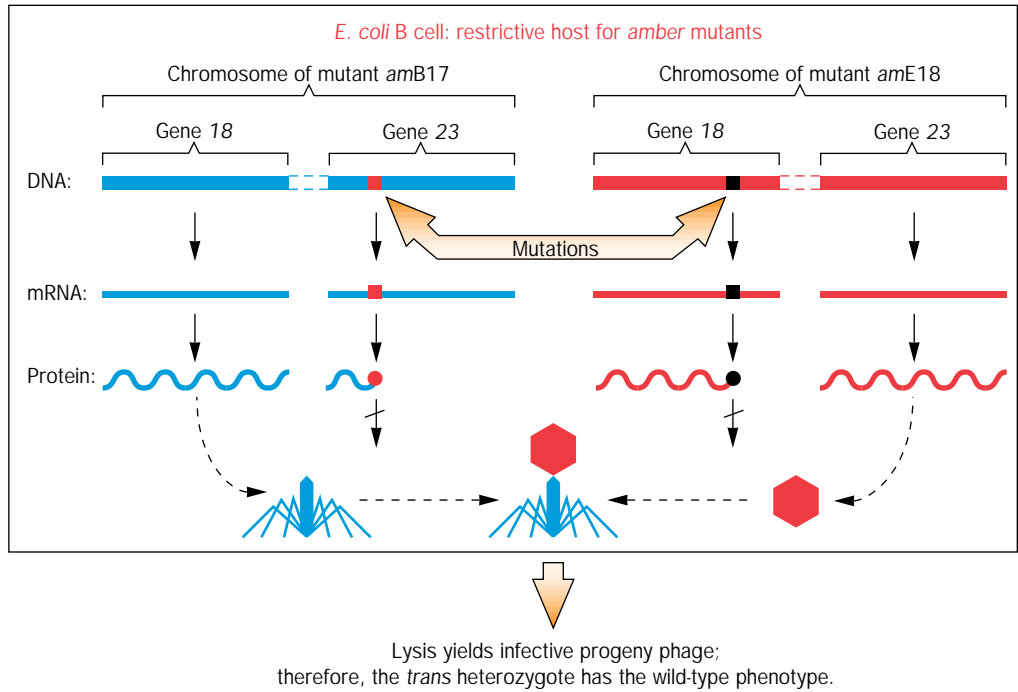
Complementation should occur in every *trans* heterozygote containing mutations in two different genes. Recombination is detected by examining the genotypes of the progeny of heterozygotes, not the phenotypes of the *trans* heterozygotes themselves. Moreover, a *trans* heterozygote will never produce more than 25 percent gametes (or progeny for haploids) with wild-type chromosomes. If the mutations are closely linked, like the *amber* mutations shown in Figure 14.18b, the frequency of wild-type recombinant chromosomes will be much lower than 25 percent.

Structural allelism is the occurrence of two or more different mutations at the same site and is determined by the recombination test. Two mutations that do not recombine are structurally allelic; the mutations either occur at the same site or overlap a common site. Functional allelism is determined by the complementation test as just described; two mutations that do not complement are in the same unit of function, the same gene. Mutations that are both structurally and functionally allelic are called **homoalleles**; they do not complement or recombine with each other. Mutant homoalleles have defects at the same site or overlap a common site in the same gene. Mutations that are functionally allelic, but structurally non-allelic, are called **heteroalleles**; they recombine with each other but do not complement one another. Mutant heteroalleles occur at different sites but within the same gene.

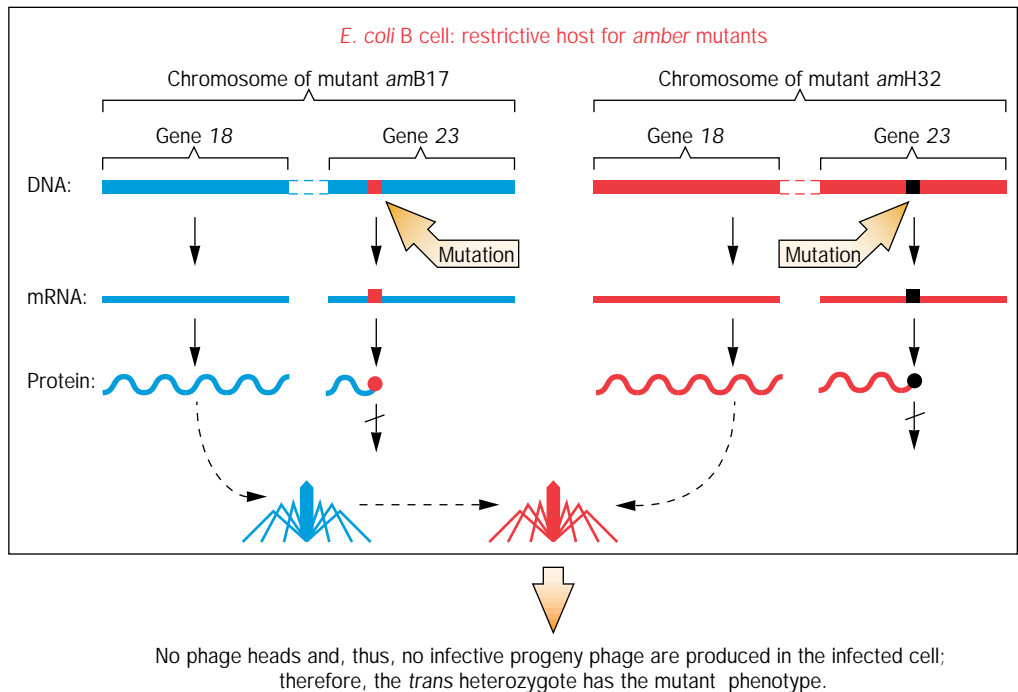
Intragenic Complementation

The results of complementation tests are usually unambiguous when mutations that result in the synthesis of no gene product, partial gene products, or totally defective gene products are used—for example, deletions of segments of genes, frameshift mutations, or polypeptide chain-terminating mutations. Of course, the mutations must be recessive. When mutations causing amino acid substitutions are used, the results of complementation tests are sometimes ambiguous because of the occurrence of a phenomenon called **intragenic complementation**.

The functional forms of some proteins are dimers or higher multimers consisting of two or more polypeptides. These polypeptides may be either homologous, the products of a single gene, or nonhomologous, the products of two or more distinct genes.



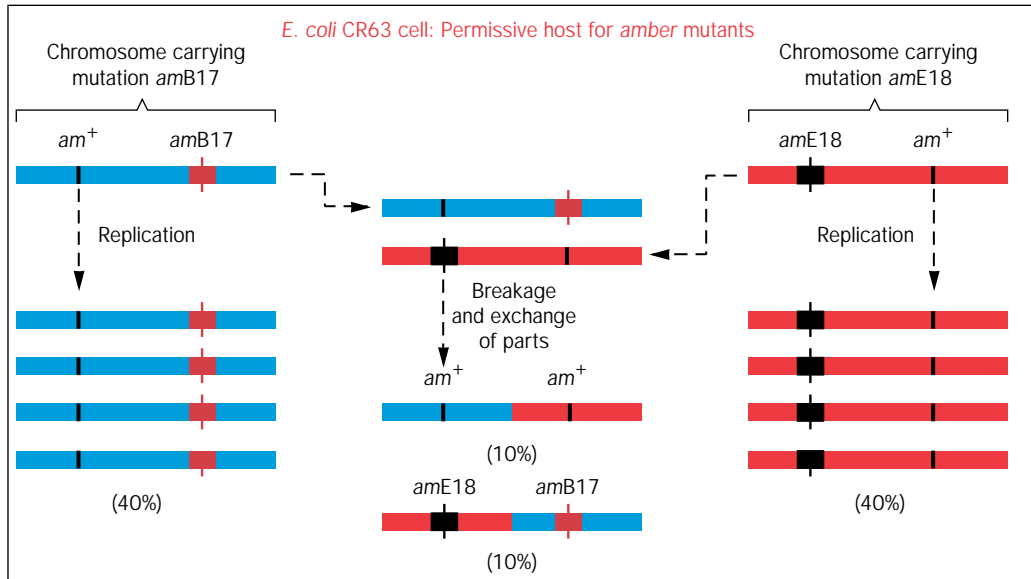
(a) Complementation between mutations *amB17* and *amE18*.



(b) Lack of complementation between mutations *amB17* and *amH32*.

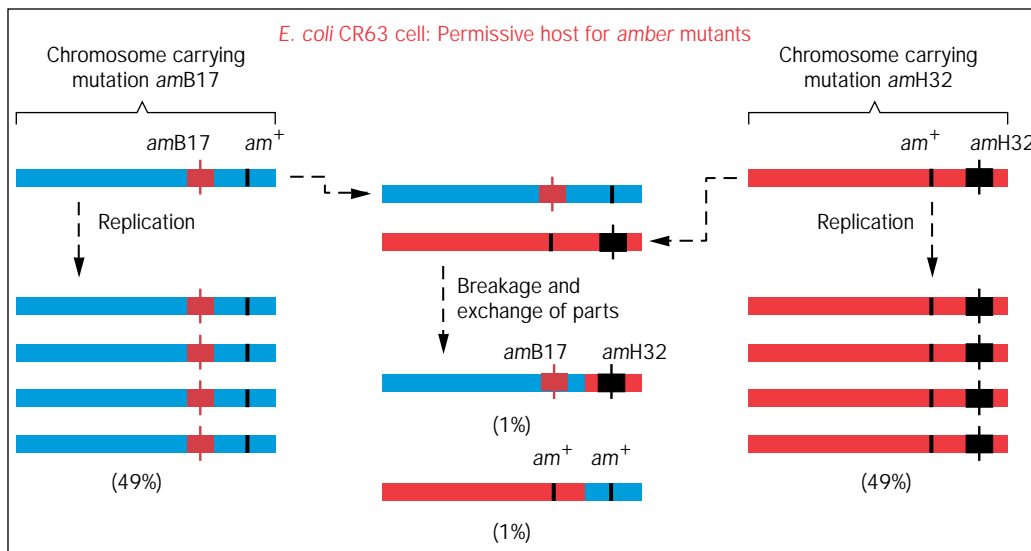
Figure 14.17 Complementation and noncomplementation in *trans* heterozygotes. (a) Complementation between mutation *amB17* in gene 23, which encodes the major structural protein of the phage T4 head, and mutation *amE18* in gene 18, which encodes the major structural protein of the phage tail. Both heads and tails are synthesized in the cell,

with the result that infective progeny phage are produced. (b) When the *trans* heterozygote contains two mutations (*amB17* and *amH32*) in gene 23, no heads are produced, and no infective progeny phage can be assembled. Compare with Figure 14.18.



Lysis yields progeny phage of four genotypes:
 Parental genotypes : $\approx 40\%$ *amB17* and $\approx 40\%$ *amE18*;
 Recombinant genotypes: $\approx 10\%$ wild-type (*am⁺*) and $\approx 10\%$ double mutant (*amE18-amB17*).

(a) Recombination between phage T4 chromosomes carrying mutations *amB17* and *amE18*.



Lysis yields progeny phage of four genotypes:
 Parental genotypes : $\approx 49\%$ *amB17* and $\approx 49\%$ *amH32*;
 Recombinant genotypes: $\approx 1\%$ wild-type (*am⁺*) and $\approx 1\%$ double mutant (*amB17-amH32*).

(b) Recombination between phage T4 chromosomes carrying mutations *amB17* and *amH32*.

Figure 14.18 Recombination between (a) the complementing mutations *amB17* (gene 23) and *amE18* (gene 18), and (b) the noncomplementing mutations *amB17* and *amH32* (both in gene 23). Recombination occurs in both cases; how-

ever, fewer recombinants are produced in cells infected with *amB17* and *amH32* because the two mutations are located closer together on the phage T4 chromosome. Compare with Fig. 14.17.

When the active form of the protein contains two or more homologous polypeptides (it may or may not also contain nonhomologous polypeptides), intragenic complementation may occur. Intergenic complementation (discussed in the preceding section) and intra-

genic complementation (described below) are two distinct phenomena.

Let us consider an enzyme that functions as a homodimer, that is, a protein containing two copies of a specific gene product (Figure 14.19). In organisms

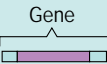
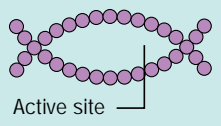
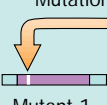
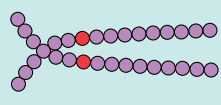
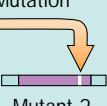
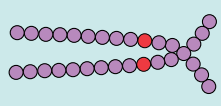
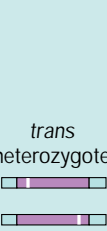
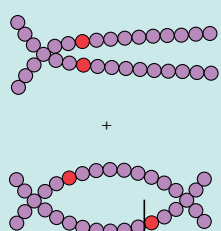
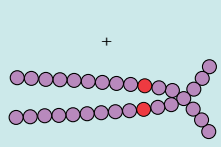
| Genotype | Protein | Phenotype |
|---|---|---------------------------------|
| Gene  Wild-type |  Active | Wild-type |
| Mutation  Mutant 1 |  Inactive | Mutant |
| Mutation  Mutant 2 |  Inactive | Mutant |
| <i>trans</i> heterozygote  |  Active | Wild-type or intermediate |
| |  Inactive | |

Figure 14.19 Intragenic complementation sometimes occurs when the active form of an enzyme or structural protein is a multimer that contains at least two copies of any one gene product. Here, the functional form of the enzyme is a dimer composed of two polypeptides encoded by one gene. The amino acids altered by the mutations are shown as red circles in the polypeptide chains.

that are homozygous for the wild-type allele of the gene, all the protein dimers will contain identical wild-type polypeptides. Similarly, organisms that are homozygous for any mutation in the gene will contain dimers with two mutant polypeptides. An organism that is heterozygous for two different mutations in the gene will produce some dimers that contain the two different mutant polypeptides. We call these heterodimers. Such heterodimers may have partial or complete (wild-type) function. If they do, intragenic complementation has occurred, and the *trans* heterozygote has a wild-type phenotype or a phenotype intermediate between mutant and wild-type (Figure 14.19, bottom). In the case of noncomplementing mutations in a gene encoding a multimeric protein, the heteromultimers are nonfunctional, just like the mu-

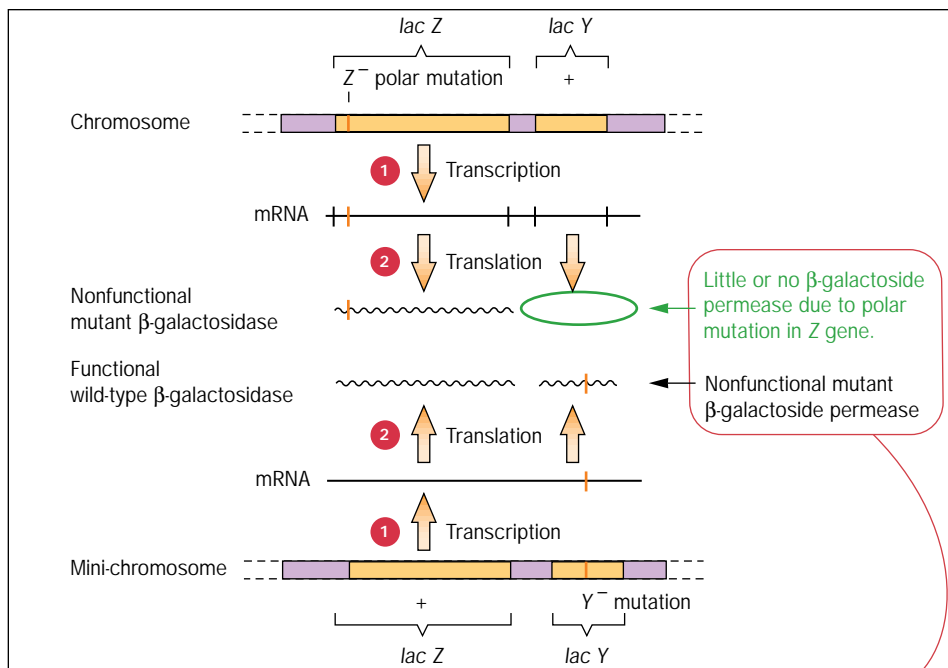
tant homomultimers (protein multimers composed of two or more identical mutant polypeptides).

In several known cases of intragenic complementation, the active form of the protein in the heterozygote has been purified and shown to be a heterodimer or heterotetramer containing two different mutant polypeptides. Why such heteromultimers should be active when the two corresponding homomultimers are inactive is not clear. Apparently, the wild-type sequence of amino acids in the nonmutant segment of one mutant polypeptide somehow compensates for the mutant segment of the polypeptide encoded by the second mutant allele, and vice versa (Figure 14.19). However, most proteins have complex three-dimensional structures, and until the exact structures of a wild-type homomultimer, two mutant homomultimers, and an active heteromultimer composed of the two mutant polypeptides have been determined, the molecular basis of intragenic complementation will continue to be subject to speculation.

Limitations on the Use of the Complementation Test

The complementation test has been very useful in operationally delimiting genes. Usually, two or more mutations that produce the same phenotype can be assigned to one or more genes based on the results of complementation tests. However, in some cases, the results of complementation tests cannot be used to delimit genes. As previously mentioned, complementation tests are not informative in studies of dominant or codominant mutations or in cases where intragenic complementation occurs. In addition, complementation tests are sometimes uninformative because of epistatic interactions between the mutant gene products. If the *cis* test is done, such interactions are readily detected because the *cis* heterozygotes will have mutant phenotypes rather than the required wild-type phenotype.

Another limitation of the complementation test is encountered in working with so-called polar mutations. A **polar mutation** is a mutation that not only results in a defective product of the gene in which it is located, but also interferes with the expression of one or more adjacent genes. The adjacent genes are always located on one side of the gene carrying the mutation (thus the term *polar mutation*). Such polar mutations are frequently observed in prokaryotes in coordinately regulated sets of genes called operons (Chapter 21). They usually are mutations resulting in polypeptide chain-termination signals (nucleotide-pair triplets yielding UAA, UAG, and UGA codons in mRNA) within genes. These polar mutations interfere with the expression of genes located downstream (relative to the direction of transcription) of the mutant gene. As

Trans heterozygote: *E. coli* partial diploid.

Little or no functional β -galactoside permease is produced. Therefore, the *trans* heterozygote has a mutant phenotype — that is, it is unable to utilize lactose as an energy source.

Figure 14.20 Lack of complementation between a polar mutation and a mutation in a downstream gene in the same transcription unit. The ability of *E. coli* cells to utilize lactose as an energy source depends on the products of two co-transcribed genes: *lacY*, which encodes β -galactoside permease, and *lacZ*, which encodes β -galactosidase. Transcription of the *lacY* and *lacZ* genes produces a multigenic

mRNA, which is translated to provide the two proteins. A translation-termination mutation near the translation-start site in *lacZ* has a polar effect on translation of the *lacY* gene, reducing its translation efficiency to 1–2% of the normal level. Thus the polar mutation in *lacZ* will not complement a null mutation in *lacY*, even though the two point mutations are in two different genes.

as a result, polar mutations fail to complement mutations in genes subject to the polar effect (Figure 14.20). Thus the results of complementation tests performed with polar mutations are often ambiguous.

Key Points: The complementation or *trans* test provides an operational definition of the gene; it is used to determine whether mutations are in the same gene or different genes. Intragenic complementation may occur when a protein is a multimer containing at least two copies of one gene product.

COMPLEX GENE-PROTEIN RELATIONSHIPS

Most prokaryotic genes consist of continuous sequences of nucleotide pairs, which specify colinear sequences of amino acids in the polypeptide gene products. As we discussed in Chapter 11, most eukaryotic

genes are split into coding sequences (exons) and non-coding sequences (introns). However, because the spliceosomes usually excise introns from primary transcripts by *cis*-splicing mechanisms (processes that join exons from the same RNA molecule), the presence of introns in genes does not invalidate the complementation test as an operational definition of the gene. Nevertheless, in some cases, transcripts of split genes may undergo several different types of splicing, making the relationships between genes and proteins more complex than the usual one gene-one polypeptide. In other cases, expressed genes are assembled from “gene pieces” during the development of the specialized cells in which they are expressed.

Alternate Pathways of Transcript Splicing: Protein Isoforms

Many interrupted eukaryotic genes, such as the mammalian hemoglobin genes and the chicken ovalbumin

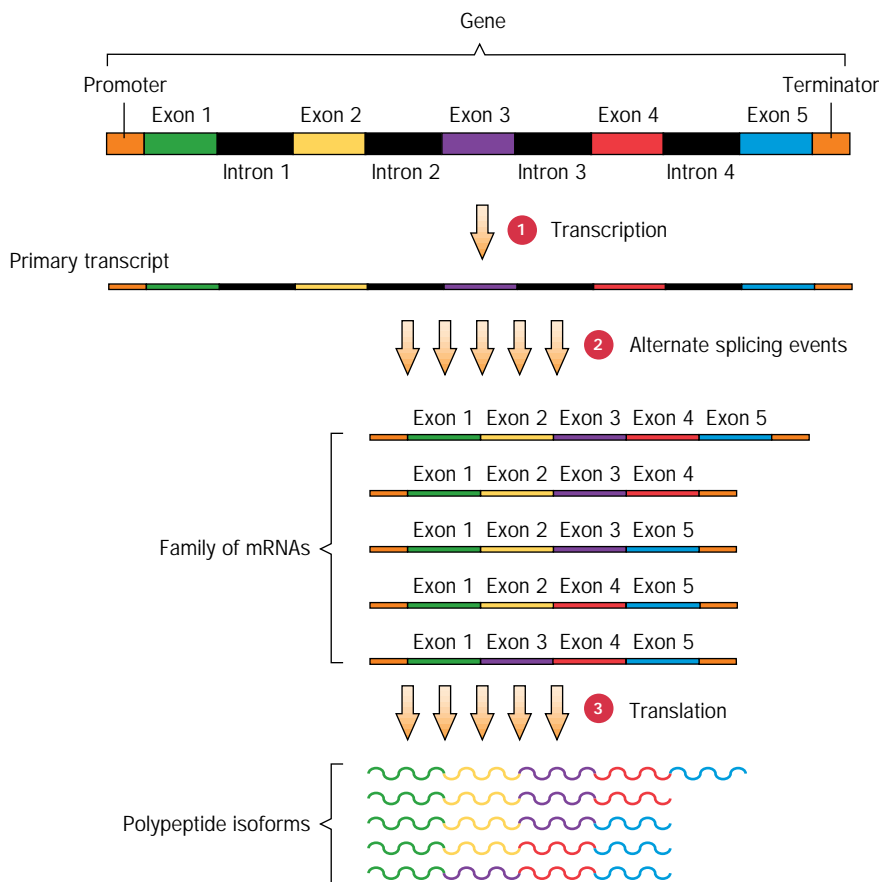


Figure 14.21 A single gene may produce a family of closely related polypeptides by using alternate pathways of exon splicing.

and $1\alpha 2$ collagen genes discussed in Chapter 11, each encode a single polypeptide chain with a specific function. In these cases, the mRNA produced from a given gene contains all the exons of the gene joined together in the same order as they occur in the gene. However, the transcripts of some interrupted genes undergo alternate pathways of transcript splicing. That is, different exons of a gene may be joined to produce a related set of mRNAs encoding a small family of closely related polypeptides called **protein isoforms** (Figure 14.21). The alternate splicing pathways are often tissue-specific, producing related proteins that carry out similar, but not necessarily identical, functions in different types of cells. The mammalian tropomyosin genes provide striking examples of genes which each produce a family of protein isoforms. Tropomyosins are proteins involved in the regulation of muscle contraction in animals. Because the various organs of an animal contain different muscle types, all of which need to be regulated, the availability of a family of related tropomyosins might be beneficial. In any case, one mouse tropomyosin gene is known to produce at least 10 different tropomyosin polypeptides as a result of alternate pathways of transcript splicing. Genes of this type obviously do not fit the one gene–one

polypeptide concept very well. For such genes, where alternate splicing pathways give rise to two or more different polypeptides, the gene can be defined as a DNA sequence that is a single unit of transcription and encodes a set of protein isoforms.

Assembly of Genes During Development: Human Antibody Chains

Genetic information is not always organized into genes of the type described in the preceding sections of this chapter. In rare cases, genes are assembled from a storehouse of **gene segments** during the development of an organism. The immune system of vertebrate animals depends on the synthesis of proteins called **antibodies** to provide protection against infections by viruses, bacteria, toxins, and other foreign substances. Each antibody contains four polypeptides, two identical heavy chains, and two identical light chains. The light chains are of two types: kappa and lambda. Each antibody chain contains a variable region, which exhibits extensive diversity from antibody to antibody, and a constant region, which is largely the same in all antibodies. In germ-line chromosomes,

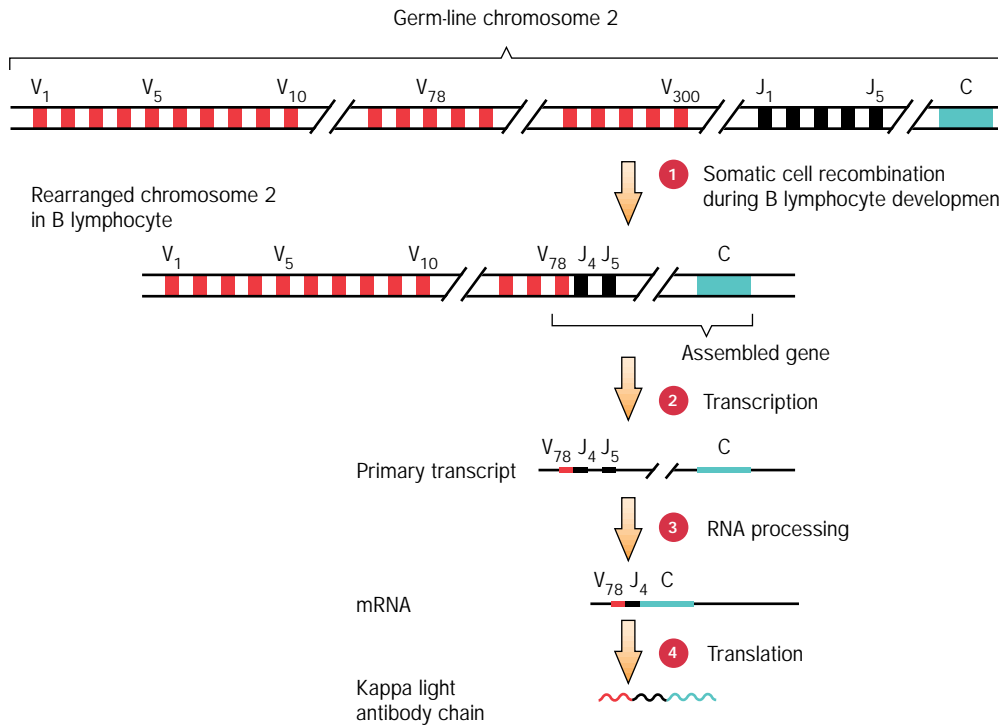


Figure 14.22 Assembly of a gene encoding an antibody kappa light chain from gene segments during the development of a B lymphocyte in humans.

the DNA sequences encoding these antibody chains are present in gene segments, and the gene segments are joined together to produce genes during the differentiation of the antibody-producing cells (B lymphocytes) from progenitor cells.

To illustrate this process of gene assembly during development, let us briefly consider the DNA sequences encoding kappa light chains in humans. (We discuss this topic in detail in Chapter 24.) A kappa light chain gene is assembled from three gene segments: V_k (V for variable region), J_k (J for joining segment), and C_k (C for constant region), during B lymphocyte development. Together, the V_k and J_k gene segments encode the variable region of the kappa light chain, whereas the C_k gene segment encodes the constant region. No functional V_k - J_k - C_k kappa light chain gene is present in any human germ-line chromosome. Instead, human chromosome 2 contains a cluster of about 300 V_k gene segments, another cluster of five J_k gene segments, and a single C_k gene segment (Figure

14.22). During the differentiation of each B lymphocyte, recombination joins one of the V_k gene segments to one of the J_k gene segments. Any J_k segments remaining between the newly formed V_k - J_k exon and the C_k gene segment become part of an intron that is removed during the processing of the primary transcript. Similar somatic recombination events are responsible for the assembly of the genes encoding antibody heavy chains, lambda light chains, and T-lymphocyte receptor proteins (Chapter 24).

Key Points: The transcripts of some genes undergo alternate pathways of splicing to produce mRNAs with different exons joined together. Translation of these mRNAs produces closely related polypeptides called protein isoforms. Other genes, such as those encoding antibody chains, are assembled from gene segments during development by regulated processes of somatic recombination.

TESTING YOUR KNOWLEDGE

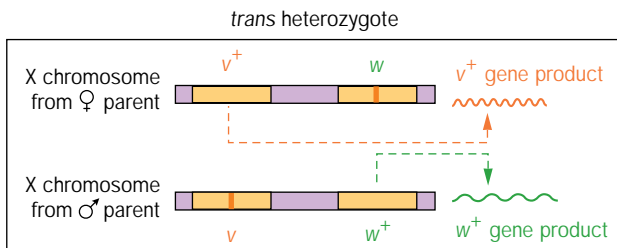
1. In *Drosophila*, *white*, *cherry*, and *vermillion* are all sex-linked mutations affecting eye color. All three mutations are recessive to their wild-type allele(s) for red eyes. A white-eyed female crossed with a vermilion-eyed male produces white-eyed male offspring and red-eyed (wild-type) female

offspring. A white-eyed female crossed with a cherry-eyed male produces white-eyed sons and light cherry-eyed daughters. Do these results indicate whether or not any of the three mutations affecting eye color are located in the same gene? If so, which mutations?

ANSWER

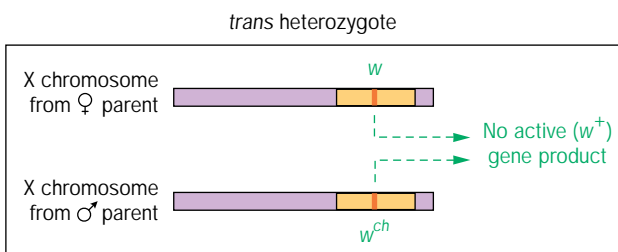
The complementation test for allelism involves placing mutations pairwise in a common protoplasm in the *trans* configuration and determining whether the resulting *trans* heterozygotes have mutant or wild-type phenotypes. If the two mutations are in the same gene, both copies of the gene in the *trans* heterozygote will produce defective gene products, resulting in a mutant phenotype (see Figure 14.16a). However, if the two mutations are in different genes, the two mutations will complement each other, because the wild-type copies of each gene will produce functional gene products (see Figure 14.16b). When complementation occurs, the *trans* heterozygote will have the wild-type phenotype. Thus the complementation test allows one to determine whether any two recessive mutations are located in the same gene or in different genes.

If the *trans* heterozygote has the mutant phenotype, the two mutations are in the same gene. If the *trans* heterozygote has the wild-type phenotype, the two mutations are in two different genes. Because the mutations of interest are sex-linked, all the male progeny will have the same phenotype as the female parent. They are hemizygous, with one X chromosome obtained from their mother. In contrast, the female progeny are *trans* heterozygotes. In the cross between the white-eyed female and the vermilion-eyed male, the female progeny have red eyes, the wild-type phenotype. Thus the *white* and *vermilion* mutations are in different genes, as illustrated in the following diagram:



Complementation yields wild-type phenotype; both v^+ and w^+ gene products are produced in the *trans* heterozygote.

In the cross between a white-eyed female and a cherry-eyed male, the female progeny have light cherry-colored eyes (a mutant phenotype), not wild-type red eyes as in the first cross. Since the *trans* heterozygote has a mutant phenotype, the two mutations, *white* and *cherry*, are in the same gene:



No w^+ gene product; therefore, mutant phenotype.

2. Suppressor-sensitive (*sus*) mutants of bacteriophage $\phi 29$ can grow on *Bacillus subtilis* strain L15, but cannot grow (that is, are lethal) on *B. subtilis* strain 12A. Wild-type (sus^+) $\phi 29$ phage can reproduce on both strains, L15 and 12A. Thus the $\phi 29$ *sus* mutants are conditional lethal mutants like the *amber* mutants of bacteriophage T4 (see Figure 12.23). Seven different *sus* mutants of phage $\phi 29$ were analyzed for complementation by simultaneously infecting the restrictive host (*B. subtilis* strain 12A) with each possible pair of mutants. Single infections with each of the mutants and with wild-type $\phi 29$ were also done as controls. The results of these complementation or *trans* tests and the controls are given as progeny phage per infected cell in the accompanying table. Several infections performed with wild-type $\phi 29$ phage yielded 300 to 400 progeny phage per infected cell. The results of the *cis* controls are not given, but assume that all of the *cis* heterozygote controls yielded over 300 progeny phage per infected cell. Also assume that no intragenic complementation occurs between any of the *sus* mutants studied.

Phage $\phi 29$ Progeny per Infected Bacterium

| Mutant: | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---------|------|--------|--------|------|--------|-----|------|
| 7 | 365 | 384 | 344 | 371 | 347 | 333 | 0.01 |
| 6 | 341 | 301 | 351 | 369 | 329 | 0.1 | |
| 5 | 386 | 326 | 322 | 0.04 | < 0.01 | | |
| 4 | 327 | 398 | 374 | 0.06 | | | |
| 3 | 354 | 387 | < 0.01 | | | | |
| 2 | 0.01 | < 0.01 | | | | | |
| 1 | 0.02 | | | | | | |

(a) Based on these data, how many genes are identified by the seven *sus* mutants? (b) Which *sus* mutations are located in the same gene(s)?

ANSWER

The seven *sus* mutants yielded from <0.01 to 0.1 progeny phage per infected *B. subtilis* strain 12A cell; those data define the mutant phenotype (basically no progeny). Infections of strain 12A cells with wild-type $\phi 29$ produced 300 to 400 progeny phage per infected cell, defining the wild-type phenotype. We then examine the phenotypes of the *trans* heterozygotes to determine whether any of the mutations are located in the same gene(s). In each case, we must ask whether the *trans* heterozygote has the mutant or the wild-type phenotype. If a *trans* heterozygote has the mutant phenotype, the two *sus* mutations are in the same gene. If it has the wild-type phenotype, the two *sus* mutations are in different genes. If you are unsure of why this is true, review Figure 14.16. Of the 21 *trans* heterozygotes examined, 19 exhibited the wild-type phenotype, indicating that in each case the two mutations are in different genes. Two *trans* heterozygotes, (1) *sus1* on one chromosome and *sus2* on a second chromosome and (2) *sus4* on one chromosome and *sus5* on the another, had the mutant phenotype. Therefore, (a) the seven *sus* mutations are located in five different genes, with (b) mutations *sus1* and *sus2* in one gene and mutations *sus4* and *sus5* in another gene.

QUESTIONS AND PROBLEMS

14.1 In what ways does our present concept of the gene differ from the pre-1940 or classical concept of the gene?

14.2 What was the first evidence that indicated that the unit of function and the unit of structure of genetic material were not the same?

14.3 What is the currently accepted operational definition of the gene?

14.4 Of what value are conditional lethal mutations for genetic fine structure analysis?

14.5 Eight independently isolated mutants of *E. coli*, all of which are unable to grow in the absence of histidine (*his*), were examined in all possible *cis* and *trans* heterozygotes (partial diploids). All of the *cis* heterozygotes were able to grow in the absence of histidine. The *trans* heterozygotes yielded two different responses: some of them grew in the absence of histidine; others did not. The experimental results, using + to indicate growth and 0 to indicate no growth, are given in the accompanying table. How many genes are defined by these eight mutations? Which mutant strains carry mutations in the same gene(s)?

Growth of Trans Heterozygotes (without Histidine)

| Mutant: | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---------|---|---|---|---|---|---|---|---|
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | + | 0 |
| 7 | + | + | + | + | + | + | 0 | |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | | |
| 5 | 0 | 0 | 0 | 0 | 0 | | | |
| 4 | 0 | 0 | 0 | 0 | | | | |
| 3 | 0 | 0 | 0 | | | | | |
| 2 | 0 | 0 | | | | | | |
| 1 | 0 | | | | | | | |

14.6 Assume that the mutants described in Problem 14.5 yielded the following results. How many genes would they have defined? Which mutations would have been in the same gene(s)?

Growth of Trans Heterozygotes (without Histidine)

| Mutant: | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---------|---|---|---|---|---|---|---|---|
| 8 | + | + | + | + | + | + | 0 | 0 |
| 7 | + | + | + | + | + | + | 0 | |
| 6 | + | + | + | + | 0 | 0 | | |
| 5 | + | + | + | + | 0 | | | |
| 4 | + | + | 0 | 0 | | | | |
| 3 | + | + | 0 | | | | | |
| 2 | 0 | 0 | | | | | | |
| 1 | 0 | | | | | | | |

14.7 What determines the maximum number of different alleles that can exist for a given gene?

14.8 What is the difference between a pair of homoalleles and a pair of heteroalleles?

14.9 Two different inbred varieties of a particular plant species have white flowers. All other varieties of this species have red flowers. What experiments might be done to obtain evidence to determine whether the difference in flower color in these varieties is the result of different alleles of a single gene or the result of genetic variation in two or more genes?

14.10 The *amber* mutants of phage T4 are conditional lethal mutants. They grow on *E. coli* strain CR63 but are lethal on *E. coli* strain B. An *amber* mutant almost never exhibits *intra*-genetic complementation with any other *amber* mutant; for this problem, assume that no *intra*-genetic complementation occurs between any of the mutants involved. The following results were obtained when eight *amber* mutants were analyzed for complementation by infecting the restrictive host (*E. coli* strain B) with each possible pair of mutants. The results of mixed infections by pairs of mutants are shown as 0 if no progeny are produced and as + if progeny phage resulted from the infection with that particular pair of mutants.

| Mutant: | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---------|---|---|---|---|---|---|---|---|
| 8 | + | + | + | + | + | + | 0 | 0 |
| 7 | + | + | + | + | + | + | 0 | |
| 6 | + | + | + | + | + | 0 | | |
| 5 | 0 | + | 0 | + | 0 | | | |
| 4 | + | + | + | 0 | | | | |
| 3 | 0 | + | 0 | | | | | |
| 2 | + | 0 | | | | | | |
| 1 | 0 | | | | | | | |

- (a) These data indicate that the eight *amber* mutations are located in how many different genes?
- (b) Which mutations are located in the same gene or genes?

14.11 Considering only base-pair substitutions, how many different mutant homoalleles can occur at one site in a gene?

14.12 Are the following statements concerning the genetic element referred to as the gene true or false?

- (a) The classical (pre-1940) conception of the gene was that it was (1) a unit of physiological function or expression, (2) the smallest unit that could undergo mutation, and (3) a unit not subdivisible by recombination.
- (b) In bacteria, the *cis-trans* test provides an operational definition by which we usually can identify a gene as the unit that specifies one mRNA molecule.
- (c) Our present knowledge of the structure of the gene indicates that the units defined by criteria (2) and (3) in statement (a) above are both equivalent to a single nucleotide pair.
- (d) Studies in the 1940s demonstrated the existence of heteroalleles, clearly indicating that many mutations that were allelic by the functional criterion could be separated by recombination, and thereby indicating that the

units of function, mutation, and recombination are not equivalent.

- (e) Homoalleles are functionally and structurally allelic; heteroalleles are functionally allelic but structurally nonallelic.

14.13 The *rosy* (*ry*) gene of *Drosophila* encodes the enzyme xanthine dehydrogenase; the active form of xanthine dehydrogenase is a dimer containing two copies of the *rosy* gene product. Mutations *ry*² and *ry*^{A2} are both located within the region of the *rosy* gene that encodes the *rosy* polypeptide gene product. However, *ry*²/*ry*^{A2} *trans* heterozygotes have wild-type eye color. How can the observed complementation between *ry*² and *ry*^{A2} be explained given that these two mutations are located in the same gene?

14.14 Both *temperature-sensitive* (*ts*) mutant alleles and *amber* (*am*) mutant alleles have been identified and studied for many of the genes of bacteriophage T4. Different *ts* mutations within the same gene are frequently found to complement each other, whereas different *am* mutations within the same gene practically never complement one another. Why is this difference to be expected?

14.15 Suppressor-sensitive (*sus*) mutants of bacteriophage lambda can grow on *E. coli* strain C600 but cannot grow (that is, are lethal) on *E. coli* strain W3350. In other words, *sus* mutants are conditional-lethal mutants. Seven *sus* mutants were analyzed for complementation by simultaneously infecting the restrictive host (*E. coli* strain W3350) with each possible pair of mutants. Single infections with each mutant and with wild-type lambda were also done as controls. The results of these complementation or *trans* tests and the controls are given as progeny per infected cell in the accompanying table. Several infections with wild-type lambda yielded 120 to 150 progeny phage per infected cell. The results of the *cis* heterozygote controls are not given, but assume that all of the *cis* heterozygotes yielded over 100 progeny phage per infected cell. Also assume that no intragenic complementation occurs between any of these *sus* mutants.

Lambda Progeny per Infected Cell

| Mutant: | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---------|-------|-------|-------|------|-------|-----|------|
| 7 | 0.01 | 133 | 0.01 | 146 | 134 | 128 | 0.01 |
| 6 | 131 | 142 | 161 | 0.06 | 0.1 | 0.1 | |
| 5 | 120 | 126 | 134 | 0.05 | <0.01 | | |
| 4 | 147 | 129 | 134 | 0.06 | | | |
| 3 | <0.01 | 147 | <0.01 | | | | |
| 2 | 170 | <0.01 | | | | | |
| 1 | 0.02 | | | | | | |

- (a) Based on the above data, how many genes are defined by the seven *sus* mutants?
 (b) Which *sus* mutations are located in the same gene(s)?

14.16 Is the number of potential alleles of a gene directly related to the number of nucleotide pairs in the gene? Is such a relationship more likely to occur in prokaryotes or in eukaryotes? Why?

14.17 In *Drosophila*, *white*, *eosin*, and *carnation* are all sex-

linked recessive mutations affecting eye color. A white-eyed female crossed with a carnation-eyed male produced white-eyed male progeny and red-eyed (wild-type) female offspring. A white-eyed female crossed with an eosin-eyed male produced white-eyed sons and light eosin-eyed daughters. Based on these data, which of the three mutations (*white*, *eosin*, and *carnation*), if any, are located in the same gene(s)?

14.18 Suppressor-sensitive (*sus*) mutants of bacteriophage ϕ 29 can grow on *Bacillus subtilis* strain L15 but cannot grow (that is, are lethal) on *B. subtilis* strain 12A. Wild-type (*sus*⁺) ϕ 29 phage can reproduce on both strains, L15 and 12A. Thus the ϕ 29 *sus* mutants are conditional lethal mutants like the *amber* mutants of bacteriophage T4. Seven different *sus* mutants of phage ϕ 29 were analyzed for complementation by simultaneously infecting the restrictive host (*B. subtilis* strain 12A) with each possible pair of mutants. Single infections with each of the mutants and with wild-type ϕ 29 were also done as controls. The results of these complementation or *trans* tests and the controls are given as progeny phage per infected cell in the accompanying table. Several infections performed with wild-type ϕ 29 phage yielded 300 to 400 progeny phage per infected cell. The results of the *cis* controls are not given, but assume that all of the *cis* heterozygote controls yielded over 300 progeny phage per infected cell. Also assume that no intragenic complementation occurs between any of the *sus* mutants studied.

Phage ϕ 29 Progeny per Infected Bacterium

| Mutant | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|--------|-------|-------|-------|------|-------|-----|------|
| 7 | 0.01 | 384 | 0.01 | 371 | 347 | 333 | 0.01 |
| 6 | 341 | 301 | 351 | 0.06 | 329 | 0.1 | |
| 5 | 386 | 326 | 322 | 367 | <0.01 | | |
| 4 | 327 | 398 | 374 | 0.06 | | | |
| 3 | <0.01 | 387 | <0.01 | | | | |
| 2 | 354 | <0.01 | | | | | |
| 1 | 0.02 | | | | | | |

- (a) Based on these data, how many genes are identified by the seven *sus* mutants?
 (b) Which *sus* mutations are located in the same gene(s)?

14.19 Assume that the mutants described in Problem 14.18 had yielded the following results.

Phage ϕ 29 Progeny per Infected Bacterium

| Mutant | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|--------|-------|-------|-------|------|-------|------|------|
| 7 | 0.01 | 0.01 | 0.01 | 0.03 | <0.01 | 0.01 | 0.01 |
| 6 | 0.08 | 0.09 | 0.05 | 0.06 | 0.1 | 0.1 | |
| 5 | 0.02 | <0.01 | <0.01 | 0.04 | <0.01 | | |
| 4 | 0.05 | 0.06 | 0.03 | 0.06 | | | |
| 3 | <0.01 | <0.01 | <0.01 | | | | |
| 2 | 0.01 | <0.01 | | | | | |
| 1 | 0.02 | | | | | | |

How many genes would they have defined? Which mutations would have been in the same gene(s)?

14.20 The recessive mutations *bl* (*black*) and *e* (*ebony*) in *Drosophila* both produce flies with black bodies rather than gray bodies like wild-type flies. Mapping studies showed that *bl* is located on chromosome 2, whereas *e* is on chromosome 3. When homozygous *bl/bl* flies are crossed with homozygous *e/e* flies, the heterozygous *bl/e* progeny have gray bodies. The observed complementation indicates that the two mutations are in two different genes. Was it necessary to perform a complementation test to conclude that the *bl* and *e* mutations were located in two different genes? If so, why? If not, why not?

14.21 Why was it necessary to modify Beadle and Tatum's one gene–one enzyme concept of the gene to one gene–one polypeptide?

14.22 In their analysis of gene function, Beadle and Tatum used *Neurospora* as an experimental organism, whereas Garrod had studied gene function in humans. What advantages does *Neurospora* have over humans for such studies?

14.23 Based on the information provided in Figure 14.11, (a) are mutations *trpA3* and *trpA33* heteroalleles or homoalleles? (b) Are mutations *trpA78* and *trpA78* heteroalleles or homoalleles?

14.24 Based on the information given in Figure 14.11, what is the maximum number of nucleotide pairs separating mutations *trpA78* and *trpA78*?

14.25 Arthur Chovnick and colleagues have mapped a large number of recessive mutations that produce fruit flies with rose-colored eyes in the homozygous state. They also have performed complementation tests on these *ry* (*rosy*) mutations. Heterozygotes that carried mutations *ry*⁴² and *ry*⁴⁰⁶ in the *trans* configuration had wild-type eyes, whereas *trans* heterozygotes that harbored *ry*⁵ and *ry*⁴¹ had rose-colored eyes. The results of two- and three-factor crosses unambiguously demonstrated that mutations *ry*⁴² and *ry*⁴⁰⁶ both map between mutations *ry*⁵ and *ry*⁴¹. How can these results be explained?

14.26 The sequences of nucleotide pairs that encode human antibody chains are usually referred to as gene segments rather than genes. Why?

14.27 Tropomyosins are proteins that mediate the interac-

tions between actin and troponin and regulate muscle contractions. In *Drosophila*, six different tropomyosins that have some amino acid sequences in common, but differ in other sequences, are encoded by two tropomyosin genes (*Tm1* and *TmII*). How can two genes encode six different polypeptides?

14.28 In *Drosophila*, *car* (*carnation*) and *g* (*garnet*) are sex-linked mutations that produce brown eyes, in contrast to the dark red eyes of wild-type flies. The *g* and *car* mutations map at positions 44.4 and 62.5, respectively, on the linkage map of the X chromosome. Is a complementation test needed to determine whether these two mutations are in the same gene or two different genes? If so, why? If not, why not?

14.29 The *loz* (*lethal on Z*) mutants of bacteriophage X are conditional lethal mutants that can grow on *E. coli* strain Y but cannot grow on *E. coli* strain Z. The results shown in the following table were obtained when seven *loz* mutants were analyzed for complementation by infecting *E. coli* strain Z with each possible pair of mutants. A + indicates that progeny phage were produced in the infected cells, and a 0 indicates that no progeny phage were produced. All possible *cis* tests were also done, and all *cis* heterozygotes produced wild-type yields of progeny phage.

| Mutant | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|--------|---|---|---|---|---|---|---|
| 7 | + | + | 0 | + | 0 | 0 | 0 |
| 6 | + | + | + | + | + | 0 | |
| 5 | + | + | 0 | + | 0 | | |
| 4 | 0 | 0 | + | 0 | | | |
| 3 | + | + | 0 | | | | |
| 2 | 0 | 0 | | | | | |
| 1 | 0 | | | | | | |

Given that intragenic complementation does not occur between any of the seven *loz* mutants analyzed here, (a) propose four plausible explanations for the apparently anomalous complementation behavior of *loz* mutant number 7. (b) What simple genetic experiments can be used to distinguish between the four possible explanations? (c) Explain why specific outcomes of the proposed experiments will distinguish between the four possible explanations.

BIBLIOGRAPHY

- BEADLE, G. W., AND E. L. TATUM. 1942. Genetic control of biochemical reactions in *Neurospora*. *Proc. Natl. Acad. Sci. USA* 27:499–506.
- CARLSON, E. A. 1966. *The Gene: A Critical History*. W. B. Saunders, Philadelphia.
- FINCHAM, J.R.S. 1966. *Genetic Complementation*. Benjamin, Menlo Park, CA.
- GARROD, A. E. 1909. *Inborn Errors of Metabolism*. Oxford University Press, New York. (Reprinted in H. Harris, 1963, *Garrod's Inborn Errors of Metabolism*, Oxford Monographs on Medical Genetics, Oxford University Press, London.)
- OLIVER, C. P. 1940. A reversion to wild type associated with crossing over in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 26:452–454.
- STURTEVANT, A. H. 1965. *A History of Genetics*. Harper & Row, New York.
- YANOFSKY, C., AND V. HORN. 1972. "Tryptophan synthetase α chain positions affected by mutations near the ends of the genetic map of *trpA* of *Escherichia coli*." *J. Biol. Chem.* 247:4494–4498.