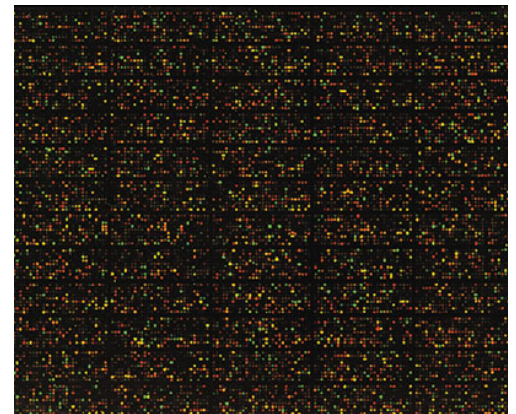


## Exploring Genes and Genomes



Processes such as the development from a caterpillar into a butterfly entail dramatic changes in patterns of gene expression. The expression levels of thousands of genes can be monitored through the use of DNA arrays. At the right, a DNA microarray reveals the expression levels of more than 12,000 human genes; the brightness of each spot indicates the expression level of the corresponding gene. [(Left) Cathy Keifer/istockphoto.com. (Right) Agilent Technologies.]

Since its emergence in the 1970s, recombinant DNA technology has revolutionized biochemistry. The genetic endowment of organisms can now be precisely changed in designed ways. Recombinant DNA technology is the fruit of several decades of basic research on DNA, RNA, and viruses. It depends, first, on having enzymes that can cut, join, and replicate DNA and those that can reverse transcribe RNA. Restriction enzymes cut very long DNA molecules into specific fragments that can be manipulated; DNA ligases join the fragments together. Many kinds of restriction enzymes are available. By applying this assortment cleverly, researchers can treat DNA sequences as modules that can be moved at will from one DNA molecule to another. Thus, recombinant DNA technology is based on the use of enzymes that act on nucleic acids as substrates.

A second foundation is the base-pairing language that allows complementary sequences to recognize and bind to each other. Hybridization with complementary DNA (cDNA) or RNA probes is a sensitive means of detecting specific nucleotide sequences. In recombinant DNA technology, base-pairing is used to construct new combinations of DNA as well as to detect and amplify particular sequences.

Third, powerful methods have been developed for determining the sequence of nucleotides in DNA. These methods have been harnessed to

### OUTLINE

- 5.1** The Exploration of Genes Relies on Key Tools
- 5.2** Recombinant DNA Technology Has Revolutionized All Aspects of Biology
- 5.3** Complete Genomes Have Been Sequenced and Analyzed
- 5.4** Eukaryotic Genes Can Be Quantitated and Manipulated with Considerable Precision

sequence complete genomes: first, small genomes from viruses; then, larger genomes from bacteria; and, finally, eukaryotic genomes, including the 3-billion-base-pair human genome. Scientists are just beginning to exploit the enormous information content of these genome sequences.

Finally, recombinant DNA technology critically depends on our ability to deliver foreign DNA into host organisms. For example, DNA fragments can be inserted into plasmids, where they can be replicated within a short period of time in their bacterial hosts. In addition, viruses efficiently deliver their own DNA (or RNA) into hosts, subverting them either to replicate the viral genome and produce viral proteins or to incorporate viral DNA into the host genome.

These new methods have wide-ranging benefits across a broad spectrum of disciplines, including biotechnology, agriculture, and medicine. Among these benefits is the dramatic expansion of our understanding of human disease. Throughout this chapter, a specific disorder, amyotrophic lateral sclerosis (ALS), will be used to illustrate the effect that recombinant DNA technology has had on our knowledge of disease mechanisms. ALS was first described clinically in 1869 by the French neurologist Jean-Martin Charcot as a fatal neurodegenerative disease of progressive weakening and atrophy of voluntary muscles. ALS is commonly referred to as Lou Gehrig's Disease, for the baseball legend whose career and life were prematurely cut short as a result of this devastating disease. For many years, little progress had been made in the study of the mechanisms underlying ALS. As we shall see, significant advances have been made with the use of research tools facilitated by recombinant DNA technology.

## 5.1 The Exploration of Genes Relies on Key Tools

The rapid progress in biotechnology—indeed its very existence—is a result of a few key techniques.

1. *Restriction-Enzyme Analysis.* Restriction enzymes are precise molecular scalpels that allow an investigator to manipulate DNA segments.
2. *Blotting Techniques.* Southern and northern blots are used to separate and characterize DNA and RNA, respectively. The western blot, which uses antibodies to characterize proteins, was described in Chapter 3.
3. *DNA Sequencing.* The precise nucleotide sequence of a molecule of DNA can be determined. Sequencing has yielded a wealth of information concerning gene architecture, the control of gene expression, and protein structure.
4. *Solid-Phase Synthesis of Nucleic Acids.* Precise sequences of nucleic acids can be synthesized de novo and used to identify or amplify other nucleic acids.
5. *The Polymerase Chain Reaction (PCR).* The polymerase chain reaction leads to a billionfold amplification of a segment of DNA. One molecule of DNA can be amplified to quantities that permit characterization and manipulation. This powerful technique can be used to detect pathogens and genetic diseases, determine the source of a hair left at the scene of a crime, and resurrect genes from the fossils of extinct organisms.

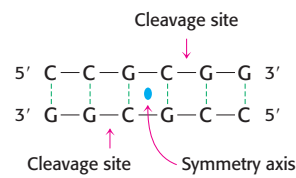
A final set of techniques relies on the computer, without which, it would be impossible to catalog, access, and characterize the abundant information

generated by the techniques just outlined. Such uses of the computer will be presented in Chapter 6.

### Restriction enzymes split DNA into specific fragments

*Restriction enzymes*, also called *restriction endonucleases*, recognize specific base sequences in double-helical DNA and cleave, at specific places, both strands of that duplex. To biochemists, these exquisitely precise scalpels are marvelous gifts of nature. They are indispensable for analyzing chromosome structure, sequencing very long DNA molecules, isolating genes, and creating new DNA molecules that can be cloned. Werner Arber and Hamilton Smith discovered restriction enzymes, and Daniel Nathans pioneered their use in the late 1960s.

Restriction enzymes are found in a wide variety of prokaryotes. Their biological role is to cleave foreign DNA molecules. Many restriction enzymes recognize specific sequences of four to eight base pairs and hydrolyze a phosphodiester bond in each strand in this region. A striking characteristic of these cleavage sites is that they almost always possess *twofold rotational symmetry*. In other words, the recognized sequence is *palindromic*, or an inverted repeat, and the cleavage sites are symmetrically positioned. For example, the sequence recognized by a restriction enzyme from *Streptomyces achromogenes* is



In each strand, the enzyme cleaves the C–G phosphodiester bond on the 3' side of the symmetry axis. As we shall see in Chapter 9, this symmetry corresponds to that of the structures of the restriction enzymes themselves.

Several hundred restriction enzymes have been purified and characterized. Their names consist of a three-letter abbreviation for the host organism (e.g., *Eco* for *Escherichia coli*, *Hin* for *Haemophilus influenzae*, *Hae* for *Haemophilus aegyptius*) followed by a strain designation (if needed) and a roman numeral (if more than one restriction enzyme from the same strain has been identified). The specificities of several of these enzymes are shown in Figure 5.1.

Restriction enzymes are used to cleave DNA molecules into specific fragments that are more readily analyzed and manipulated than the entire parent molecule. For example, the 5.1-kb circular duplex DNA of the tumor-producing SV40 virus is cleaved at one site by *EcoRI*, at four sites by *HpaI*, and at 11 sites by *HindIII*. A piece of DNA, called a restriction fragment, produced by the action of one restriction enzyme can be specifically cleaved into smaller fragments by another restriction enzyme. The pattern of such fragments can serve as a *fingerprint* of a DNA molecule, as will be considered shortly. Indeed, complex chromosomes containing hundreds of millions of base pairs can be mapped by using a series of restriction enzymes.

### Restriction fragments can be separated by gel electrophoresis and visualized

Small differences between related DNA molecules can be readily detected because their restriction fragments can be separated and displayed by gel electrophoresis. In Chapter 3, we considered the use of gel electrophoresis

#### Palindrome

A word, sentence, or verse that reads the same from right to left as it does from left to right.

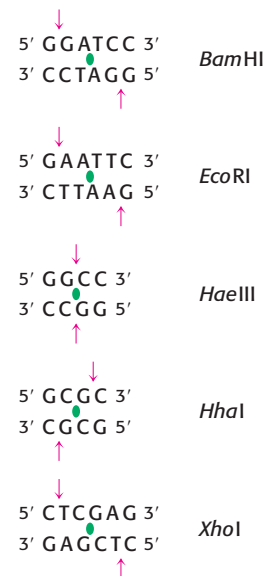
Radar

Senile felines

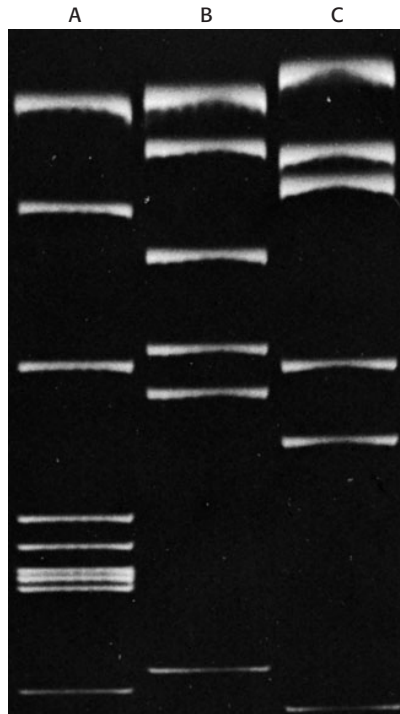
Do geese see God?

Roma tibi subito motibus ibit amor

Derived from the Greek *palindromos*, "running back again."



**Figure 5.1 Specificities of some restriction endonucleases.** The sequences that are recognized by these enzymes contain a twofold axis of symmetry. The two strands in these regions are related by a 180-degree rotation about the axis marked by the green symbol. The cleavage sites are denoted by red arrows. The abbreviated name of each restriction enzyme is given at the right of the sequence that it recognizes. Note that the cuts may be staggered or even.

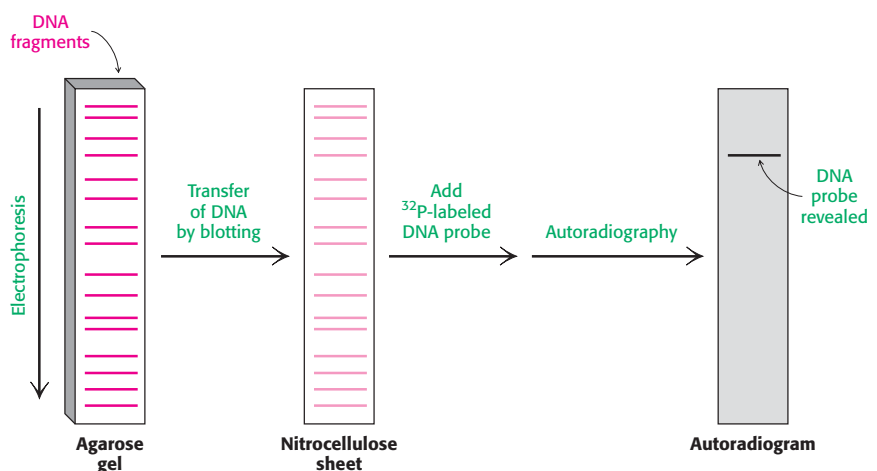


**Figure 5.2 Gel-electrophoresis pattern of a restriction digest.** This gel shows the fragments produced by cleaving SV40 DNA with each of three restriction enzymes. These fragments were made fluorescent by staining the gel with ethidium bromide. [Courtesy of Dr. Jeffrey Sklar.]

to separate protein molecules (Section 3.1). Because the phosphodiester backbone of DNA is highly negatively charged, this technique is also suitable for the separation of nucleic acid fragments. For most gels, the shorter the DNA fragment, the farther the migration. Polyacrylamide gels are used to separate, by size, fragments containing as many as 1000 base pairs, whereas more-porous agarose gels are used to resolve mixtures of larger fragments (as large as 20 kb). An important feature of these gels is their high resolving power. In certain kinds of gels, fragments differing in length by just one nucleotide of several hundred can be distinguished. Bands or spots of radioactive DNA in gels can be visualized by autoradiography. Alternatively, a gel can be stained with ethidium bromide, which fluoresces an intense orange when bound to a double-helical DNA molecule (Figure 5.2). A band containing only 50 ng of DNA can be readily seen.

A restriction fragment containing a specific base sequence can be identified by hybridizing it with a labeled complementary DNA strand (Figure 5.3). A mixture of restriction fragments is separated by electrophoresis through an agarose gel, denatured to form single-stranded DNA, and transferred to a nitrocellulose sheet. The positions of the DNA fragments in the gel are preserved on the nitrocellulose sheet, where they are exposed to a  $^{32}\text{P}$ -labeled single-stranded DNA probe. The probe hybridizes with a restriction fragment having a complementary sequence, and autoradiography then reveals the position of the restriction-fragment-probe duplex. A particular fragment amid a million others can be readily identified in this way. This powerful technique is named *Southern blotting*, for its inventor Edwin Southern.

Similarly, RNA molecules can be separated by gel electrophoresis, and specific sequences can be identified by hybridization subsequent to their transfer to nitrocellulose. This analogous technique for the analysis of RNA has been whimsically termed *northern blotting*. A further play on words accounts for the term *western blotting*, which refers to a technique for detecting a particular protein by staining with specific antibody (Section 3.3). Southern, northern, and western blots are also known respectively as *DNA*, *RNA*, and *protein blots*.

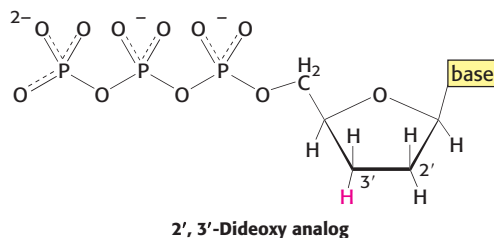


**Figure 5.3 Southern blotting.** A DNA fragment containing a specific sequence can be identified by separating a mixture of fragments by electrophoresis, transferring them to nitrocellulose, and hybridizing with a  $^{32}\text{P}$ -labeled probe complementary to the sequence. The fragment containing the sequence is then visualized by autoradiography.



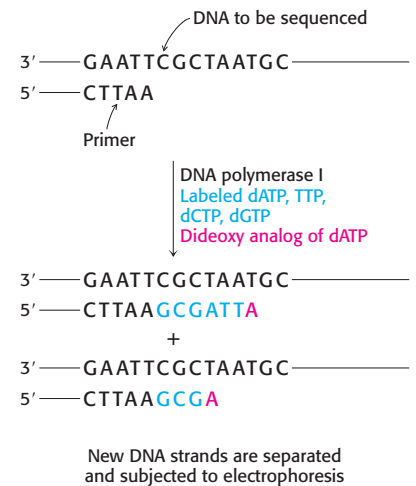
## DNA can be sequenced by controlled termination of replication

The analysis of DNA structure and its role in gene expression also have been markedly facilitated by the development of powerful techniques for the *sequencing* of DNA molecules. The key to DNA sequencing is the generation of DNA fragments whose length depends on the last base in the sequence. Collections of such fragments can be generated through the *controlled termination of replication* (Sanger dideoxy method), a method developed by Frederick Sanger and coworkers. This technique has superseded alternative methods because of its simplicity. The same procedure is performed on four reaction mixtures at the same time. In all these mixtures, a DNA polymerase is used to make the complement of a particular sequence within a single-stranded DNA molecule. The synthesis is primed by a chemically synthesized fragment that is complementary to a part of the sequence known from other studies. In addition to the four deoxyribonucleoside triphosphates (radioactively labeled), each reaction mixture contains a small amount of the 2',3'-*dideoxy analog* of one of the nucleotides, a different nucleotide for each reaction mixture.



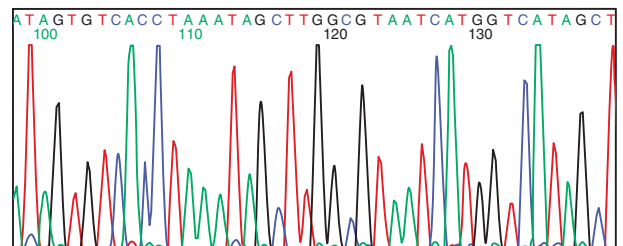
The incorporation of this analog blocks further growth of the new chain because it lacks the 3'-hydroxyl terminus needed to form the next phosphodiester bond. The concentration of the dideoxy analog is low enough that chain termination will take place only occasionally. The polymerase will insert the correct nucleotide sometimes and the dideoxy analog other times, stopping the reaction. For instance, if the dideoxy analog of dATP is present, fragments of various lengths are produced, but all will be terminated by the dideoxy analog (Figure 5.4). Importantly, this dideoxy analog of dATP will be inserted only where a T was located in the DNA being sequenced. Thus, the fragments of different length will correspond to the positions of T. Four such sets of *chain-terminated fragments* (one for each dideoxy analog) then undergo electrophoresis, and the base sequence of the new DNA is read from the autoradiogram of the four lanes.

Fluorescence detection is a highly effective alternative to autoradiography because it eliminates the use of radioactive reagents and can be readily automated. A fluorescent tag is incorporated into each dideoxy analog—a differently colored one for each of the four chain terminators (e.g., a blue emitter for termination at A and a red one for termination at C). With the use of a mixture of terminators, a single reaction can be performed and the resulting fragments are separated by a technique known as *capillary electrophoresis*, in which the mixture is passed through a very narrow tube at high voltage to achieve efficient separation within a short time. As the DNA fragments emerge from the capillary, they are detected by their fluorescence; the sequence of their colors directly gives the base sequence (Figure 5.5). Sequences of as many as 500 bases can be determined in



**Figure 5.4 Strategy of the chain-termination method for sequencing DNA.**

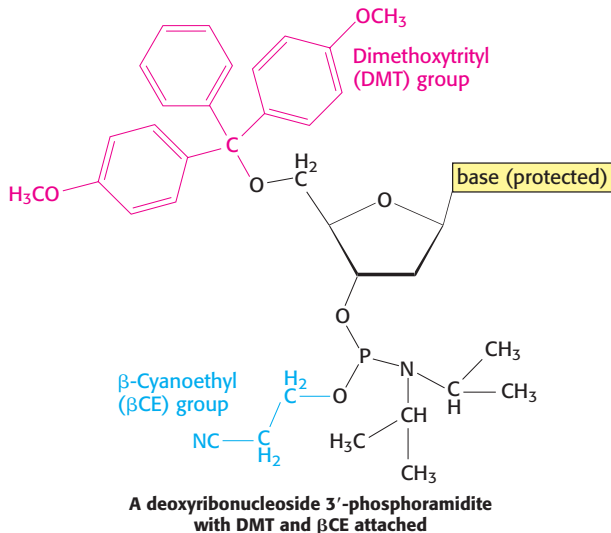
Fragments are produced by adding the 2',3'-dideoxy analog of a dNTP to each of four polymerization mixtures. For example, the addition of the dideoxy analog of dATP (shown in red) results in fragments ending in A. The strand cannot be extended past the dideoxy analog.



**Figure 5.5 Fluorescence detection of oligonucleotide fragments produced by the dideoxy method.** A sequencing reaction is performed with four chain-terminating dideoxy nucleotides, each labeled with a tag that fluoresces at a different wavelength (e.g., red for T). Each of the four colors represents a different base in a chromatographic trace produced by fluorescence measurements at four wavelengths. [After A. J. F. Griffiths et al., *An Introduction to Genetic Analysis*, 8th ed. (W. H. Freeman and Company, 2005).]

this way. Indeed, modern DNA-sequencing instruments can sequence more than 1 million bases per day with the use of this method.

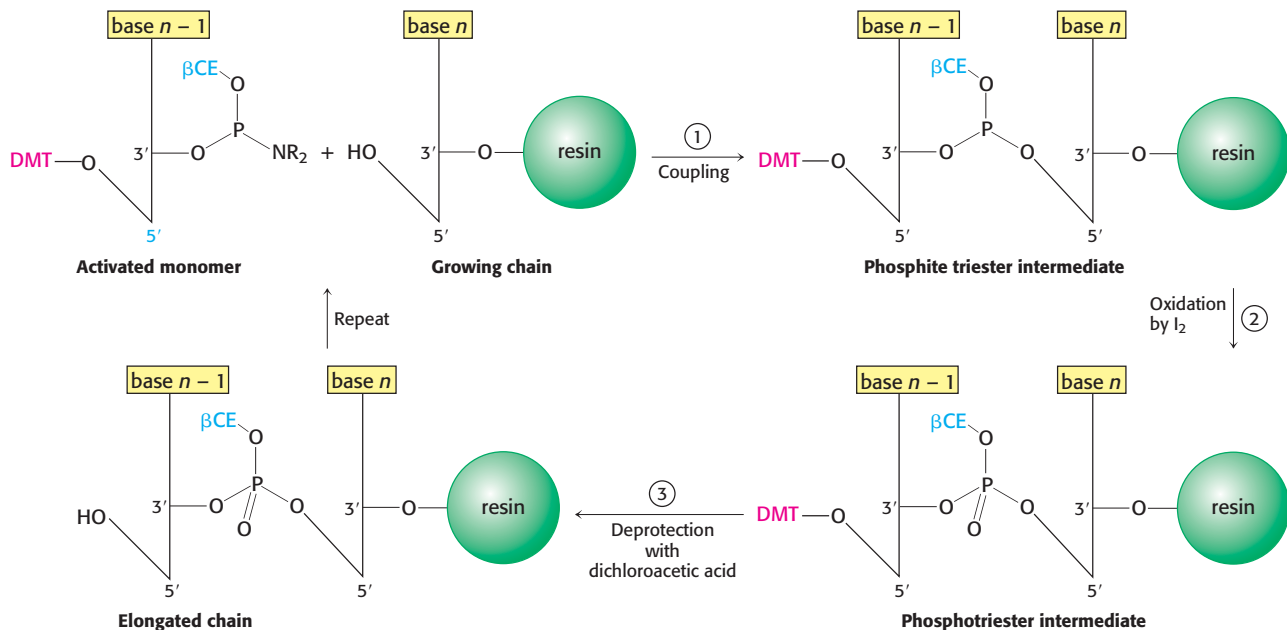
### DNA probes and genes can be synthesized by automated solid-phase methods



DNA strands, like polypeptides (Section 3.4), can be synthesized by the sequential addition of activated monomers to a growing chain that is linked to an insoluble support. The activated monomers are protected *deoxyribonucleoside 3'-phosphoramidites*. In step 1, the 3'-phosphorus atom of this incoming unit becomes joined to the 5'-oxygen atom of the growing chain to form a *phosphite triester* (Figure 5.6). The 5'-OH group of the activated monomer is unreactive because it is blocked by a dimethoxytrityl (DMT) protecting group, and the 3'-phosphoryl group is rendered unreactive by attachment of the β-cyanoethyl (βCE) group. Likewise, amino groups on the purine and pyrimidine bases are blocked.

Coupling is carried out under anhydrous conditions because water reacts with phosphoramidites. In step 2, the phosphite triester (in which P is trivalent) is oxidized by iodine to form a *phosphotriester* (in which P is pentavalent). In step 3, the DMT protecting group on the 5'-OH group of the growing chain is removed by the addition of dichloroacetic acid, which leaves other protecting groups intact. The DNA chain is now elongated by one unit and ready for another cycle of addition. Each cycle takes only about 10 minutes and usually elongates more than 99% of the chains.

This solid-phase approach is ideal for the synthesis of DNA, as it is for polypeptides, because the desired product stays on the insoluble support



**Figure 5.6 Solid-phase synthesis of a DNA chain by the phosphite triester method.** The activated monomer added to the growing chain is a deoxyribonucleoside 3'-phosphoramidite containing a dimethoxytrityl (DMT) protecting group on its 5'-oxygen atom, a β-cyanoethyl (βCE) protecting group on its 3'-phosphoryl oxygen atom, and a protecting group on the base.

until the final release step. All the reactions take place in a single vessel, and excess soluble reagents can be added to drive reactions to completion. At the end of each step, soluble reagents and by-products are washed away from the resin that bears the growing chains. At the end of the synthesis,  $\text{NH}_3$  is added to remove all protecting groups and release the oligonucleotide from the solid support. Because elongation is never 100% complete, the new DNA chains are of diverse lengths—the desired chain is the longest one. The sample can be purified by high-pressure liquid chromatography or by electrophoresis on polyacrylamide gels. DNA chains of as many as 100 nucleotides can be readily synthesized by this automated method.

The ability to rapidly synthesize DNA chains of any selected sequence opens many experimental avenues. For example, a synthesized oligonucleotide labeled at one end with  $^{32}\text{P}$  or a fluorescent tag can be used to search for a complementary sequence in a very long DNA molecule or even in a genome consisting of many chromosomes. The use of labeled oligonucleotides as DNA probes is powerful and general. For example, a DNA probe that can base-pair to a known complementary sequence in a chromosome can serve as the starting point of an exploration of adjacent uncharted DNA. Such a probe can be used as a *primer* to initiate the replication of neighboring DNA by DNA polymerase. An exciting application of the solid-phase approach is the synthesis of new tailor-made genes. New proteins with novel properties can now be produced in abundance by the expression of synthetic genes. Finally, the synthetic scheme heretofore described can be slightly modified for the solid-phase synthesis of RNA oligonucleotides, which can be very powerful reagents for the degradation of specific mRNA molecules in living cells by a technique known as RNA interference (Section 5.4).

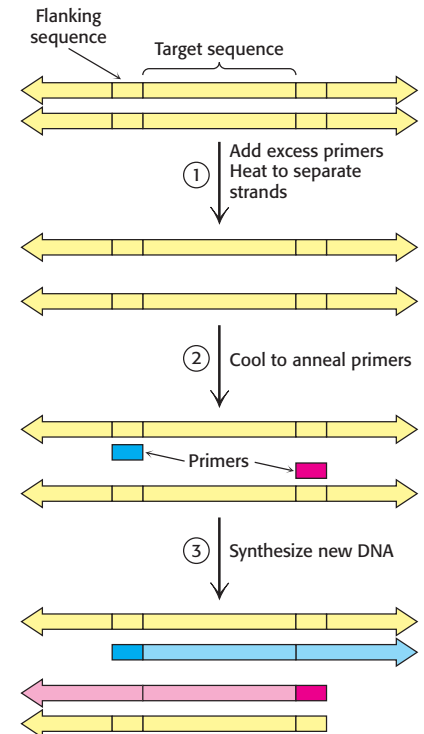
### Selected DNA sequences can be greatly amplified by the polymerase chain reaction

In 1984, Kary Mullis devised an ingenious method called the *polymerase chain reaction* (PCR) for amplifying specific DNA sequences. Consider a DNA duplex consisting of a target sequence surrounded by nontarget DNA. Millions of copies of the target sequences can be readily obtained by PCR if the flanking sequences of the target are known. PCR is carried out by adding the following components to a solution containing the target sequence: (1) a pair of primers that hybridize with the flanking sequences of the target, (2) all four deoxyribonucleoside triphosphates (dNTPs), and (3) a heat-stable DNA polymerase. A PCR cycle consists of three steps (Figure 5.7).

1. *Strand Separation.* The two strands of the parent DNA molecule are separated by heating the solution to  $95^\circ\text{C}$  for 15 s.

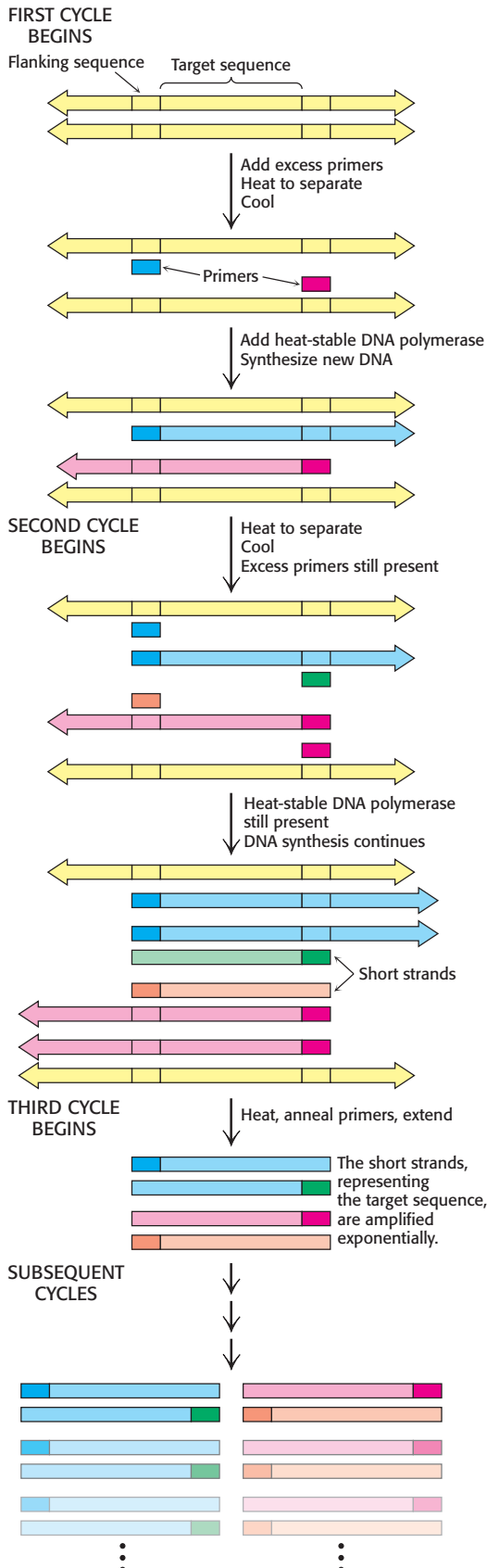
2. *Hybridization of Primers.* The solution is then abruptly cooled to  $54^\circ\text{C}$  to allow each primer to hybridize to a DNA strand. One primer hybridizes to the 3' end of the target on one strand, and the other primer hybridizes to the 3' end on the complementary target strand. Parent DNA duplexes do not form, because the primers are present in large excess. Primers are typically from 20 to 30 nucleotides long.

3. *DNA Synthesis.* The solution is then heated to  $72^\circ\text{C}$ , the optimal temperature for heat-stable polymerases. One such enzyme is *Taq* DNA polymerase, which is derived from *Thermus aquaticus*, a thermophilic bacterium that lives in hot springs. The polymerase elongates both primers in the direction of the target sequence because DNA synthesis is in the 5'-to-3'



**Figure 5.7** The first cycle in the polymerase chain reaction (PCR). A cycle consists of three steps: strand separation, the hybridization of primers, and the extension of primers by DNA synthesis.

direction. DNA synthesis takes place on both strands but extends beyond the target sequence.



These three steps—strand separation, hybridization of primers, and DNA synthesis—constitute one cycle of the PCR amplification and can be carried out repetitively just by changing the temperature of the reaction mixture. The thermostability of the polymerase makes it feasible to carry out PCR in a closed container; no reagents are added after the first cycle. At the completion of the second cycle, four duplexes containing the targeting sequence have been generated (Figure 5.8). Of the eight DNA strands comprising these duplexes, two short strands constitute only the target sequence—the sequence including and bounded by the primers. Subsequent cycles will amplify the target sequence exponentially. Ideally, after  $n$  cycles, the desired sequence is amplified  $2^n$ -fold. The amplification is a millionfold after 20 cycles and a billionfold after 30 cycles, which can be carried out in less than an hour.

Several features of this remarkable method for amplifying DNA are noteworthy. First, the sequence of the target need not be known. All that is required is knowledge of the flanking sequences so that complementary primers can be synthesized. Second, the target can be much larger than the primers. Targets larger than 10 kb have been amplified by PCR. Third, primers do not have to be perfectly matched to flanking sequences to amplify targets. With the use of primers derived from a gene of known sequence, it is possible to search for variations on the theme. In this way, families of genes are being discovered by PCR. Fourth, PCR is highly specific because of the stringency of hybridization at relatively high temperature. *Stringency* is the required closeness of the match between primer and target, which can be controlled by temperature and salt. At high temperatures, only the DNA between hybridized primers is amplified. A gene constituting less than a millionth of the total DNA of a higher organism is accessible by PCR. Fifth, PCR is exquisitely sensitive. A single DNA molecule can be amplified and detected.

### PCR is a powerful technique in medical diagnostics, forensics, and studies of molecular evolution

PCR can provide valuable diagnostic information in medicine. Bacteria and viruses can be readily detected with the use of specific primers. For example, PCR can reveal the presence of small amounts of DNA from the human immunodeficiency virus (HIV) in persons who have not yet mounted an immune response to this pathogen. In these patients, assays designed to detect antibodies against the virus would yield a false negative test result. Finding *Mycobacterium tuberculosis* bacilli in tissue specimens is slow and laborious. With PCR, as few as 10 tubercle bacilli per million human cells can be readily detected. PCR is a promising method for the early detection of certain cancers. This technique can identify mutations of certain growth-control genes, such as the *ras*

**Figure 5.8 Multiple cycles of the polymerase chain reaction.** The two short strands produced at the end of the third cycle (along with longer strands not shown) represent the target sequence. Subsequent cycles will amplify the target sequence exponentially and the parent sequence arithmetically.



genes (Chapter 14). The capacity to greatly amplify selected regions of DNA can also be highly informative in monitoring cancer chemotherapy. Tests using PCR can detect when cancerous cells have been eliminated and treatment can be stopped; they can also detect a relapse and the need to immediately resume treatment. PCR is ideal for detecting leukemias caused by chromosomal rearrangements.

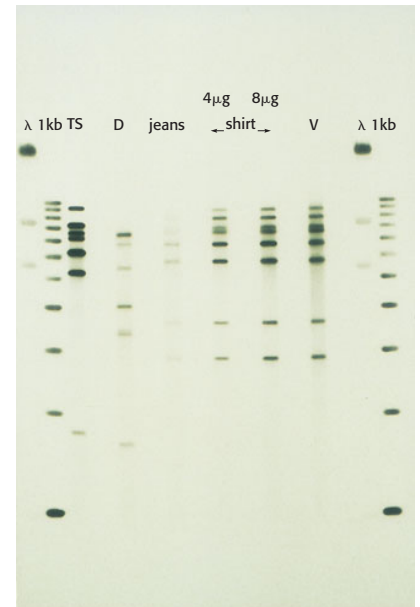
PCR is also having an effect in forensics and legal medicine. An individual DNA profile is highly distinctive because many genetic loci are highly variable within a population. For example, variations at one specific location determines a person's HLA type (human leukocyte antigen type; Section 34.5); organ transplants are rejected when the HLA types of the donor and recipient are not sufficiently matched. PCR amplification of multiple genes is being used to establish biological parentage in disputed paternity and immigration cases. Analyses of blood stains and semen samples by PCR have implicated guilt or innocence in numerous assault and rape cases. The root of a single shed hair found at a crime scene contains enough DNA for typing by PCR (Figure 5.9).

DNA is a remarkably stable molecule, particularly when shielded from air, light, and water. Under such circumstances, large fragments of DNA can remain intact for thousands of years or longer. PCR provides an ideal method for amplifying such ancient DNA molecules so that they can be detected and characterized (Section 6.5). PCR can also be used to amplify DNA from microorganisms that have not yet been isolated and cultured. As will be discussed in Chapter 6, sequences from these PCR products can be sources of considerable insight into evolutionary relationships between organisms.

### The tools for recombinant DNA technology have been used to identify disease-causing mutations

Let us consider how the techniques just described have been utilized in concert to study ALS, introduced at the beginning of this chapter. Five percent of all patients suffering from ALS have family members who also have been diagnosed with the disease. A heritable disease pattern is indicative of a strong genetic component of disease causation. To identify these disease-causing genetic alterations, researchers identify *polymorphisms* (instances of genetic variation) within an affected family that correlate with the emergence of disease. Polymorphisms may themselves cause disease or be closely linked to another genetic alteration that does. One class of polymorphisms are *restriction-fragment-length polymorphisms* (RFLPs), which are mutations within restriction sites that change the sizes of DNA fragments produced by the appropriate restriction enzyme. Using restriction digests and Southern blots of the DNA from members of ALS-affected families, researchers identified RFLPs that were found preferentially in those family members with a positive diagnosis. For some of these families, strong evidence was obtained for the disease-causing mutation within a specific region of chromosome 21.

After the probable location of one disease-causing gene had been identified, this same research group compared the locations of the ALS-associated RFLPs with the known sequence of chromosome 21. They noted that this chromosomal locus contains the *SOD1* gene, which encodes the Cu/Zn superoxide dismutase protein SOD1, an enzyme important for the protection of cells against oxidative damage (Section 18.3). PCR amplification of regions of the *SOD1* gene from the DNA of affected family members, followed by Sanger dideoxy sequencing of the targeted fragment, enabled the identification of 11 disease-causing mutations from 13 different families.



**Figure 5.9 DNA and forensics.** DNA isolated from bloodstains on the pants and shirt of a defendant was amplified by PCR, then compared with DNA from the victim as well as the defendant by using gel electrophoresis and autoradiography. DNA from the bloodstains on the defendant's clothing matched the pattern of the victim but not that of the defendant. The frequency of a coincidental match of the DNA pattern on the clothing and the victim is approximately 1 in 33 billion. Lanes λ, 1kb, and TS refer to control DNA samples; lane D, DNA from the defendant; jeans and shirt, DNA isolated from bloodstains on defendant's pants and shirt (two different amounts analyzed); V, DNA sample from victim's blood. [Courtesy of Cellmark Diagnostics, Germantown, Maryland.]

This work was pivotal for focusing further inquiry into the roles that superoxide dismutase and its corresponding mutant forms play in the pathology of ALS.

## 5.2 Recombinant DNA Technology Has Revolutionized All Aspects of Biology

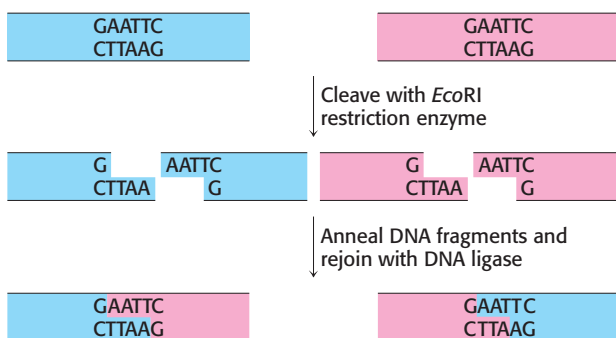
The pioneering work of Paul Berg, Herbert Boyer, and Stanley Cohen in the early 1970s led to the development of recombinant DNA technology, which has taken biology from an exclusively analytical science to a synthetic one. New combinations of unrelated genes can be constructed in the laboratory by applying recombinant DNA techniques. These novel combinations can be cloned—amplified many-fold—by introducing them into suitable cells, where they are replicated by the DNA-synthesizing machinery of the host. The inserted genes are often transcribed and translated in their new setting. What is most striking is that the genetic endowment of the host can be permanently altered in a designed way.

### Restriction enzymes and DNA ligase are key tools in forming recombinant DNA molecules

Let us begin by seeing how novel DNA molecules can be constructed in the laboratory. An essential tool for the manipulation of recombinant DNA is a *vector*, a DNA molecule that can replicate autonomously in an appropriate host organism. Vectors are designed to enable the rapid, covalent insertion of DNA fragments of interest. *Plasmids* (naturally occurring circles of DNA that act as accessory chromosomes in bacteria) and bacteriophage lambda ( $\lambda$  phage), a virus, are choice vectors for cloning in *E. coli*. The vector can be prepared for accepting a new DNA fragment by cleaving it at a single specific site with a restriction enzyme. For example, the plasmid pSC101, a 9.9-kb double-helical circular DNA molecule, is split at a unique site by the *EcoRI* restriction enzyme. The staggered cuts made by this enzyme produce *complementary single-stranded ends*, which have specific affinity for each other and hence are known as *cohesive* or *sticky ends*. Any DNA fragment can be inserted into this plasmid if it has the same cohesive ends. Such a fragment can be prepared from a larger piece of DNA by using the same restriction enzyme as was used to open the plasmid DNA (Figure 5.10).

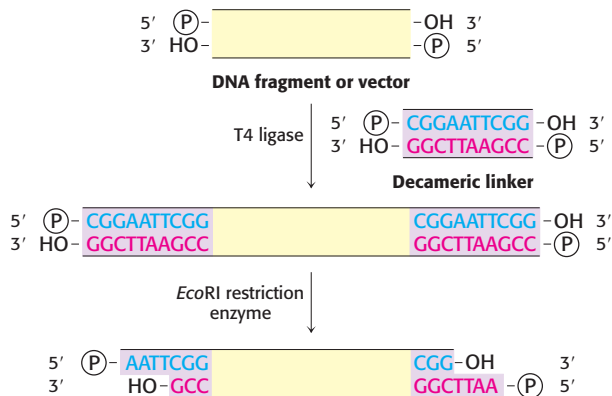
The single-stranded ends of the fragment are then complementary to those of the cut plasmid. The DNA fragment and the cut plasmid can be annealed and then joined by *DNA ligase*, which catalyzes the formation of a phosphodiester bond at a break in a DNA chain. DNA ligase requires a free 3'-hydroxyl group and a 5'-phosphoryl group. Furthermore, the chains joined by ligase must be in a double helix. An energy source such as ATP or  $\text{NAD}^+$  is required for the joining reaction, as will be discussed in Chapter 28.

What if the target DNA is not naturally flanked by the appropriate restriction sites? How is the fragment cut and annealed to the vector? The cohesive-end method for joining DNA molecules can still be used in these cases by adding a *short, chemically synthesized DNA linker* that can be cleaved by restriction enzymes. First, the linker is covalently joined to the ends of a DNA fragment. For example,



**Figure 5.10 Joining of DNA molecules by the cohesive-end method.** Two DNA molecules, cleaved with a common restriction enzyme such as *EcoRI*, can be ligated to form recombinant molecules.

the 5' ends of a decameric linker and a DNA molecule are phosphorylated by polynucleotide kinase and then joined by the ligase from T4 phage (Figure 5.11). This ligase can form a covalent bond between blunt-ended (flush-ended) double-helical DNA molecules. Cohesive ends are produced when these terminal extensions are cut by an appropriate restriction enzyme. Thus, *cohesive ends corresponding to a particular restriction enzyme can be added to virtually any DNA molecule*. We see here the fruits of combining enzymatic and synthetic chemical approaches in crafting new DNA molecules.



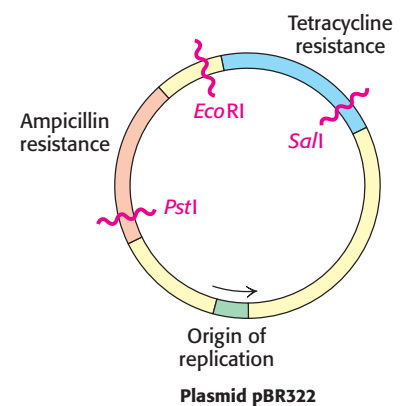
**Figure 5.11 Formation of cohesive ends.**

Cohesive ends can be formed by the addition and cleavage of a chemically synthesized linker.

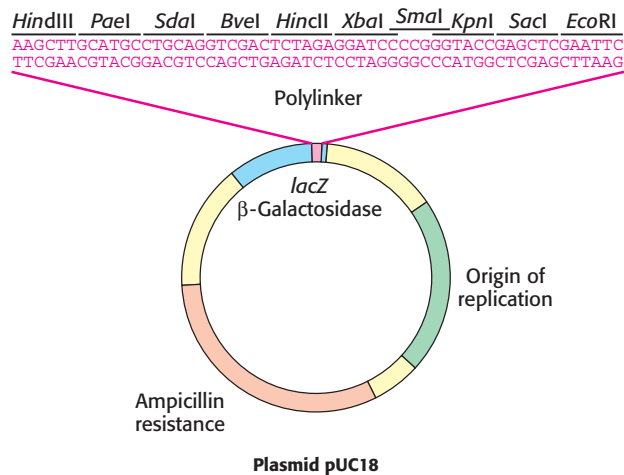
### Plasmids and lambda phage are choice vectors for DNA cloning in bacteria

Many plasmids and bacteriophages have been ingeniously modified by researchers to enhance the delivery of recombinant DNA molecules into bacteria and to facilitate the selection of bacteria harboring these vectors. As already mentioned, plasmids are circular double-stranded DNA molecules that occur naturally in some bacteria. They range in size from two to several hundred kilobases. Plasmids carry genes for the inactivation of antibiotics, the production of toxins, and the breakdown of natural products. These accessory chromosomes can replicate independently of the host chromosome. In contrast with the host genome, they are dispensable under certain conditions. A bacterial cell may have no plasmids at all or it may house as many as 20 copies of a plasmid.

Many plasmids have been optimized for a particular experimental task. For example, one class of plasmids, known as *cloning vectors*, is particularly suitable for the rapid insertion and replication of a collection of DNA fragments. The creative placement of antibiotic-resistance genes or reporter genes or both within these plasmids enables the rapid identification of those vectors that harbor the desired DNA insert. For example, in *pBR322*, one of the first plasmids used for this purpose, insertion of DNA at the *SalI* or *BamHI* restriction site (Figure 5.12) inactivates the gene for tetracycline resistance, an effect called *insertional inactivation*. Cells containing *pBR322* with a DNA insert at one of these restriction sites are resistant to ampicillin but sensitive to tetracycline, and so they can be readily selected. Another class of plasmids have been optimized for use as *expression vectors* for the production of large amounts of protein. In addition to antibiotic-resistance genes, they contain promoter sequences designed to drive the transcription of large amounts of a protein-coding DNA sequence. Often, these vectors contain sequences flanking the cloning site that simplify the addition of



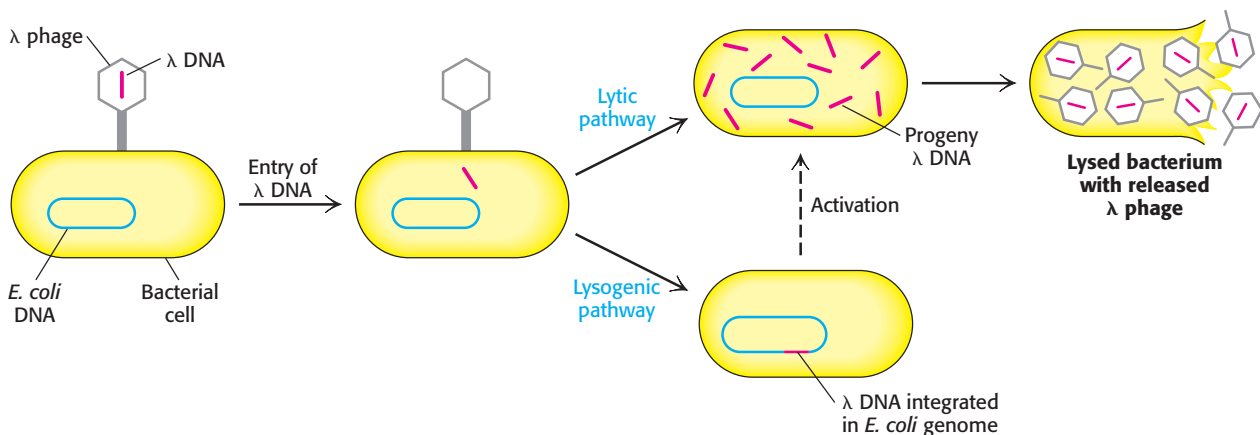
**Figure 5.12 Genetic map of the plasmid *pBR322*.** This plasmid carries two genes for antibiotic resistance. Like all other plasmids, it is a circular duplex DNA.



**Figure 5.13 A polylinker in the plasmid pUC18.** The plasmid pUC18 includes a polylinker within a gene for  $\beta$ -galactosidase (often called the *lacZ* gene). Insertion of a DNA fragment into one of the many restriction sites within this polylinker can be detected by the absence of  $\beta$ -galactosidase activity.

fusion tags to the protein of interest (Section 3.1), greatly facilitating the purification of the overexpressed protein. Both types of plasmid vectors often feature a *polylinker* region that includes many unique restriction sites within its sequence (Figure 5.13). This polylinker can be cleaved with many different restriction enzymes or combinations of enzymes, providing great versatility in the DNA fragments that can be inserted.

Another widely used vector,  $\lambda$  phage, enjoys a choice of life styles: this bacteriophage can destroy its host or it can become part of its host (Figure 5.14). In the *lytic pathway*, viral functions are fully expressed: viral DNA and proteins are quickly produced and packaged into virus particles, leading to the lysis (destruction) of the host cell and the sudden appearance of about 100 progeny virus particles, or *virions*. In the *lysogenic pathway*, the phage DNA becomes inserted into the host-cell genome and can be replicated together with host-cell DNA for many generations, remaining inactive. Certain environmental changes can trigger the expression of this dormant viral DNA, which leads to the formation of progeny viruses and lysis of the host. Large segments of the 48-kb DNA of  $\lambda$  phage are not essential for productive infection and can be replaced by foreign DNA, thus making  $\lambda$  phage an ideal vector.



**Figure 5.14 Alternative infection modes for  $\lambda$  phage.** Lambda phage can multiply within a host and lyse it (lytic pathway) or its DNA can become integrated into the host genome (lysogenic pathway), where it is dormant until activated.