

Chapter 7

Bacteriophage

Bacteriophage or **phage** for short are viruses that infect only bacteria. In contrast to cells that grow from an increase in the number of their components and reproduce by division, **viruses** are assembled from pre-made components. Viruses are nucleic acid molecules surrounded by a protective coating. They are not capable of generating energy and reproduce inside of cells. The nucleic acid inside the coating, called the **phage genome** in a bacteriophage, encodes most of the gene products needed for making more phage. The phage genome can be made of either double- or single-stranded DNA or RNA, depending on the bacteriophage in question. The genome can be circular or linear. The protective coating or **capsid** surrounding the phage genome is composed of phage-encoded proteins.

Many important discoveries have been made using phage as model systems. From the discovery that a nonsense codon stopped protein synthesis to the first developmental switch to be understood at the molecular level, phage have proven to be very useful. In this chapter, we will look at phage development using T4, λ (lambda), P1, and M13 as examples. Each of these phage infect *E. coli*. We will examine specific discoveries using these phage or specific properties of the phage that have made them particularly useful to biologists.

The structure of phage

All phage have a chromosome encased in a capsid that is composed of phage-encoded proteins. For many phage types, the capsid is attached to a tail structure that is also made from phage-encoded proteins. T4 and P1 contain a linear double-stranded DNA genome enclosed in a capsid and attached to a tail (Fig. 7.1a). The T4 genome is 172 kb, while P1 is a smaller phage with a genome of 90 kb. The T4 capsid is an elongated icosahedron. T4 has a very elaborate tail structure including a collar at the base of the head and a rigid tail core surrounded by a contractile sheath. The core and sheath are attached to a hexagonal base plate. Also attached to the tail plate are tail pins and six kinked tail fibers. P1 also has an icosahedral capsid, a tail with a contractile sheath, a base plate, and tail fibers. λ contains a linear double-stranded DNA genome of 48.5 kb, a capsid, and a tail (Fig. 7.1b). The finished capsid is again shaped like an icosahedron whereas the tail is a thin flexible tube that ends in a small conical part and a single tail fiber. M13 contains a circular single-stranded DNA genome of 6407 nucleotides surrounded by five phage-encoded proteins (Fig. 7.1c). The M13 chromosome is coated

FYI 7.1

Discovery of phage

Phage were first described in 1915 by Frederick Twort and 1917 by Felix d'Herelle. Both men discovered phage when the bacteria they were working with lysed. The agent responsible for the lysis was transferable from culture to culture, invisible by light microscopy, and would go through the smallest filter they had. d'Herelle coined the term "bacteriophage", signifying an entity that eats bacteria.

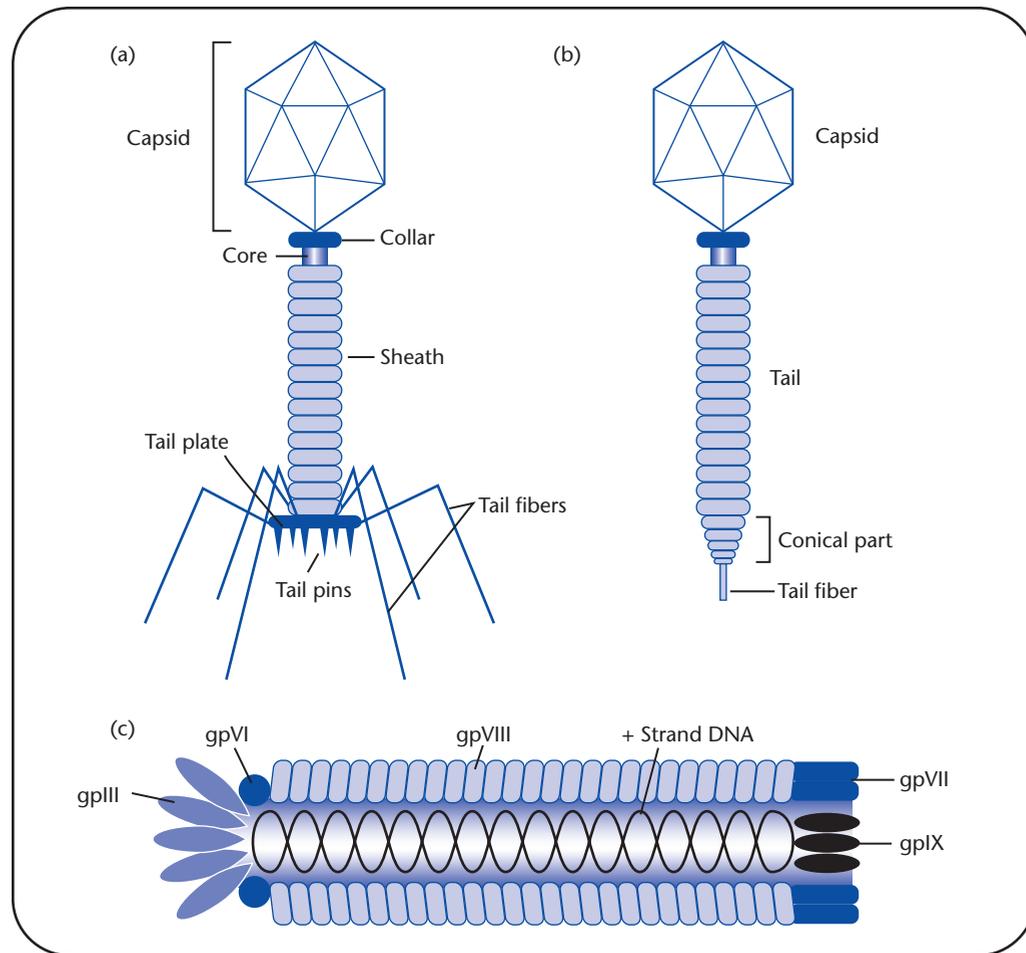


Fig. 7.1 The structures of (a) T4, (b) λ , and (c) M13.

by a single layer of ~2700 subunits of gene VIII encoded protein (gpVIII) giving it a filamentous appearance, the reason M13 is also known as a **filamentous phage**. At one end of the filament are bound the M13 proteins encoded by genes VII and IX (gpVII and gpIX) and at the other end are bound the M13-encoded gene III and VI proteins (gpIII and gpVI).

The lifecycle of a bacteriophage

All phage must carry out a specific set of reactions in order to make more of themselves. First, the phage must be able to recognize a bacterium that it can multiply in by binding to the bacterial cell surface. Next, the phage must inject its genome and the genome must be protected from the bacterial nucleases in the cytoplasm. The phage genome must be replicated, transcribed, and translated so that a large number of genomes, capsid proteins, and tail proteins, if present, are produced at the same or nearly the same time. Complete phage particles are then assembled and the phage must get back out of the bacterium. Different phage use different strategies to carry out each of these reactions.

Phage are very choosy as to what bacteria they infect. This is referred to as the **host range** of the phage. For example, λ only infects certain *E. coli*, whereas Spo1 phage infect only *Bacillus subtilis*. Several phage types may infect a single bacterial species. *E. coli* can be infected by λ , M13, P1, T4, and Mu phages, to name a few.

The number of phage that can be released from one bacterium after infection and growth by one phage is known as the **burst size**. Every phage has a characteristic burst size. Different phage also take different amounts of time to go through one growth cycle. We know when a phage has successfully reproduced when we are able to detect **plaques** or circular areas with little or no bacterial growth on an agar plate covered with a thin layer of bacteria.

Once bound to the cell, the phage must get its genome into the cytoplasm. The rate of phage DNA transport can be very rapid. It is different for different phages but can reach values as high as 3000 base pairs per second. In contrast, two other methods for getting DNA from the outside of the cell to the cytoplasm (conjugation, Chapter 10 and transformation, Chapter 11) transfer the DNA at a rate of approximately 100 bases per second. In many cases the details of how a phage genome gets into the cytoplasm are not known. From the information we do have, it is clear that not just one mechanism is used.

Lytic–Lysogenic options

The process of a phage infecting a bacterium and producing progeny is referred to as a **lytic** infection. Some phage, like T4, are only capable of lytic growth. Some phage are also capable of maintaining their chromosome in a stable, silent state within the bacteria. This is called **lysogeny**. Phage that are capable of both a lytic and lysogenic pathway are called **temperate phage**. P1 and λ are temperate phage. M13 is unusual because phage continually exit from a bacterium without killing it. For this reason, M13 is not considered to have a true lysogenic state and is not a temperate phage. When the bacterium contains a silent phage chromosome, it is referred to as a **lysogen**. The incorporated phage genome is referred to as a **prophage**.

The λ lifecycle

λ adsorption

Phage identify a host bacterium by binding or adsorbing to a specific structure on the surface of the cell. Many different cell surface structures can be used as binding sites. The basics of adsorption are that a specific structure on the surface of the phage interacts with a specific structure on the surface of the bacterium. λ binds to an outer membrane protein called LamB via a protein that resides at the tip of the λ tail called the J protein. LamB normally functions in the binding and uptake of the sugars maltose and maltodextrin.

λ DNA injection

Initially, λ binds to LamB and the binding is reversible. This step requires only the λ tail and the LamB protein. Next, the bound phage undergoes a change and the binding to LamB becomes irreversible. The nature of the change is unknown but it requires

that a phage head be attached to the phage tail. Next the λ DNA is ejected from the phage and taken up by the bacterium. The DNA in the phage head is very tightly packed. If the condensed state of the phage DNA is stabilized, ejection of the DNA does not occur. The addition of small positively charged molecules such as putrescine to the phage counteracts the negatively charged DNA and stabilizes the DNA in the phage head. This implies that the tight packing of the DNA is used to help eject the DNA from the phage particle. When λ DNA is put into the capsid, one end known as the left end is inserted first. When the λ DNA comes out of the phage head, the right end exits first. Unlike phage T4 (see below), no change in the λ tail structure is seen

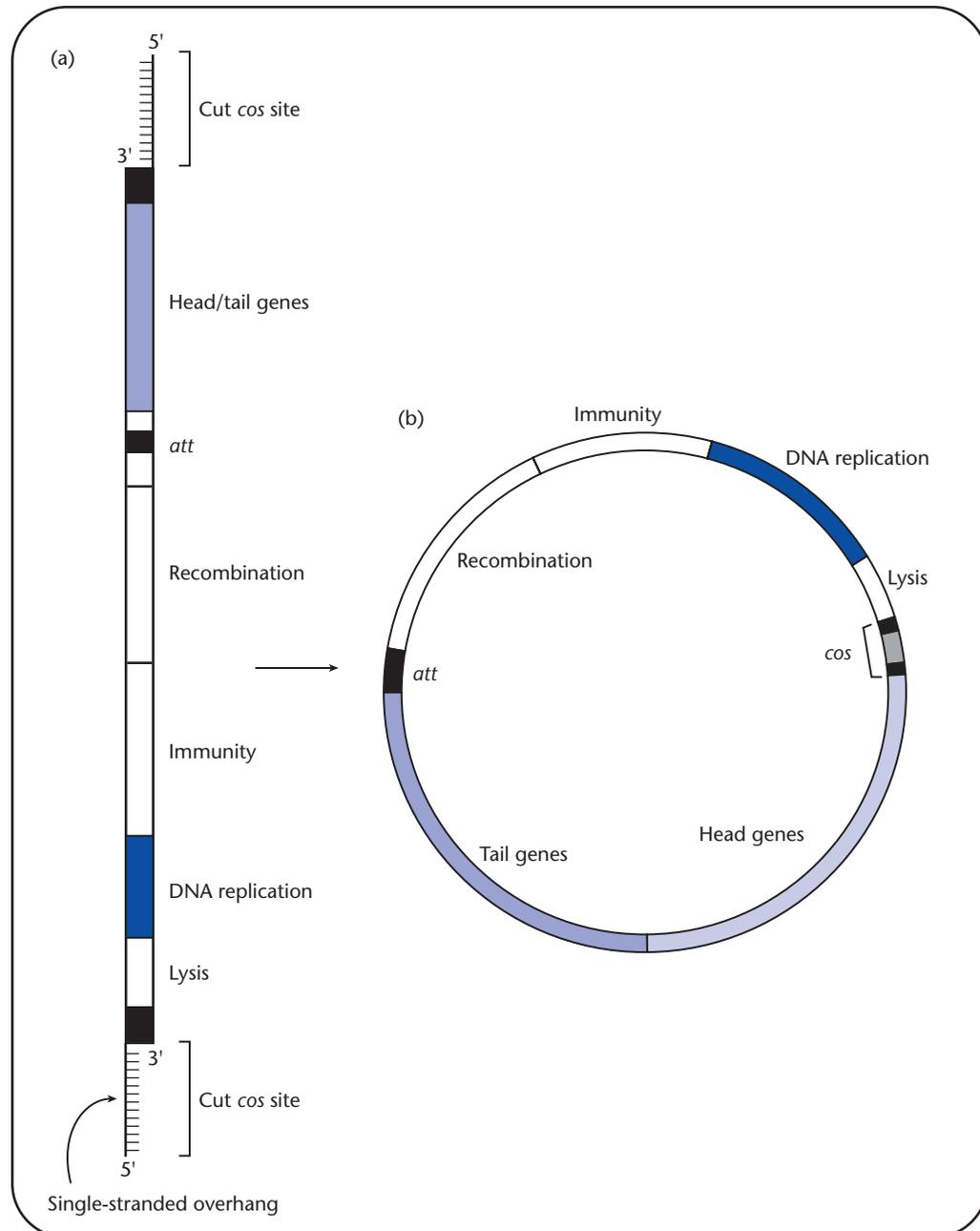


Fig. 7.2 The structure of the λ DNA in the phage capsid (a) and after circularization in the cytoplasm (b). The DNA circularizes via the *cos* site.

when the DNA is ejected. In addition to LamB, λ also uses an inner membrane protein called PstM to gain entry to the cytoplasm. How the λ DNA physically traverses the peptidoglycan and periplasm and gets through PstM is not known.

Protecting the λ genome in the bacterial cytoplasm

What protection the phage genome needs in the cytoplasm depends on the physical state of the injected nucleic acid. λ contains a linear double-stranded DNA molecule in its capsid. In the bacterial cytoplasm, dsDNA molecules are subject to degradation by exonucleases that need a free end to digest the DNA. The first event that happens to newly injected λ DNA is that the DNA circularizes to prevent it from being degraded.

λ has a specific site on its DNA, termed the *cos* site, which it uses to circularize the DNA (Fig. 7.2). The *cos* site is a 22 bp sequence that is cut asymmetrically when the λ DNA is packaged (see below). The cut *cos* site has a 12 bp overhang. There is one cut *cos* site at the left end of the λ genome and another cut *cos* site at the right end of the λ genome (Fig. 7.2a). When the λ DNA is injected into the cytoplasm, the cut *cos* sites at either end of the linear λ genome anneal (Fig. 7.2b). A host enzyme, DNA ligase, seals the nicks at either end of the *cos* site generating a covalently closed, circular λ genome. The host encoded enzyme, DNA gyrase, supercoils the λ molecule.

What happens to the λ genome after it is stabilized?

The λ genome contains six major promoters known as P_L for promoter leftward, P_R for promoter rightward, P_{RE} for promoter for repressor establishment, P_{RM} for promoter for repressor maintenance, P_I for promoter for integration, and $P_{R'}$ for secondary rightward promoter (Fig. 7.3). After the genome is circularized and supercoiled, transcription begins from P_L and P_R . A series of genes known as early genes are transcribed and translated. These gene products are the initial proteins needed for further phage development. *E. coli* RNA polymerase interacts with P_L to give rise to a short mRNA transcript that is translated into the N protein (Fig. 7.4a). *E. coli* RNA polymerase interacts with P_R to give rise to a short mRNA transcript that is translated into the Cro protein (Fig. 7.4a).

The N protein is able to extend transcription when RNA polymerase encounters a sequence in the DNA that tells it to stop. For this reason, N is called an anti-termination protein. N allows RNA polymerase to transcribe through the t_L and t_{R1} termination signals resulting in the synthesis of longer mRNA transcripts (Fig. 7.4b). The longer transcripts from P_R encode the O, P, and CII proteins, and a small amount of another anti-terminator, the Q protein. From P_L , CIII, the recombination proteins Gam and Red and a small amount of Xis and Int are made.

The N protein anti-terminates by binding to RNA polymerase after a specific base pair sequence, located upstream of the transcriptional termination site, has been

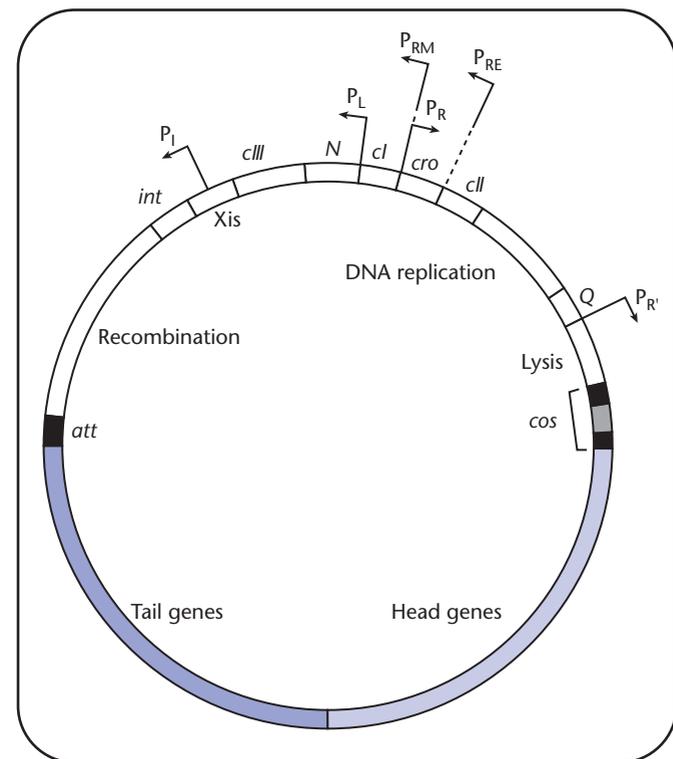


Fig. 7.3 The location of the six major promoters on the λ genome and the direction in which they specify mRNA production.

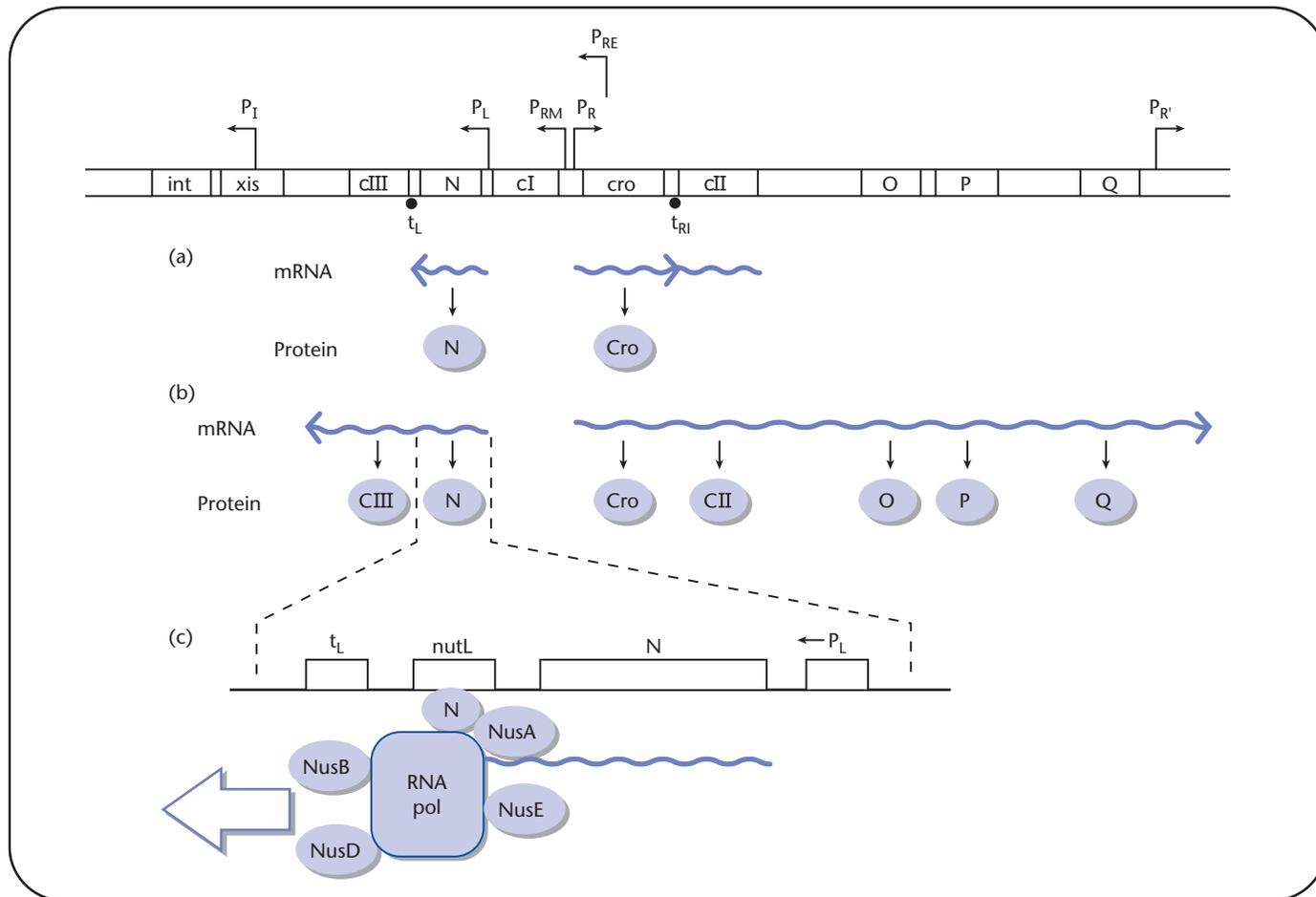


Fig. 7.4 The first transcription and translation events that take place on the λ genome after infection. (a) Transcription from P_L leads to the production of N protein. Transcription from P_R leads to Cro protein. (b) N is an anti-terminator that allows RNA polymerase to read through the t_L and t_{RI} terminators. From P_L , N and CIII proteins will

be produced. From P_R , Cro, CII, O, P, and Q proteins will be produced. (c) N binds to the *nutL* site on the DNA. In conjunction with four bacterial proteins, NusA, NusB, NusD, and NusE, N allows RNA polymerase to read through the terminator t_L .

transcribed into mRNA (Fig. 7.4c). This sequence is called *nut* for N ut ilization. Other *E. coli* proteins contribute to anti-termination. These proteins have been named Nus, for N ut ilization substance.

At this point, all of the players needed to make the lytic-lysogenic decision have been made. CII and CIII are needed for lysogenic growth. Cro and Q are needed for lytic growth. The O and P proteins are used for replicating the λ DNA.

λ and the lytic-lysogenic decision

The decision between lytic or lysogenic growth for λ was the first developmental switch understood at the molecular level. At the most basic level, the decision depends on the amounts of two phage-encoded proteins called CI (pronounced C-one) and Cro, and their binding to their promoter control regions (Fig. 7.5). When CI is bound, the expression of the lytic genes is repressed and the phage follows the

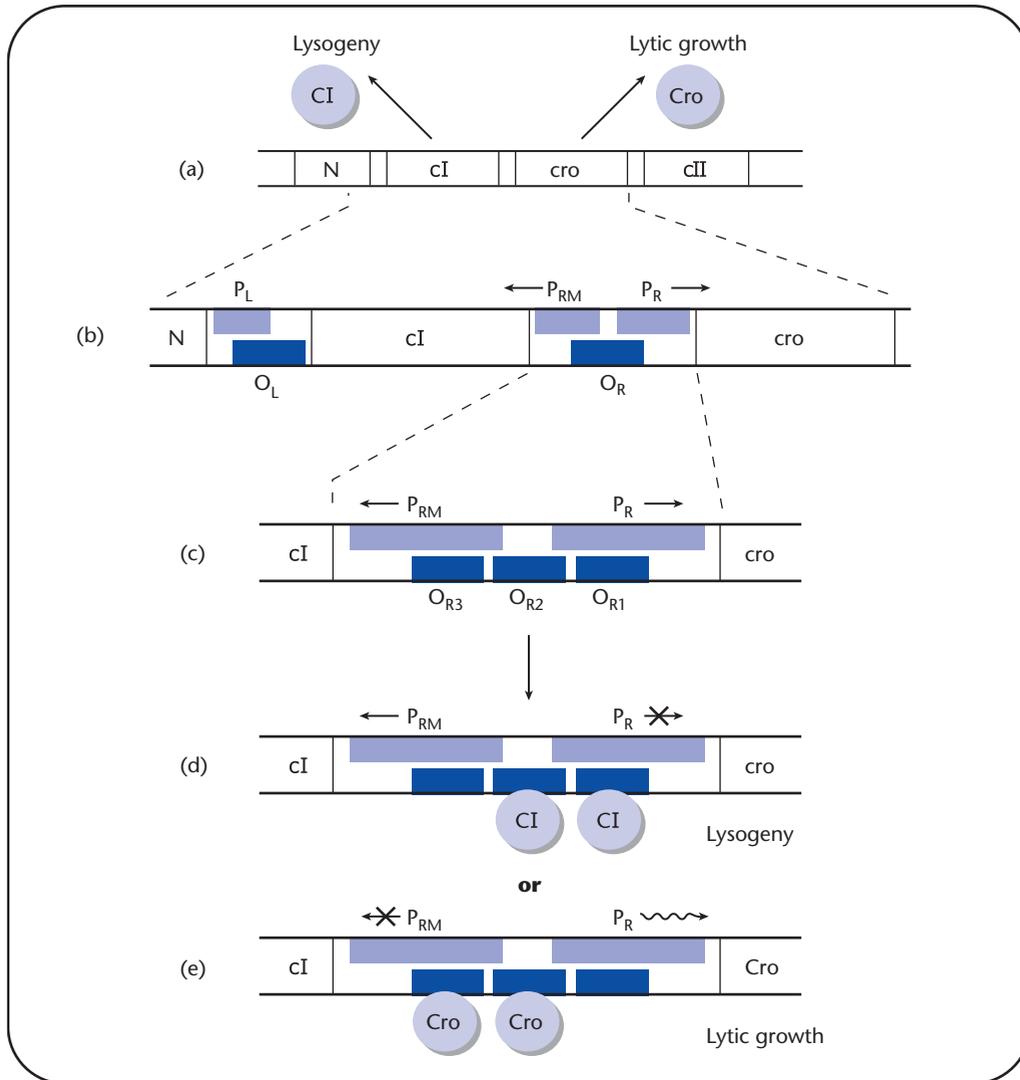


Fig. 7.5 CI and Cro are the proteins responsible for the two developmental fates of λ . (a) CI leads to lysogeny and Cro leads to lytic growth. (b) Both CI and Cro bind to two operator regions, O_R and O_L. O_R overlaps with both P_R and P_{RM}. O_L overlaps with P_L. (c) O_R is required for the switch between developmental pathways. It is composed of three 17 base pair sequences called O_{R1}, O_{R2}, and O_{R3}. They are similar in sequence but not identical. (d) CI binds to O_{R1} first then O_{R2}. It will bind to O_{R3} but only at very high concentrations. When CI binds to O_R, it represses transcription from P_R and activates it from P_{RM}. CI binding to O_R is actually required for P_{RM} to be activated. CI binding leads to lysogeny. (e) Cro also binds to O_{R1}, O_{R2}, and O_{R3} but in the opposite order from CI. Cro binds to O_{R3} first then O_{R2} and at high concentrations O_{R1}. Cro binding to O_{R3} inhibits P_{RM} and leads to lysogeny.

lysogenic pathway (Fig. 7.5a). For this reason, CI is also known as CI repressor or λ repressor. The expression and binding of Cro leads to lytic development.

Cro is made from P_R and CI is made from either P_{RE} or P_{RM}. Both Cro and CI bind to the same DNA sequences called operators (Fig. 7.5b). λ contains two operators that bind Cro and CI. One, called O_R, overlaps the P_{RM} and P_R promoters. The other, called

O_L is behind the P_L promoter. O_R is a major player in the lytic–lysogenic decision, while O_L is not part of the decision.

O_R is composed of three 17 base pair sequences called O_{R1} , O_{R2} , and O_{R3} (Fig. 7.5c). CI repressor binds to O_{R1} 10 times better than it binds to O_{R2} or O_{R3} . At increasing concentrations of CI, it will bind to O_{R2} and eventually to O_{R3} . When CI is bound to O_R , it stimulates the P_{RM} promoter and the production of CI repressor and inhibits the P_R promoter and the production of Cro, leading to lysogeny (Fig. 7.5d). Cro also binds to O_{R1} , O_{R2} , and O_{R3} but in the reverse order from CI repressor. Cro binds to O_{R3} first, then O_{R2} , and finally at high concentrations to O_{R1} . When Cro is bound to O_R , it inhibits the P_{RM} promoter and the production of CI, leading to lytic growth (Fig. 7.5e). This is the basis for either lytic or lysogenic growth.

How does the phage switch between these two developmental pathways? The major protein involved in the switch is another phage-encoded protein called CII (pronounced C-two, Fig. 7.6). CII activates the P_{RE} and P_I promoters. This leads to the production of repressor and the Integrase protein, which is also needed for lysogeny (Fig. 7.6b). The gene for CII (*cII*) resides just to the right of the *cro* gene. When λ infects a cell, transcription automatically begins from P_L and P_R using host proteins. Transcription from P_R leads to production of both the Cro and CII proteins. If CII is active it will lead to production of CI and Integrase and lysogeny. If CII is inactive then Cro will repress P_{RM} , preventing expression of CI and leading to lytic growth.

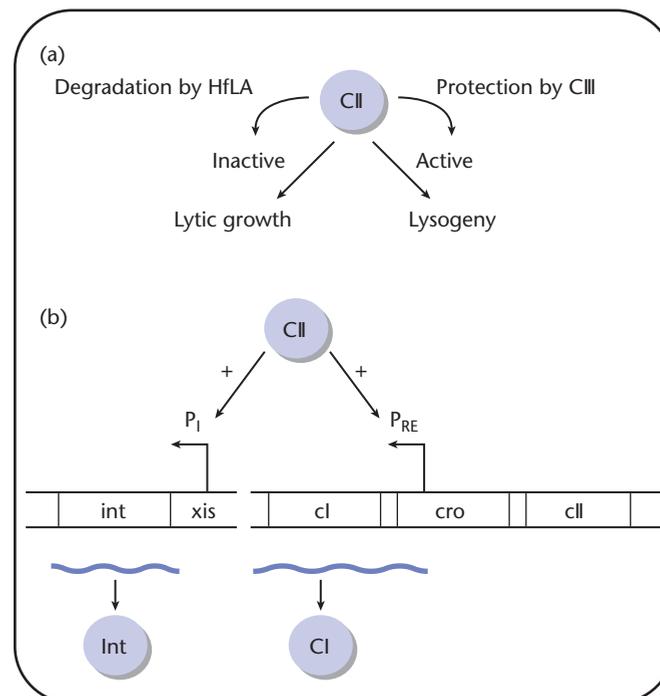
The CII protein is inherently unstable. Several factors influence this feature of the protein. CII is degraded by the bacterial-encoded HflA protease. When cells are actively growing in nutrient-rich conditions, the amount of HflA in the cell is high, leading to degradation of CII and lytic growth. When cells are growing slowly, HflA levels are low, leading to stabilization of CII, production of CI, and lysogeny. In this manner, CII is used to monitor the health of the cell and impact the lytic–lysogenic decision accordingly. It is thought that λ wants to produce more phage when cells are healthy, nutrients are plentiful, and the prospect of completing phage development

is good. Lysogeny is a better bet when cells are growing poorly. CII is also stabilized by a phage-encoded protein called CIII. CIII is produced from P_L by infecting phage.

The λ lysogenic pathway

If CII prevails, CI will be produced, initially from the P_{RE} promoter and eventually from the P_{RM} promoter. CI activates P_{RM} ensuring that a continuous supply of CI is made. CI also activates the P_I promoter, leading to the production of the Integrase protein. The recombination of λ DNA into the chromosome occurs at a

Fig. 7.6 The CII protein is the major player in the switch between lytic and lysogenic growth. CII is unstable and rapidly degraded by the host-encoded HflA protease. Inactive CII leads to lytic growth. CII can be protected by the phage-encoded CIII protein. Active CII leads to lysogenic growth.



specific site in the λ DNA called *attP* and at a specific site in the bacterial chromosome called *attB* (Fig. 7.7). The recombination of λ DNA into the chromosome requires Integrase and the host-encoded IHF protein (for *integration host factor*). Once in the chromosome, the phage DNA is bounded by hybrid *att* sites called *attL* and *attR*. The reverse of this reaction, recombination of λ DNA out of the chromosome requires Int, IHF, and a third protein Xis (for excision and pronounced excise). Because the recombination always occurs at specific sites and requires very specific enzymes, it is known as a **site-specific recombination** event (Chapter 5). Once the λ DNA is recombined into the chromosome, it is replicated and stably inherited by daughter cells as part of the bacterial chromosome. The *attB* site on the chromosome lies between the *gal* and *bio* genes and does not disrupt either gene. When λ DNA has recombined into the bacterial chromosome it is quiescent, except for the continued production of CI from P_{RM} .

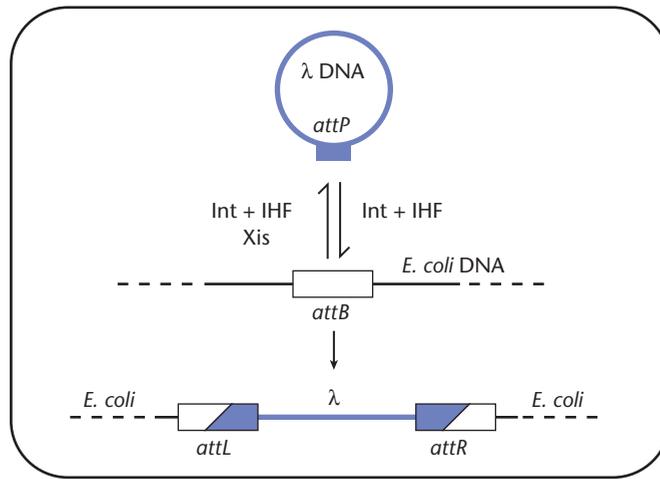


Fig. 7.7 λ recombines into the chromosome using a specific site on the phage called *attP* and a specific site on the bacterial chromosome called *attB*. When the λ DNA is in the chromosome, it is bounded by *attL* and *attR*, which are hybrid *attP/attB* sites.

What prevents the expression of the late genes coding for lytic function? The expression of late genes is prevented by the action of the λ repressor. λ repressor binding to the operator sequences O_R and O_L blocks transcription from P_L and P_R . Since P_R is blocked, the λ Q protein is not made and transcription of the late genes does not occur.

The λ lytic pathway

If enough of the Q protein accumulates in the cell, RNA polymerase will continue its transcription from a third promoter, $P_{R'}$, located in front of the Q gene (Fig. 7.8). This extends transcription into the late genes located downstream of Q. The late genes encode the proteins needed to complete the lytic infection including the head, tail, and lysis proteins.

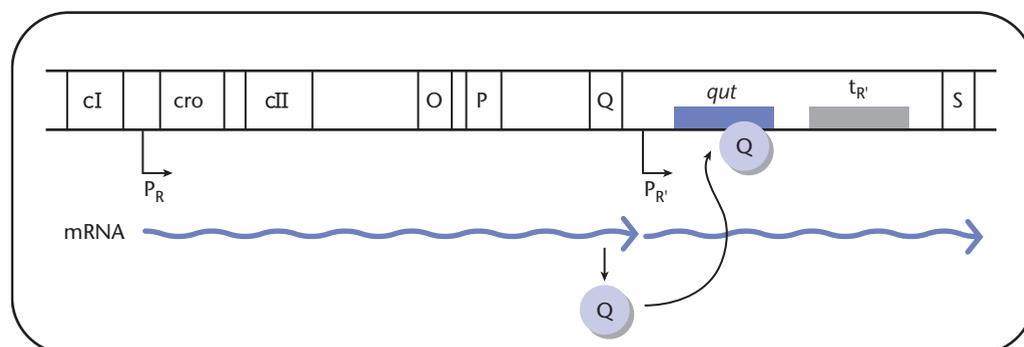


Fig. 7.8 The Q protein which is made from P_R when N is present is a second anti-termination protein. It acts on the *qut* site and allows transcription through $t_{R'}$. Q is necessary for synthesis of the head and tail genes.

DNA replication during the λ lytic pathway

After the infecting λ DNA has been converted to a double-stranded circular molecule, it replicates from a specific origin using both the phage-encoded O and P proteins and bacterial-encoded proteins. Replication proceeds bidirectionally, much like the *E. coli* chromosome. This form of replication produces molecules that look like the Greek letter theta and is called **theta replication** (Fig. 7.9a). Later in lytic development, λ switches to a second mode of replication called rolling circle replication.

Rolling circle replication of λ DNA commences when an endonuclease, encoded by λ *exo*, cuts one strand of the covalently closed circular double-stranded DNA molecule (Fig. 7.9b). The cut strand is called the plus strand. The 5' end of the cut plus strand is peeled away from the intact minus strand. DNA polymerase adds deoxyribonucleotides to the free 3' OH of the cut plus strand using the intact circular minus strand as the template. This produces new plus strands through a process of continually elongating the original plus strand. The new plus strands are used as a template to synthesize new minus strands. Rolling circle replication produces long DNA molecules containing multiple phage genomes called **concatamers**.

Making λ phage

The structure of the finished capsid is determined by the physical characteristics of the structural proteins that they are made from and the phage and host proteins used for assembly. Assembly of the capsids requires at least 10 phage-encoded proteins and two host-encoded proteins. The final capsid is made up of eight proteins, E, D, B, W, FII, B*, X1, and X2. Initially, B, C, and Nu3 (all phage proteins) form a small, ill-defined initiator structure (Fig. 7.10a). This structure is a substrate for the host-encoded GroEL and GroES proteins. GroEL and GroES act on proteins or protein complexes and help remodel them. The major coat protein, E, is added to this structure to form an immature phage head (Fig. 7.10b). The immature phage head is converted to the mature

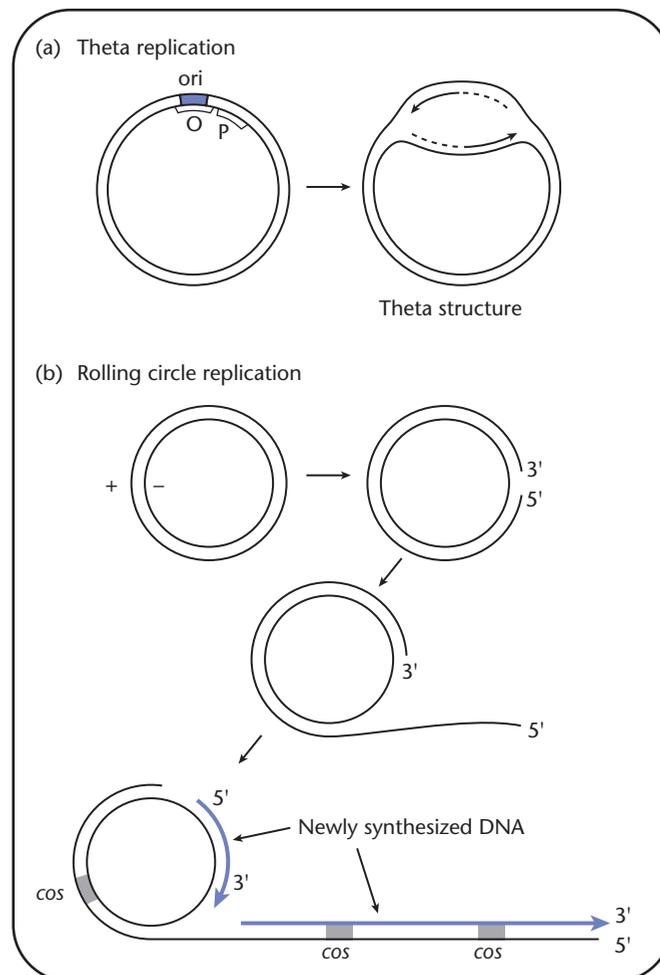


Fig. 7.9 λ has two modes of DNA replication: theta replication (a) and rolling circle replication (b). Theta replication occurs early in infection and rolling circle replication occurs late in infection. Rolling circle replication produces concatamers for packaging into phage heads.

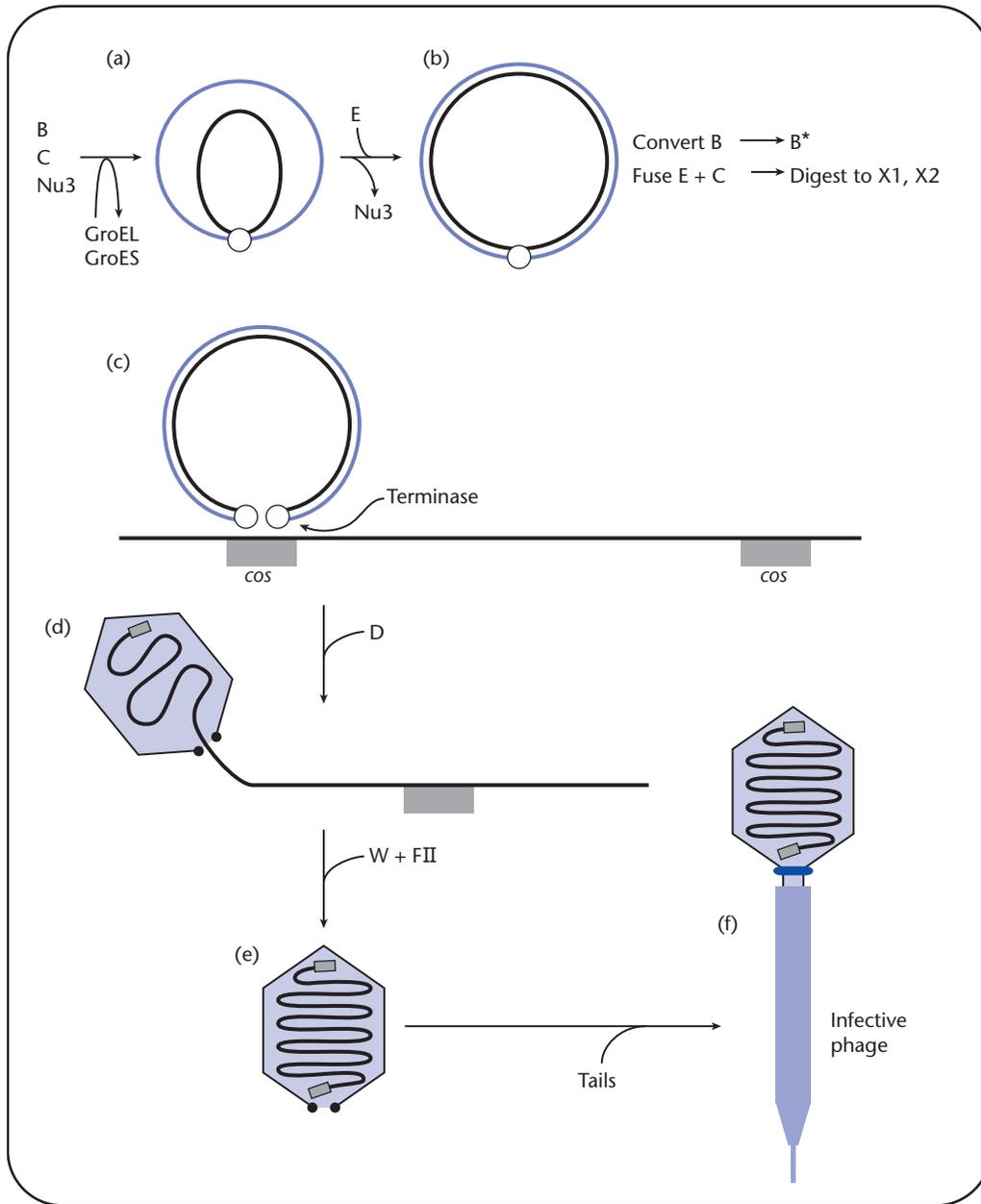


Fig. 7.10 The assembly pathway for λ . (a) The initiator structure for the head is composed of the B, C, and Nu3 proteins. (b) E, the major head protein, is added to this structure. Nu3 is degraded, B is cleaved to a smaller form (B*), and E and C are fused and cleaved at a new position to form X1 and X2. This forms the immature phage head. (c) The immature phage head is now ready for DNA from a concatamer. The D protein is added to the capsid at this point. (d) Packaging starts at a *cos* site and proceeds to the next *cos* site. (e) The DNA is inserted into the capsid and sealed inside by the W and FII proteins. (f) Tails are added to the full capsid to form a phage.

phage head by the degradation of Nu3, the cleavage of B to B*, and the fusion of C protein and some E protein followed by the cleavage of the fused protein into two new proteins, X1 and X2 (Fig. 7.10c).

The mature phage head is now ready for DNA. As the DNA is inserted into the phage head, it expands and the D protein is added to the surface of the capsid. λ DNA cannot be packaged from a monomer of λ DNA but only from concatamers usually produced by rolling circle replication (Fig. 7.10c). The DNA is cut at one *cos* site by a λ encoded enzyme and put into the phage head. Terminase binds to a *cos* site and to a phage head, cuts the *cos* site, and inserts that end of the λ DNA into the phage head (Fig. 7.10d). Terminase cuts the *cos* site asymmetrically, leaving the 12 base pair overhang. The terminase enzyme then tracks along the concatamer of λ DNA until it reaches a second *cos* site. As terminase tracks, the DNA is inserted into the phage head. When a second *cos* site is reached, terminase cuts the DNA and the last bit of DNA is inserted into the phage head (Fig. 7.10e).

This phage head with newly inserted DNA is unstable and not able to join to phage tails. The W and FII proteins are added to the base of the full head (Fig. 7.10e). This both stabilizes the DNA-containing head and builds the connector to which the tail will bind. Tails add spontaneously to this structure (Fig. 7.10f).

Tails are constructed from 12 gene products. Like the capsids, the tails are formed from an ill-defined initiator complex. This complex requires the J, I, L, K, H, G, and M phage proteins. They are added to the complex in the order listed beginning with the J protein. For this reason, it is thought that tails are built from the tip that recognizes the bacterium towards the end that binds to the phage head. Once the initiator structure is formed, the major tail protein, V, is added. The H protein is used as a measuring stick and determines how long the tail will be. Once the tails reaches the correct length, the U protein is added to prevent further growth and the H protein is cleaved. The Z protein is added last and is required to make an infectious phage. A tail without Z will bind to a full phage head but the resulting particle is not infectious.

The λ phage packaging system packages DNA molecules on the basis of the *cos* sites rather than on the basis of the length of the DNA molecule. Varying lengths of DNA molecules, within set limits, can be packaged as long as the molecule contains a *cos* site at both ends. If the distance between the two *cos* sites is less than ~37 kb, the resulting phage particle will be unstable. When the DNA is inside the capsid, it exerts pressure on the capsid. Likewise the capsid exerts an inward force on the DNA. If there is not enough DNA inside the capsid, it will implode from the inward force of the capsid. If the distance between the two *cos* sites is too far (~52 kb), then the capsid will be filled before the second *cos* is reached. The tail cannot be added because the DNA hanging out of the capsid is in the way and no infectious phage particle is produced.

Getting out of the cell—the λ S and R proteins

The λ R and S proteins are required for λ to release progeny phage into the environment. The R protein is an endolysin that degrades the peptidoglycan cell wall and allows the phage to escape from the cell. The S protein forms a hole in the inner membrane to allow the endolysin to gain access to the cell wall. After the hole is formed, approximately 100 intact λ phage particles are released into the environment. The entire lytic cycle lasts ~35 minutes.

Induction of λ by the SOS System

When a λ lysogen is treated with ultraviolet light (UV), ~35 minutes later the cells lyse and release phage. What does the UV do to the cell? UV damages the DNA and triggers a cellular response called the SOS response to deal with this damage (Fig. 7.11). The RecA protein, which is normally used for homologous recombination, is activated and binds to the LexA protein. The activated RecA complexed to LexA induces LexA to cleave itself. This leads to the activation of a number of genes whose products repair the DNA damage in the cell. λ has tapped into this system through the CI protein. CI repressor can interact with activated RecA, leading to the cleavage of CI. This leads to expression of the phage lytic genes and phage production. The rationale for this response is that λ does not want to risk staying in a cell that has DNA damage and may not survive.

Superinfection

If a cell is a λ lysogen, another λ phage that infects is not able to undergo lytic development and produce phage.

The incoming phage can inject its DNA, however, the DNA is immediately shut down and no transcription or translation of the λ initiates. λ lysogens are immune to infection by another λ phage particle, which is called **superinfection**. Superinfection is blocked because the lysogen is continuously producing CI repressor. The lysogen actually produces more repressor than it needs to shut down one phage. This extra repressor binds to the superinfecting phage DNA at O_L and O_R and prevents transcription from P_L and P_R .

Restriction and modification of DNA

A simple experiment with λ leads to the discovery of how bacteria tell their own DNA from foreign DNA. λ is capable of making plaques on two different types of *E. coli*, *E. coli* K12 and *E. coli* C. If λ is grown on *E. coli* K12, it will form plaques on *E. coli* K12 or *E. coli* C with equal efficiency. If λ is grown on *E. coli* C, it will form plaques on *E. coli* C but if it is plated on *E. coli* K12, only a few phage will form plaques. The efficiency of forming plaques or **efficiency of plating** (EOP) is decreased by 10,000-fold. This is known as **restriction**. If the *E. coli* C grown phage that did plaque on *E. coli* K12 are replated on *E. coli* K12, the EOP is 1. This is known as **modification**. The few phage that survive the replating on *E. coli* K12 have been modified so that they can efficiently plate on *E. coli* K12.

While this originated as a curiosity of phage growth, it has proven to be essential for many molecular techniques. Further investigation showed that the protein responsible for restriction, a **restriction enzyme** or **restriction endonuclease**, actually recognizes a specific DNA sequence and cleaves the DNA on both strands. The cut or digested DNA is sensitive to nucleases that degrade DNA. The modification part of the system is a protein that specifically modifies the DNA sequence recognized by the restriction enzyme and prevents the DNA from being digested. *E. coli* K12 has a restric-

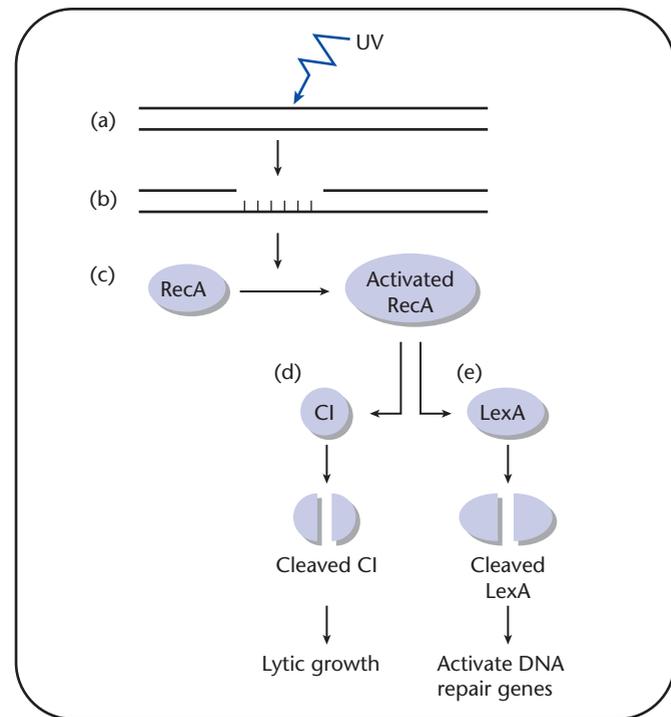


Fig. 7.11 The SOS response induces λ . UV treatment of cells (a) damages the DNA and leaves stretches of single-stranded DNA (b). The single-stranded DNA activates RecA (c). Activated RecA interacts with CI, leading to cleavage of CI and induction of the λ lysogen (d). Activated RecA also interacts with LexA and leads to LexA inactivation (e). LexA inactivation leads to expression of a number of genes, including some DNA repair enzymes.

FYI 7.2**Pathogenicity in *Vibrio cholera***

Cholera is caused by the bacterium, *Vibrio cholera*. Many of the genes that make this bacteria pathogenic or disease causing are part of a prophage located in the *V. cholera* chromosome. This prophage bears striking resemblance to M13 and other filamentous phage. It is possible that the transmission of these pathogenic genes is as simple as the phage moving from one bacterial species to another sensitive bacterial species.

tion/modification system and *E. coli* C does not. This explains the original observation with λ growth. If a bacterium carries the restriction enzyme, it must also carry the modification enzyme so that the bacterial chromosome is not digested and degraded. The restriction/modification system allows a bacterium to tell DNA from its own species from foreign DNA. Many different bacteria contain restriction/modification systems that recognize different DNA sequences. The restriction enzymes are purified and used in vitro to cleave DNA at specific DNA sequences, depending on the recognition sequence of the enzyme in question. Restriction enzymes are used to cleave and clone DNA fragments as described in Chapter 14.

The lifecycle of M13**M13 adsorption and injection**

M13 adsorbs to the tip of the F pilus, a hair-like structure on the surface of some bacteria. It can only infect bacteria that carry an F or F-like conjugative plasmid that encodes the proteins that make up the F pilus (see Chapter 10). For the filamentous phage, it is known that infection is initiated by the binding of gpIII to the tip of the F pilus. GpIII then interacts with the inner membrane protein TolA. Two additional facts about gpIII suggest a mechanism for phage DNA entry. GpIII contains amino acid sequences that are fusogenic or promote localized fusion of two membranes and gpIII is capable of forming pores in membranes that are large enough for DNA to go through. If each of these properties of gpIII are important for phage entry, then the phage could bind to the F pilus, promote fusion of the membranes, and use gpIII to form holes in the membrane to gain entry into the cytoplasm.

Protection of the M13 genome

The M13 DNA that ends up in the cytoplasm is a circular single-stranded DNA molecule. The strand present in phage particle is known as the plus or + strand. After entry into the cytoplasm, the + strand DNA is immediately coated with an *E. coli* single-stranded DNA binding protein known as SSB. The SSB coating protects the DNA from degradation.

M13 DNA replication

The M13 plus strand is converted to a double-stranded molecule immediately upon entry into *E. coli* (Fig. 7.12). Synthesis of the complementary strand is carried out entirely by *E. coli*'s DNA synthesis machinery. The complementary strand is called the minus or – strand. Only the minus strand is used as the template for mRNA synthesis and ultimately it is the template for the translation of the encoded M13 gene products. The SSB that coats the plus strand upon entry of the DNA into the *E. coli* cytoplasm fails to bind to ~60 nucleotides of the molecule (Fig 7.12c). These nucleotides form a hairpin loop that is protected from nuclease degradation. M13 gpIII from the phage is found associated with the hairpin loop. The hairpin loop is recognized by *E. coli* RNA polymerase as a DNA replication origin and is used to initiate transcription of a short RNA primer (Fig. 7.12d). The RNA primer is extended by *E. coli* DNA polymerase III to create the minus strand (Fig. 7.12e). The RNA primer is eventually removed by the exonuclease activities of *E. coli* DNA polymerase I. The gap is filled in by

the 5' to 3' polymerizing activity of the same DNA polymerase. *E. coli* ligase forms the final phosphodiester bond resulting in a covalently closed double-stranded circular M13 chromosome. The double-stranded form of M13 chromosome is called the replicative form (RF) DNA.

The RF form is replicated by rolling circle replication similar to the mechanism used by the λ chromosome (see Fig. 7.9b). The M13 gene II encoded protein is an endonuclease that nicks the plus strand of the RF DNA at a specific place to initiate the replication process for M13 RF DNA. Approximately 100 copies of M13 RF DNA are made. While the M13 chromosome is being replicated, the genes encoding the coat proteins are being transcribed and translated. When M13 gpV protein accumulates to sufficient levels, a switch from synthesizing RF DNA to synthesizing the plus strand occurs. GpV blocks the synthesis of the minus strand, presumably by displacing SSB on the plus strand and preventing the plus strand from being used as a template. The plus strand is circularized.

M13 phage production and release from the cell

M13 phage particles are assembled and released from *E. coli* cells through a process that does not involve lysing *E. coli* or disrupting cell division (Fig. 7.13). The gpV coated plus strand makes contact with the bacterial inner membrane (Fig. 7.13a). This interaction requires a specific packaging sequence on the DNA and gpVII and gpIX. The protein-coated DNA traverses the membrane and gpV is replaced by gpVIII in the process (Fig. 7.13b). GpVIII is found in the membrane. When the last of the phage particle crosses the membrane, gpIII and gpVI are added. M13 phage are continually released from actively growing infected *E. coli*.

The lifecycle of P1

Adsorption, injection, and protection of the genome

P1 adsorbs to the terminal glucose on the lipopolysaccharide present on the outer surface of the outer membrane. The P1 tail can contract, suggesting that P1 might inject its DNA into the cell like T4 (see below). Once inside the cell, P1 DNA circularizes by homologous recombination. Circularization can occur by recombination because when the phage DNA is packaged, 107% to 112% of the phage genome is incorporated into a capsid. This ensures that every phage DNA molecule has between 7 and 12% homology at its ends; a property called **terminal redundancy** (Fig. 7.14). The terminal redundancy is used to circularize the genome.

P1 DNA replication and phage assembly

Like λ , early P1 replication takes place by the theta mode of replication. Later in infection, P1 switches to rolling circle replication, again like λ (see Fig. 7.9). At

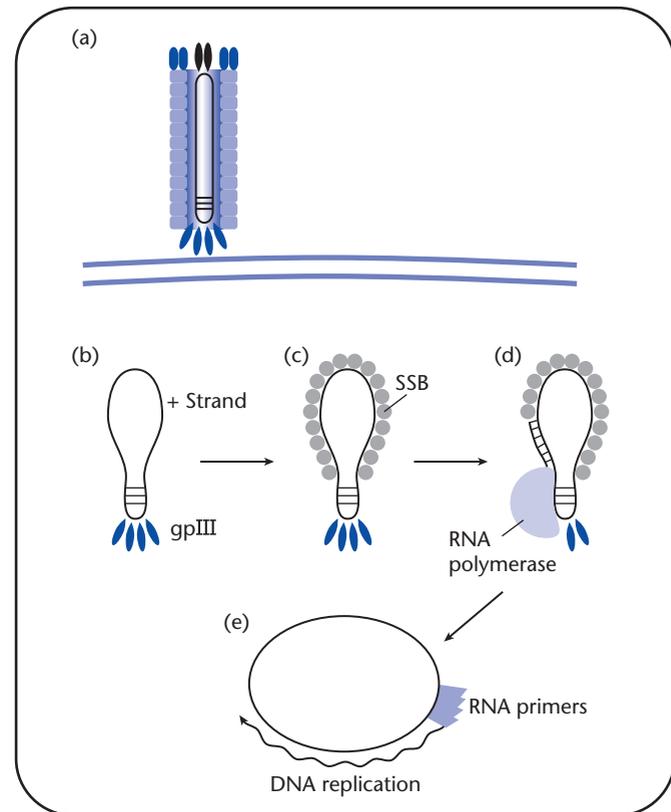
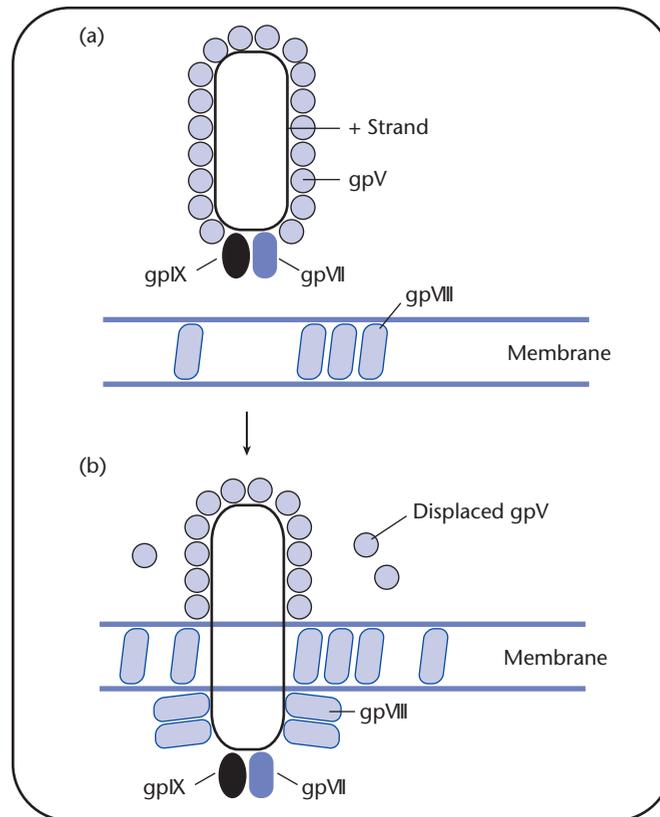


Fig. 7.12 The conversion of the M13 plus strand to a double-stranded DNA molecule. The plus strand enters the cell (a and b) with gpIII attached. It is immediately coated with host SSB (c). RNA polymerase synthesizes a short primer (d) and DNA polymerase synthesizes the minus strand.

Fig. 7.13 M13 is released from the cell without lysing the bacterium. (a) The plus strand, coated with gpV interacts with the membrane through gpVII and gpIX. (b) As the DNA traverses the membrane, the gpV is replaced by gpVIII.



approximately 45 minutes after infection, the cells are filled with concatamers of phage DNA, assembled phage heads, and assembled phage tails. Now assembly of the complete phage must take place. A protein made from the phage genome recognizes a site on the concatamers of phage DNA called the ***pac* site** (Fig. 7.15). The protein cuts the DNA, making a double-stranded end. This end is inserted into a phage head. The DNA continues to be pushed inside the head until the head is full, a process termed **headfull packaging**. Once the first phage head is filled, another empty phage starts packaging. Experiments indicate that up to five headfulls of DNA can be

packaged sequentially from a single *pac* site at 100% efficiency. An additional five headfulls of DNA can be packaged although the efficiency gradually decreases over these last five headfulls to only about 5%. While each phage head contains the same genes, the gene order changes. This is known as **circular permutation** of the genome (Fig. 7.14). After the head is full of DNA, a double-stranded cut is made and a tail is attached. This part of phage development is very much an assembly line. P1 is thought to encode an endolysin and holin to use in lysing the cell, similar to those described for λ .

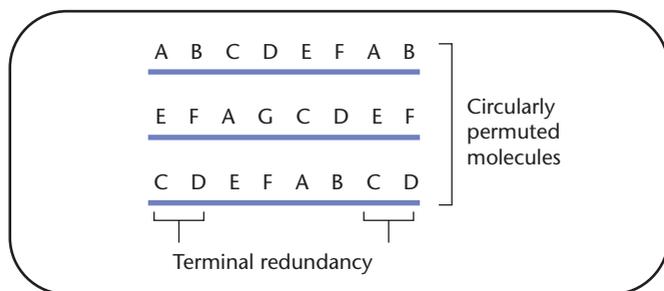


Fig. 7.14 P1 genomes are both circularly permuted and terminally redundant. Terminal redundancy means that the same sequences are present on both ends of one DNA molecule. Circular permutation means that the order of the genes on each DNA molecule is different but every DNA molecule contains the same genes.

The location of the P1 prophage in a lysogen

Prophages can be physically located in one of two places in a lysogen. In the case of λ , the phage genome is recombined into the bacterial chromosome. P1 is maintained in the cytoplasm as a stably inherited extrachromosomal piece of DNA or **plasmid** (see Chapter 9). P1 contains an origin for DNA replication and once the phage genome is converted to circular, double-stranded DNA, it can be established as a plasmid.

P1 transducing particles

One unusual aspect of P1 development is the formation of **transducing particles** or phage particles that contain chromosomal DNA instead of phage DNA. The *E. coli*

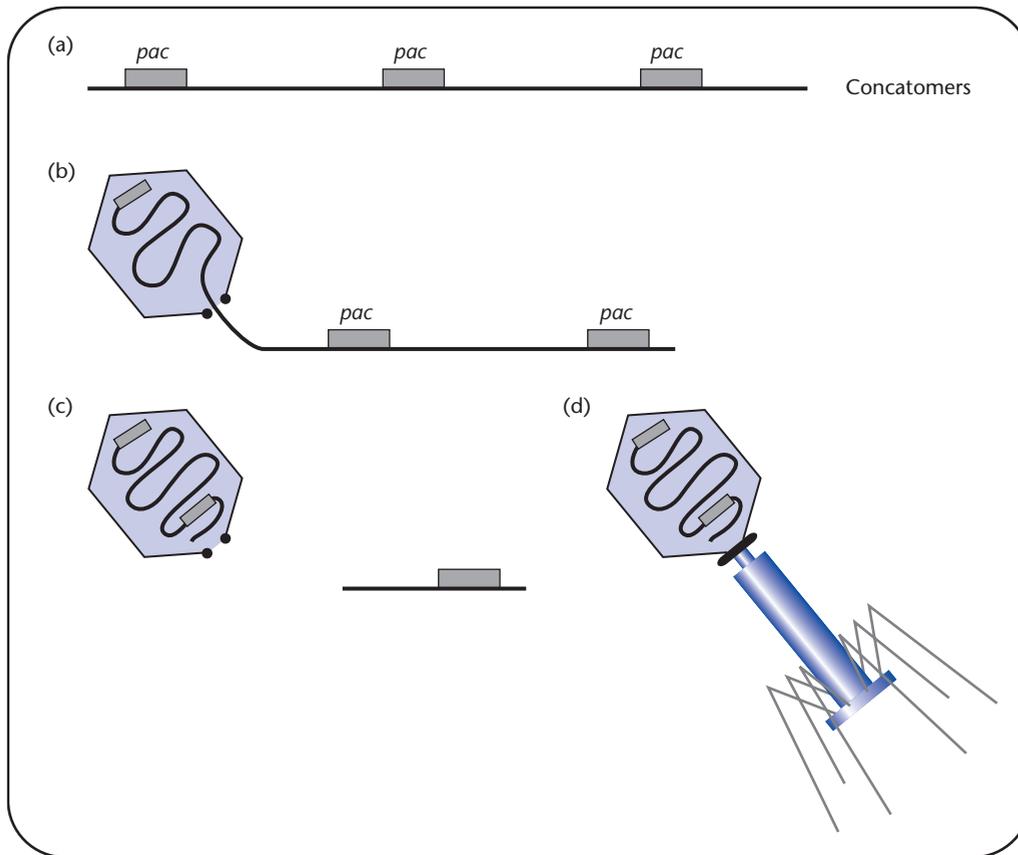


Fig. 7.15 P1 packages DNA from a *pac* site (a) and packages between 7 and 12% more than one P1 genome, until the phage head is full (b and c). Once the phage head is full, a preassembled head is added (d).

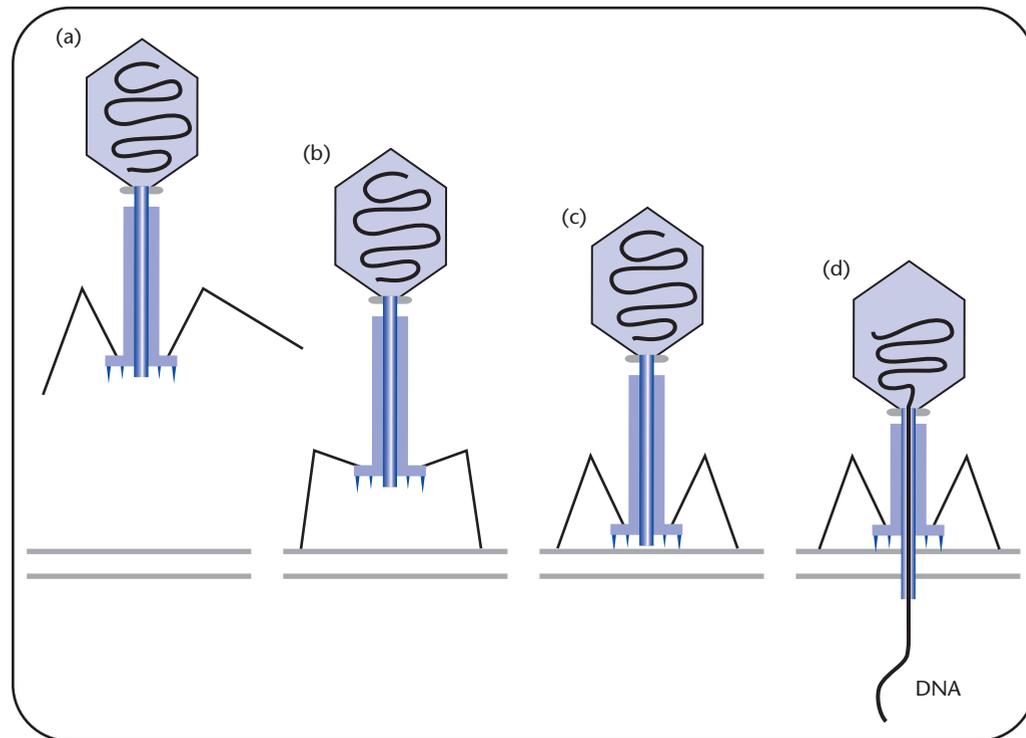
chromosome contains many **pseudopac sites** or sites that can be used to initiate packaging of host chromosomal DNA into maturing phage. These pseudopac sites are used much less frequently than the phage *pac* sites but they are used. The resulting phage carry random pieces of the chromosome in place of phage genomes. The ability to package any piece of chromosomal DNA instead of phage DNA makes P1 a **generalized transducing phage**. Transducing particles are used to move pieces of host chromosomal DNA from one strain to another for the purposes described in Chapter 8.

The lifecycle of T4

T4 adsorption and injection

For T4, the phage binds to the lipopolysaccharide. The tips of the tail fibers contact the cell first (Fig. 7.16). Once the phage has bound to the cell, the base plate rearranges creating a hole in the base plate. The outer sheath contracts and the internal tube goes through the outer membrane, peptidoglycan, and periplasm and comes close to the cytoplasmic membrane. The DNA is injected and crosses the cytoplasmic membrane in about 30 seconds. Not all phage that have the structure of T4 inject their DNA this way. Some phage such as T7, have tails that cannot contract. The T7 genome is only 40 kb but takes 9 to 12 minutes to cross into the cytoplasm. For T7, a small portion of

Fig. 7.16 Injection of T4 DNA into the cell. (a) T4 “looks” for a susceptible bacterium with its tail fibers. (b) The tail fibers recognize the membrane first. (c) The tail spikes interact with the membrane. (d) The tail sheath contracts, driving the internal tail tube through the outer membrane, peptidoglycan, and to the inner membrane where the DNA is released.



genome (about 8%) crosses both membranes, the peptidoglycan and periplasm, and enters the cytoplasm. After a 4-minute lag during which two proteins encoded by this piece of DNA are synthesized, the rest of the phage DNA enters the cytoplasm. Binding of these two phage proteins to the DNA is thought to pull the DNA into the cytoplasm.

Once T4 DNA is in the cell cytoplasm, it specifies a highly organized and coordinated program of gene expression. A group of genes with similar promoters, called the **early genes**, are transcribed and translated by host enzymes. One early gene encoded protein activates a second set of promoters for the **middle genes**. A different early gene encoded protein shuts off synthesis of the early genes. One product of middle transcription is required to activate the **late genes**. The early genes encode the proteins needed for DNA synthesis and late genes encode the proteins needed to build the capsid and tail structures. Many phage stage the expression of their genes in this temporal fashion to ensure proper construction of the phage particles.

The T4 genome does not contain cytosine residues. All of the cytosines are modified by hydroxymethylated. Several of the early genes encode proteins that degrade the cytosine-containing host DNA. The phage DNA is protected from degradation. The T4 genome, like P1 is both circularly permuted and terminally redundant. T4 has about 3% terminal redundancy. Unlike P1, T4 does not appear to use this terminal redundancy to circularize upon infection. T4 begins replication immediately after the early gene products are made. T4 replicates as a linear molecule and uses replication and recombination to both replicate the entire genome and to make concatamers to package into phage heads (Fig. 7.17). Like other phage, capsids, tails, and concatamers of phage DNA are premade and assembled into infectious phage particles late in phage development.

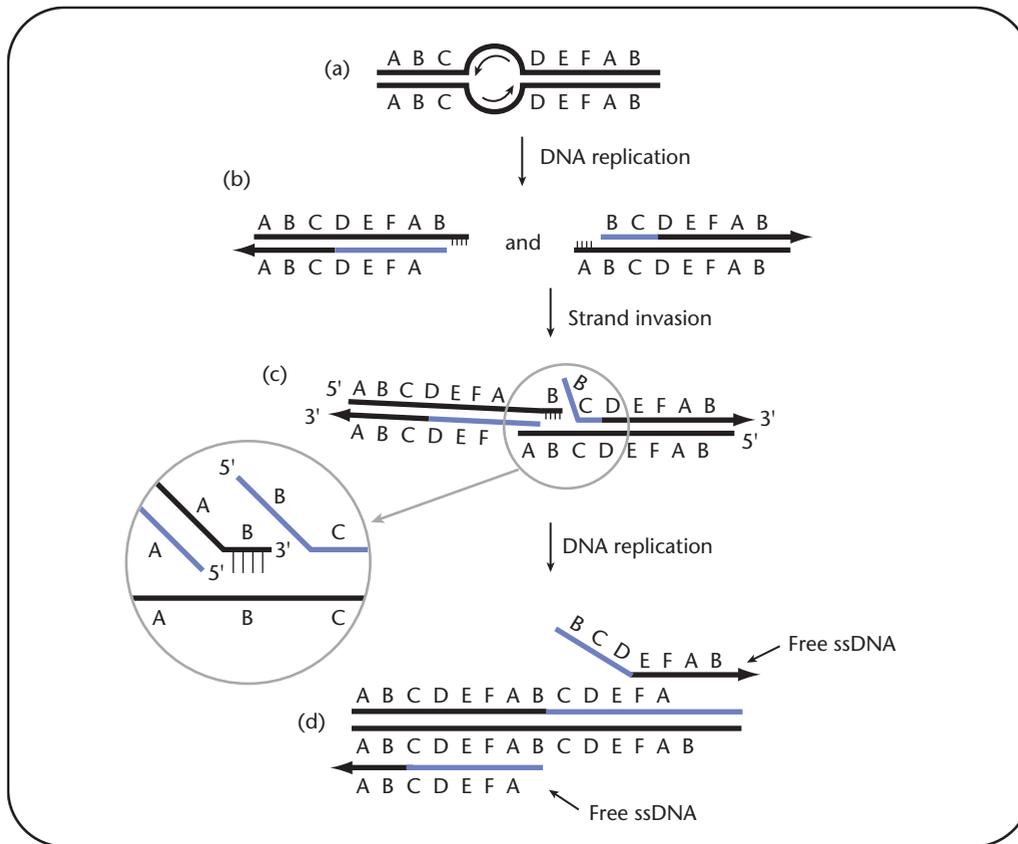


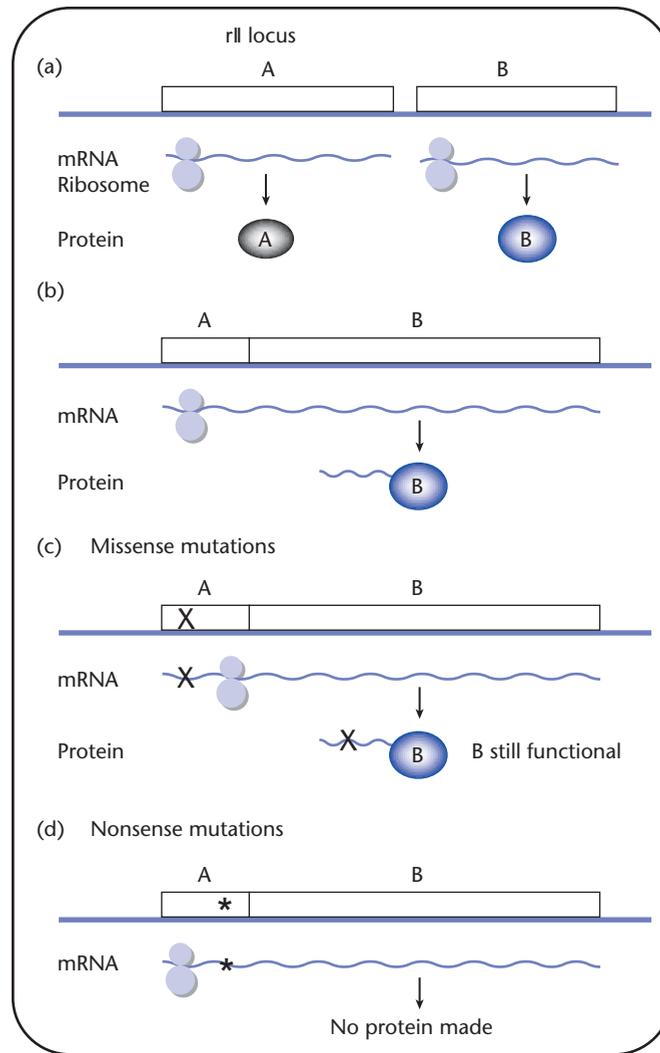
Fig. 7.17 T4 replicates its DNA using both replication and recombination. (a) Linear T4 DNA molecules are injected into the cytoplasm of the host. (b) DNA replication begins at an origin and proceeds bidirectionally to the ends. However, because of DNA polymerase's requirement for a primer, a piece of the DNA at one end of the molecule cannot be replicated and remains single stranded. (c) This piece of single-stranded DNA can invade duplexed DNA at any place where it has homology, like the initial reaction in recombination. (d) DNA replication of this molecule will lead to concatamers of the phage genome. Depending on where strand invasion takes place, branched molecules can also be formed. T4 packages its DNA out of the concatamers. The displaced single strands are free to strand invade the concatamer structures.

T4 rII mutations and the nature of the genetic code

The study of two genes in T4 has contributed significantly to our understanding of the genetic code and the nature of the gene. In the late 1950s and 1960s the understanding of the nature of the gene was in its infancy. The prevailing thought was that the gene was the smallest genetic unit and it was inherited as a unit. The chemical nature of DNA had just been described by Watson and Crick. The relationship between DNA and the gene was not understood. Seymour Benzer used the T4 rII locus and genetic logic to describe several key features of the gene.

rII encodes two proteins, A and B. Several thousand point mutations and deletions were isolated in these two genes. These mutations were put to good use. A phage carrying one mutation and a phage carrying a second mutation were mixed together and

Fig. 7.18 The T4 rII locus was used to conduct studies on the nature of the gene. (a) rII is composed of two genes A and B that are normally made into separate proteins. (b) A specific deletion was described from the mapping studies. This deletion (called r1589) fused the A and B genes, leaving A nonfunctional but B functional. (c) Some mutations in A, called missense mutations did not interfere with B function. (d) Other mutations in A did interfere with B function. These mutations were said to be “nonsense” and interfered with the production of B.



grown on an appropriate host to determine if any of the offspring had recombined back to wild type. Many of these phage crosses were carried out and used to construct a map of where the mutations resided in the rII locus.

Several conclusions were drawn from these studies. Deletion mutations were defined as mutations that could not generate wild-type recombinants when crossed with more than one of the other mutations. Deletion mutations were missing a part of the gene. In the thousands of crosses that were carried out, the frequency of obtaining wild-type recombinants was predictable based on the positions of the starting mutations. This led to the conclusion that DNA was a linear molecule across the length of the gene and not a branched molecule. If DNA was branched, then the recombination frequencies should be very different

when one mutation resided on one side of a branch and the other mutation resided on the opposite side of the branch. Using the recombination frequencies, it was determined that recombination can take place within a gene and not just outside of it as had been thought. This changed the definition of a gene from the unit of heredity that mutated to altered states and recombined with other genes. It is now recognized that the gene is a functional unit that must be intact in the DNA to lead to a specific characteristic or phenotype. Each of these studies shed more light on the behavior of the gene and the nature of mutations.

One of the deletions fused the A gene to the B gene such that the A gene was not functional but the B gene was (Fig. 7.18). Some of the point mutations in A could be crossed onto the same phage that carried the deletion. If a point mutation in A did not affect the activity of B in the fused genes then the point mutation was interpreted to be a missense mutation. Missense mutations could cause a change in the genetic code without affecting the production of the protein product. Other point mutations in A did affect the activity of B. These were interpreted to be nonsense mutations or changes that stopped the production of the B protein. These studies led directly to the modern concepts of a gene and how it functions.

FYI 7.3

The RNA phage MS2

MS2 is a typical phage containing an RNA genome. MS2 binds to the F pilus. It has a genome of 3569 nucleotides and encodes only four proteins: the coat protein, an RNA-directed RNA polymerase, a lysin and an adsorption protein. MS2 has an icosahedral capsid composed mainly of one type of protein with a few molecules of a minor protein in it.

Summary

Bacteriophage are a very diverse group of viruses. Their genomes can be made from either DNA or RNA. They can be linear or circular, single stranded or double stranded. Phage have evolved many different ways to carry out the limited number of steps in a phage infection. All phage must recognize the correct bacterium to infect, get their genome inside the cell, replicate the genome, transcribe and translate the genome, and assemble phage particles. The relative

simplicity of phage have made them favorite model systems to study many biological processes. While it may appear that phage carry out some processes using baroque mechanisms, it usually turns out that other biological systems share these mechanisms. For example, the unusual mechanism used to replicate T4 DNA is also used to help maintain bacterial and eukaryotic chromosomes.

Study questions

- 1 What processes must be carried out by all phage to produce progeny?
- 2 What is the phenotype of a λ mutant containing a defective *ci* gene?
- 3 Which regulatory proteins and promoters are crucial in λ 's decision-making process? Which regulatory proteins and promoters are crucial in λ 's lytic pathway? Which regulatory proteins and promoters are crucial in λ 's lysogenic pathway? Describe the roles for all identified participants.
- 4 A new phage from local sewage was recently isolated that infects laboratory strains of *E. coli*. How would you determine if this new phage is a temperate or lytic phage using simple genetic tests?
- 5 Contrast and compare the lytic pathway for λ and M13 phage. What do they do that is similar? What do they do that is different?
- 6 Contrast and compare rolling circle replication and theta mode replication. What components of the machinery are similar? What components of the machinery are different? When would one type of mechanism be preferable to the other type? Why?
- 7 How does T4 get its DNA from the phage head into the cytoplasm?
- 8 How do restriction/modification systems function?
- 9 How do different phage protect their DNA in the cell cytoplasm?
- 10 Why is M13 not considered a temperate phage?

Further reading

- Campbell, A.M. 1996. Bacteriophages. In *Escherichia coli* and *Salmonella typhimurium: Cellular and Molecular Biology*, 2nd edn., eds. F.C. Neidhardt, R. Curtiss III, J.L. Ingraham, E.C.C. Lin, K.B. Low, B. Hagasanik, W.S. Rexnikoff, M. Riley, M. Schaechter, and H.E. Umbarger, pp. 2325–38. Washington, DC: ASM Press.
- Ptashne, M. 1993. *A Genetic Switch*. 2nd edn. Cambridge, MA: Blackwell Scientific.
- Young, R., Wang, I.-N., and Roof, W. 2002. Phage will out: Strategies of host cell lysis. *Trends in Microbiology*, **8**: 120–8.
- Zaman, G., Smetsers, A., Kaan, A., Schoenmaters, J., and Konings, R. 1991. Regulation of expression of the genome of bacteriophage M13. Gene V protein regulated translation of the mRNAs encoded by genes I, II, V and X. *Biochimica et Biophysica Acta*, **1089**: 183–92.