Bacteriophages

ALLAN M. CAMPBELL

123

INTRODUCTION

Bacteriophages (phages) are viruses that infect bacterial hosts. Those that grow on *Escherichia coli* have a wide variety of sizes, shapes, and lifestyles. They range from phages with small single-stranded genomes (either DNA or RNA) and as few as four genes up to complex double-stranded DNA phages like T4, with over 200 genes. Some (like T4) are completely virulent, killing and lysing every cell they infect; others (like λ) are temperate, able to join the bacterial genome as quiescent prophages.

Phages have been studied for various reasons. In the decade following their discovery by Twort and d'Herelle 6, 49), research was driven by the hope (largely unfulfilled) that phages might prove useful in combating bacterial diseases. Their most important direct application to medical bacteriology has been as reagents for typing bacterial strains. Starting about 1940, phages became favored objects for basic experiments in mechanistic (eventually molecular) biology. From the 1970s onward, phages have been used as cloning vehicles for genetic engineering.

Whereas the role of phages in nature is incompletely understood, they seem to have substantial impact on the growth and evolution of bacteria in their natural habitats. It is likely that bacterial restriction enzymes developed primarily as a defense against phage infection, and selection for phage resistance has shaped other bacterial properties as well. Temperate phages affect the antigenicity and pathogenicity of hosts they lysogenize. Through transduction, they also mediate gene transfer within natural populations. Like all viruses, phages can be viewed as vehicles for packaging specific nucleic acids and delivering them to new cells, where they can direct the cell's biosynthetic machinery toward synthesis of more phage. The life cycle commences when a free phage particle (virion) encounters a susceptible cell and includes several (sometimes overlapping) phases: (i) entry of phage nucleic acid into the cell; (ii) phage gene expression (coordinated in a temporal program and frequently leading to inhibition of host macromolecular synthesis); (iii) replication of viral nucleic acid; (iv) assembly of new virions; and (v) egress of new virions from the cell (generally, but not always, achieved by cell death and lysis). A wide diversity of mechanisms are used in each of these phases.

In this chapter, the life cycles of representative phages from various groups are outlined. Some of these phages have been entirely sequenced, and in all of them, extensive genetic studies define the phage-specified determinants used. For all but the most complex phages (P1 and T4), a catalog of genes or major genes and their functions is included. These are always listed in their order of occurrence on the genome.

RNA PHAGES

Arguably the most primitive phages are those with RNA genomes. The RNA coliphages are male-specific, using the F-coded pili as attachment organs. Virions bind to the sides of pili, and their RNA is internalized by an unknown mechanism. A typical RNA coliphage (MS2) has a genome of 3,569 nucleotides (nt) of linear plus-strand (sense) RNA enclosed in an icosahedral capsid composed of a single protein species with one or a few molecules of a minor protein (Table 1) (50). The RNA encodes four proteins: coat (C), RNA-directed RNA polymerase (P), lysin (L), and adsorption protein (A) (50) (Table 2). The gene order is 5′*AC*(*L*)*P*3′, where *L* overlaps both the 3′ terminus of *C* and the 5′ terminus of *L*.

Inside the cell, phage RNA can be translated. Because genome and message are synonymous, there is no transcriptional control, yet an effective temporal program of gene expression is executed. This program depends on the secondary structure of the RNA, which has extensive self-complementarity. Thus the coat protein gene, whose product is needed in large amounts (180 copies per virion), is translated throughout the infectious cycle, whereas polymerase synthesis is positively regulated by coat gene translation (which frees the ribosome binding site from pairing with a complementary sequence within the coat gene) and negatively regulated by coat protein, which binds near the ribosome binding site and sequesters it. The former control prevents replication from initiating until conditions for translation are favorable, and the latter curtails wasteful synthesis of polymerase late in infection. The gene *A* ribosome binding site is buried in secondary structure of the folded RNA molecule and thus is accessible only on nascent RNA molecules whose 5′ termini are still open. The A protein remains associated with plus-strand RNA and is packaged with it. Translation of gene *L* is activated by termination of coat protein translation. A ribosome bound at the terminator for gene *C* can apparently loosen the secondary structure near the initiation codon for gene *L*, allowing either the same or another ribosome to initiate there (8, 31). This coordinates lysin synthesis with coat protein synthesis, so that lysis is unlikely before a full yield of intracellular virus has been assembled.

TABLE 1 Features of MS2

Virion: Icosahedral, tailless, T=3, single major capsid protein
Nucleic acid: Linear plus-strand RNA, 3,569 nt
Receptor: F pilus (lateral attachment)
Regulation: Translational; sequestration of ribosome binding sites
Effects on host synthesis: Competition for building blocks.
Replication: RNA → RNA; phage enzyme plus host translation factors
Initiation at 3′ end of template
Multiple replication forks on a single template
Packaging substrate: Linear monomer
Lysis: Phage lysin, mechanism unknown
Growth parameters: Burst size 20,000
Latent period 22 min
Related or similar phages: Qβ, f2, R17, GA, SP

Replication is catalyzed by a complex composed of the phage-coded polymerase and three host translation factors (ribosomal protein S1 and elongation factors Tu and Ts). The polymerase (most thoroughly characterized for phage Qβ) recognizes the terminal sequences of phage RNA and initiates replication from the 3′ end (24). Reinitiation can occur before the first copy has reached the 5′ end, so that multiple copies of minus-strand RNA are made concurrently from a single plus-strand. Full-length minus strands are then copied by the same polymerase into plus strands that can be packaged into virions.

As more RNA copies are made, coat protein accumulates, and plus-strand RNA with its associated A protein is packaged into virions, which typically number about 20,000. These are released by lysis triggered by the 8-kDa L protein.

TABLE 2
Genes of MS2

Gene	Product
A	Minor virion component, needed for attachment, perhaps for assembly
C	Major coat protein
L	Lysin
P	RNA-directed RNA polymerase

SINGLE-STRANDED DNA PHAGES

Isometric Phages

The prototypic isometric single-stranded DNA phage is ϕX174 (Table 3) (26). ϕX174 was the first DNA virus discovered to have a single-stranded genome, which proved to be circular. Its replication cycle was studied intensively, especially as it became clear that the first round of replication, in which the single-stranded circle is converted to a double-stranded circle, is carried out entirely by host enzymes. This provided the earliest system in which *E. coli* enzymes synthesized a biologically active DNA in vitro from a defined origin (21).

The ϕX174 DNA (5,387 nt) encodes 10 proteins, 4 of which are present in the virion. Sixty monomers of gpF form an icosahedral capsid, onto which are added two spike proteins (60 monomers of gpG and 12 monomers of gpH). The spike proteins can be physically removed without destroying the integrity of the capsid, but they are essential for infectivity. In normal infection, a molecule of gpH (the pilot protein) accompanies the DNA into the infected cell. However, successful infection can also be achieved through transfection of protoplasts by phage DNA. The fourth protein in the virion is a 4-kb internal protein (Table 4) (26).

After entering the cell, the ϕX174 DNA is used as a template for minus-strand DNA synthesis, producing double-helical DNA. Synthesis is primed by a short RNA transcript made by host primosomes, and host polymerase is used for elongation. Once made, the double-stranded DNA can be transcribed, and phage-specific proteins are made. One of these (gpA) initiates plus-strand synthesis by nicking the plus strand at a specific site within gene *A*. The 3′ end of the nicked strand then primes plus-strand synthesis of the rolling-circle type, where the 5′ end of the parental plus strand (covalently linked to gpA through a tyrosine ester linkage) is displaced by the newly synthesized DNA. When the replication fork has traversed the circle and reencounters the origin, the displaced 5′ end is ligated to the 3′ end created by a fresh nick, providing one free plus-strand circle and one double-stranded circle, both of which can initiate further rounds of replication. Transcription (from several phage promoters) proceeds in the same direction as replication.

TABLE 3 Features of ϕX174

Virion: Icosahedral, tailless, T=1; one capsid protein, two spike proteins
Nucleic acid: Circular plus-strand DNA, 5,387 nt
Receptor: Lipopolysaccharide
Regulation: None known
Effects on host synthesis: Inhibition of DNA synthesis by gpA*
Replication: DNA → DNA. First cycle (ss → ds)[a] carried out entirely by host enzymes
Rolling-circle replication of ds DNA initiated by phage protein gpA
Packaging substrate: Single-strand tails of rolling circles, processed into monomer circles
Lysis: Phage lysin, mechanism unknown
Growth parameters: Burst size 180
Latent period 13 min
Related or similar phages: G4, S13, α3

[a]ss, single stranded; ds, double stranded.

Late in infection, the single-stranded circles formed by replication are encapsidated into new virions rather than returning to the replicating pool. Single-stranded DNA at the initiation site becomes coated either with host single-stranded binding protein (Ssb), which eventually covers the entire single strand that returns to the replicating pool, or, competitively, with the phage protein gpC, which signals packaging into virions. Once gpC has interacted with the initiation complex on a double-stranded template, all further

replication from that template requires interaction with procapsids (2).
The cycle terminates by cellular lysis, mediated by viral protein gpE. A cloned copy of gpE provokes lysis of uninfected cells when expressed. The mechanism of lysis is unknown.

TABLE 4 Genes of ϕX174

Gene	Product
A	Initiator of rolling-circle replication, nicks and seals DNA at a specific site
*A** (within A)	Major coat protein
B (within A)	Assembly protein
K (overlaps A and C)	Nonessential protein
C	Single-strand DNA-binding protein directing plus strands toward packaging
D	Scaffolding protein
E (within D)	Lysin
J	Internal protein
F	Capsid protein
G	Major spike protein
H	Minor spike protein functioning in attachment; enters cell with DNA

Filamentous Phages

The filamentous DNA phages differ markedly from ϕX174 in their mode of attachment to and egress from cells. They are not detectably related to the isometric phages in DNA sequences, yet the replication cycles of the two groups are strikingly similar. M13 is a typical filamentous phage (Tables 5, 6) (33, 35).

The virion of M13 consists of a flexible helical capsid enclosing a single strand of circular DNA. The capsid is composed of a single protein species (gpVIII), except for a few molecules of minor proteins. At the leading end (the end that makes contact with the cell surface in infection) are the adsorption protein gpIII and a second protein, gpVI; at the opposite end are gpIII and gpIX. Adsorption is to the tips of pili determined by conjugative plasmids. Phages f1, M13, and fd all attach to F pili. The mechanism of internalization is unknown; apparently the whole virion, not just the DNA, is internalized.

TABLE 5 Features of M13

Virion: Filamentous, one major protein, four minor proteins at ends
Nucleic acid: Circular plus strand, 6,407 nt
Receptor: Tip of F pilus
Regulation: None known
Effects on host synthesis: No major effect; infected cells grow and divide
Replication: First cycle (ss → ds) carried out by host enzymes
Rolling-circle replication of ds DNA initiated by gpII
Packaging substrate: ss tails of rolling circles, processed into circular monomers, complexed with gpII
Egress: Through intact cell envelope
Growth parameters: Continuous release
Latent period 30 min
Related or similar phages: fd, fl, ZJ/2, Ec9, AE2, HR, δA, Jf1, Jf2, X, Ike

Once within the cell, the DNA replicates in a manner very similar to that of ϕX174. The role of ϕXgpC in directing plus-strand DNA toward virions is played by M13 proteins gpV and gpX. The gpV protein coats the single-stranded DNA to form a precapsid, whereas gpX prevents recycling of single strands into the replicating pool in a manner competitive with some activity of gpII.

The DNA enclosed in its gpV precapsid makes contact with the inner membrane, where it first interacts in an assembly step requiring gpVII and gpIX and a specific packaging sequence on the DNA, then passes through the cell envelope. In traversing the inner membrane, gpV is replaced by gpVIII (an integral membrane protein). When the end of the virion crosses the membrane, gpIII and gpVI are added, so that the leading end (in infection) exits last. Such extrusion, without lysis or inhibition of cell division, can continue indefinitely, leading to very high phage yields.

TABLE 6 Genes of M13

Gene	Product
IV	Assembly protein
I	Assembly protein
VI	Minor protein at leading end of virion
III	Minor protein at leading end of virion
VIII	Major coat protein
IX	Minor protein at virion terminus
VII	Minor protein at virion terminus
V	ss DNA-binding protein
X	Inhibitor of gpII?
II	Initiator of rolling-circle replication; nicks and seals DNA at a specific site

Filamentous DNA phages (especially M13) have become popular cloning vectors. Foreign DNA can be spliced either into double-helical replicating form M13 DNA (at a site where it does not disrupt genome function) or into plasmids that include signals for M13 replication and packaging and which then can be introduced by transformation into strains that supply M13 function from defective helper phage that is not itself packaged. The system has the advantages of high yield, unlimited packaging capacity, and single-stranded product suitable for use as a hybridization probe. Gene splicing can also create chimeras between gpVIII or gpIII and peptide sequences of interest, which are then displayed on the virion surface (4). This is useful in screening libraries for rare members with specific binding properties, because those virions that are enriched on affinity columns contain the genetic determinants for specific binding.

Because cells are not killed by infection, cloned genes that alter bacterial phenotype can be scored by observations of colonies arising from infected cells. For genes of the cloning host, this can be facilitated by using a phage with conditional mutations in gene II, where the cloned genes are perpetuated by integration into their chromosomal homologs (3).

VIRULENT DOUBLE-STRANDED DNA PHAGES

Early students of mechanistic biology agreed to focus their attention on seven natural isolates (numbered T1 through T7), all virulent phages that plate on *E. coli* B. These T phages fall into four groups: the T-even group (T2, T4, T6); T3 and T7; T1; and T5. Members of a group are similar in size, virion morphology, DNA sequence, and life cycle and can coinfect the same cell with production of viable recombinants. All the T phages have linear double-stranded DNA genomes.

The T-Even Group

The T-even group (specifically T4) have been studied most intensively. These are large phages, with DNA about 170 kb in length encoding over 200 gene products, at least 69 of which are essential for plaque

formation (Table 7) (34).

In the virion, T4 DNA is linear and circularly permuted (almost randomly) with a terminal redundancy of about 5 kb. T4 DNA differs from host DNA in that all of its cytosines are hydroxymethylated and many are glucosylated as well.

The DNA is surrounded by an elongated icosahedral protein capsid. A total of 840 copies of the major capsid protein (gp23) form a T=13 lattice with an extra row of capsomeres winding around the lateral faces. Two nonessential proteins (Hoc and Soc) are regularly distributed in the gp23 lattice. A minor protein (gp24) lies at the icosahedral vertices, and several other proteins are located at the portal vertex, which is connected to the phage tail. The tail (which is joined to the head by a collar structure) consists of a central hollow core surrounded by a contractile sheath, terminating at a hexagonal baseplate to which six long kinked fibers are attached. The tail fibers tend to stick to the sheath (in which state the virion is unable to attach to cells) and can be released by treatment with tryptophan.

In infection, the tips of the tail fibers first make specific contacts with lipopolysaccharide receptors on the cell surface, then become anchored to the cell. A remarkable series of mechanical changes ensues. The baseplate rearranges to create a central hole through which the core can pass, and contraction of the sheath drives the core through the cell envelope. DNA is then extruded from the phage head into the cell.

TABLE 7 Features of T4

Virion: Elongated icosahedral capsid, T=13, one major and several minor capsid proteins; tail with core, contractile sheath, hexagonal base plate with six kinked tail fibers
Nucleic acid: Linear ds, ~170 kb, circularly permuted, ~5-kb terminal repeat
Receptor: Lipopolysaccharide
Regulation: Early and middle transcription by host RNA polymerase Late transcription uses phage-coded factor and replicated DNA Translational repression of early genes at late times
Effects on host synthesis: Massive degradation of host DNA Additional inhibition of host transcription
Replication: Phage enzymes; first cycle bidirectional, primed by host RNA polymerase Later cycles primed by invading 3′ ends
Packaging substrate: Linear concatemer
Lysis: Holin and lysozyme
Growth parameters: Burst size 200 Latent period 23 min
Related or similar phages: T2, T6

The temporal program of gene expression is then set in motion. Early transcription is initiated at promoters controlled by the host RNA polymerase holoenzyme associated with the major host sigma factor σ^{70}. The genes expressed from these promoters include all those needed for phage DNA synthesis and also two regulatory genes: *motA*, whose product stimulates transcription from another set of promoters (middle transcripts), and *regA*, whose product is a translational repressor of early genes. Middle promoters, which have a specific nucleotide sequence (TGCTT) around –35, depend on RNA polymerase holoenzyme plus σ^{70}. One product of middle transcription is a phage-coded sigma factor (gp55), which replaces σ^{70} and initiates transcription at late promoters (with a consensus TATAAATA around –10), which control synthesis of virion components and lytic enzymes. Transcription from late promoters requires not only gp55 but also DNA replication. In vitro, the requirement for replication can be replaced by a nick or gap near the promoter in the nontranscribed strand and requires the participation of three DNA replication proteins, gp45, gp42, and gp66, which are thought to recognize the interruption and signal a polymerase-associated transcription factor, gp33 (20).

Soon after infection, some T4 early gene products cause a massive degradation of cytosine-containing host DNA, conversion of cytosine monophosphate to hydroxymethylcytosine monophosphate, and conversion of monophosphates to triphosphates. Bidirectional replication from several origins, primed by transcripts made by host RNA polymerase and executed by a complex of several T4-encoded enzymes,

proceeds to the molecular ends of the infecting DNA molecule. Because there is no source of a primer for Okazaki fragment synthesis at the 3′ ends of the template strands, the daughter molecules have projecting 3′ ends (37).

As infection proceeds, host RNA polymerase becomes associated with gp55, and several ancillary factors (not essential for plaque formation by T4) interfere with σ^{70}-requiring host transcription. At this point, the early replication origins become inactive, and further replication proceeds from a different source. The 3′ ends of linear molecules invade other molecules by forming D-loops with homologous sequences. In single infection, these will initially lie in the repeated DNA near the molecular ends. In subsequent cycles or in multiple infection, invasions can also be internal. The invading 3′ ends then prime DNA synthesis, producing (after multiple cycles) a complex DNA network that includes multigenomic linear concatemers. Either before or concomitant with packaging, some of the branches are cleaved off by a phage enzyme (gp49), yielding unbranched concatemers suitable as packaging substrates (19).

After early times, most replication is initiated by strand invasion, which mediates genetic recombination as well. The recombination rate in T4 is correspondingly high. Strand-invasion-associated replication is not the only source of genetic recombinants, because some treatments that block replication stimulate recombination still further (42).

Assembly of T4 virions has been studied in depth. Thanks to the availability of conditional mutations in the structural genes for most virion components and to the facility with which later steps take place in vitro, it was possible to deduce a great deal about the assembly pathways simply through mixing appropriate mutant extracts and observing the production of active phage. This provided a framework within which physical studies of interacting components could be placed. Heads and tails join to form head-tail complexes, to which tail fibers then attach.

In head formation, the capsid is formed first and then DNA is pumped into it from multigenomic concatemers. The first step of capsid assembly seems to be nucleation from a unique vertex, which becomes the portal vertex through which DNA enters the head and to which the tail later attaches. Nucleation requires some minor capsid proteins, some noncapsid proteins, and a host chaperonin (GroEL). The complex attracts the major capsid protein (gp23), which initiates a hexagonal surface lattice. gp23 interacts with a scaffolding protein (gp22), which fills the prohead. Another protein (gp24) locates at the icosahedral vertices. After the shell is constructed, gp22 is degraded, and DNA of the concatemeric substrate is drawn into the head. Host DNA has been degraded by this time and hence does not compete for packaging. Packaging requires two phage proteins, gp16 and gp17, where gp17 is the endonuclease responsible for cutting a headful of DNA from the concatemer when the head is filled (10). Cutting probably takes place at the portal vertex. Gp16 and gp17 are not present in the finished virion. After the DNA is cut, the collar proteins are added at the portal vertex, preparing the head to interact with the tail.

Both the cyclic permutation and the terminal repetition are consequences of headful packaging, which initiates at random positions on the concatemer and cuts out a headful length slightly longer than the genome length. This idea has been tested by independently varying the genome length (through deletions or additions) and the headful length (through alterations in capsid size.) Virions with abnormal capsid size can have isometric capsids (smaller than normal T4 and unable to package a complete genome) and elongated capsids (with additional equatorial rows of capsomeres that extend the lateral faces until, in very long capsids, they collectively approach a cylindrical configuration with the diameter of normal capsids.) A minute fraction of the virions in a wild-type lysate have abnormal capsid size, but the frequency is greatly enhanced by certain missense mutations in genes *22–24* or by certain treatments such as growth in the presence of the amino acid analog canavanine (11). All these alterations have the predicted results: if genome length is decreased while capsid size remains constant, the extent of terminal repetition increases, so that a tester gene lies within the terminal repeat in a larger than normal fraction of the phage progeny (scored by heterozygosity in mixed infection with genetically marked stocks [44]. If capsid size is decreased while genome length is held constant, the isometric virions package random segments of about 23 genome length; if it is increased, the resulting virions have more DNA than normal, a high frequency of

heterozygosity, and enhanced resistance to UV irradiation (17, 36).

Tail assembly commences with the baseplate, a complex structure made of more than 20 different protein species. Six wedges (composed of several proteins) surround the central hub. Wedge proteins assemble in a defined order, the innermost protein being added last (which should ensure that wedges cannot associate with the hub until they are complete). Core assembly is then initiated on the baseplate. After core growth is completed, the sheath is added around it. Finished tails can then join to heads.

The tail fibers are linear, with an acute-angle kink. The specificity for host cell receptors lies at the end of the fiber, at the C-terminus of gp37. The distal and proximal segments of the fiber are assembled separately, then joined at the kink. Completed fibers then attach to the baseplates of head-tail complexes. Determinants near the kink facilitate the addition by interaction with a collar component (Wac). This interaction is not essential in vivo, because *wac* mutants are viable. Addition of fibers to the baseplate is catalyzed by a phage protein (gp63), which also has RNA ligase activity. This was one of the first documented cases of catalysis of a noncovalent reaction between protein structures.

Lysis of the infected cell is brought about by two enzymes, a lysozyme (gpe) that hydrolyzes bonds between *N*-acetyglucosamine residues in the rigid murein layer of the cell envelope, and a holin (gpt) that creates holes in the inner membrane, allowing the lysozyme to contact its murein substrate and incidentally collapsing the membrane potential and killing the cell. Once the murein layer is destroyed, osmotic forces break open the cell and release progeny virions.

T4 was the first phage to be scrutinized genetically and was therefore used in many genetic experiments aimed at elucidating the structure of the T4 genome, the mechanism of T4 replication, the fine structure of the gene, and the mechanism of homologous recombination. T4 genetics was launched by the discovery that genetic recombinants are produced following mixed infection by parental phages bearing different mutations (29). Once a sufficient variety of mutants had been isolated (many of them conditional lethals, such as amber or *ts* mutations), a linkage map was constructed that fits with (and provides some of the evidence for) the replication cycle already described. Because progeny are packaged from random starting sites on a concatemer arising by strand invasions at random positions, the linkage map is circular. As the DNA is terminally repetitious, progeny of mixed infection may carry alleles from one parent at the left end and from the other parent at the right. Such heterozygous particles are recognizable because they produce a mixed yield, leading (for plaque morphology markers) to mottled plaques. The terminal repeat is about 3% of the genome, so approximately 1.5% of the progeny are heterozygous for a specific gene under examination. Individual bursts from virions heterozygous for multiple mutations have been analyzed (52). An array of recombinants has been found (indicating that recombination near the molecular ends is very frequent), and their pattern is polarized as though the markers closest to the molecular ends are preferentially lost (indicating that the termini of the infecting molecule, or copies thereof, initiate multiple-strand invasions).

The segment of the T4 genome that has been analyzed most intensively is the *rII* locus. The locus consists of two adjacent genes (operationally defined as cistrons, or units of complementation) whose exact function in T4 biology is still incompletely understood. On the normal laboratory host, *E. coli* B, *rII* mutants make larger plaques than wild type, because they are insensitive to lysis inhibition. When a cell infected with an r^+ phage is superinfected before lysis with a second T4 particle, lysis is delayed, and intracellular phage development continues. Because superinfection within plaques is common, wild-type plaques are small. When T4 *rII* mutants infect strains lysogenic for λ, the products of two genes expressed from the λ prophage (*rexA* and *rexB*) interact with the superinfecting phage to trigger a premature collapse of the membrane potential before any phage is liberated (39). This does not happen with T4 r^+. Therefore, T4 *rII* mutants make large plaques on *E. coli* B and no plaques on *E. coli* K-12(λ), whereas T4 r^+ plates on both.

The ease with which spontaneous *r* plaques can be recognized and picked, plus their failure to plate on K-12(λ), enabled Benzer to map this locus almost to saturation (6). A detailed recombinational map was constructed by pairwise infection of *E. coli* B with independent *rII* mutants and counting wild-type

recombinant progeny on *E. coli* K-12(λ). The task was simplified by the identification of some mutations as deletions rather than point mutations, on the primary criterion that deletions were unable to revert to wild type. Crosses among the deletions (scoring not for number of wild-type recombinants but only for their presence or absence) yielded results consistent with a linear genetic map and also divided the whole locus into intervals. Point mutations could be assigned to the proper interval by crossing with a set of deletions. Point mutations were found preferentially at certain hot spots, of which the most active proved to be repetitious sequences that allow slippage during replication.

Mixed infection on K-12(λ) tests for physiological complementation. The cistron was defined as a set of mutations that fail to complement one another, as judged by the ability of mixedly infected cells to produce a yield approaching that found with wild-type T4. The yield was measured by plating on B and consisted primarily of *r* mutants, indicating that the basis of the high yield was that coinfecting phages were sharing products rather than recombining to produce many r^+ progeny. On that criterion, the *rII* locus comprises two cistrons, A and B.

Besides their role in the foundations of molecular genetics, the *rII* mutants provided material for many sophisticated investigations of T4 genetics. The markers used in the studies with terminal repeat heterozygotes described earlier were *rII* mutations. More recent experiments indicated that splice-type heteroduplexes can arise from a single event (most plausibly, a strand invasion from a molecular end [41]). The experimental distinction between heterozygosis due to terminal repetition and that attributable to heteroduplex recombination intermediates, the extent to which more than two parental genomes contribute to a single recombinant, the high correlation between recombinational events in adjacent map intervals, and the number of coinfecting particles able to contribute genetically to the yield of a single infected cell were all worked out with the aid of *rII* mutants.

The T7 Group

Phages T3 and T7 are related in DNA sequence, and several other coliphages belonging to the same group have been discovered. T7 DNA is linear and double stranded, 39,936 bp in length. Its sequence is nonpermuted, with a terminal repeat of 160 bp (Table 8) (25).

All genes of T7 are transcribed rightward on the standard map, the leftmost genes being expressed at the earliest time (Table 9) (45). The class I genes (including that for T7 RNA polymerase), located at the left end and transcribed by host RNA polymerase, are expressed first. Several promoters are used, with all transcripts terminating at a common site. The class II genes, toward the middle of the map, are transcribed by T7 RNA polymerase and include the T7 DNA polymerase. The right half of the map, including genes for virion components and lysis, is also transcribed by T7 RNA polymerase, primarily at late times (although replication is not prerequisite to expression of these class III genes). Again, several promoters are used, the transcripts terminating near the right end of the map.

TABLE 8 Features of T7

Virion:	Icosahedral head
Nucleic acid:	Linear ds
Receptor:	Lipopolysaccharide
Regulation:	Early genes transcribed by host RNA polymerase
	Middle and late genes transcribed by phage RNA polymerase
Effects on host synthesis:	DNA degradation
	Inactivation of host RNA polymerase
Replication:	Bidirectional from internal origin
	Concatemers formed by joining through 160-bp terminal homology
Packaging substrate:	Linear concatemers
Lysis:	Phage lysin
Growth parameters:	Burst size 300
	Latent period 13 min
Related phage:	T3

T7 RNA polymerase is a highly efficient competitor for deoxyribonucleoside triphosphates, so much so that a cloned T7 RNA polymerase gene can convert a cell into a factory for making only those mRNAs programmed by the T7 promoter sequence. That sequence does not occur in the *E. coli* genome, so a gene controlled under a T7 promoter becomes virtually the only gene expressed in a cell that has been engineered to contain the T7 polymerase (46).

DNA replication, catalyzed by the T7 DNA polymerase, proceeds bidirectionally from a unique internal site, producing linear double-stranded daughter molecules with the 3′ ends of the template strands protruding. Because of the 160-bp repetition, these ends are complementary and can anneal. Trimming or gap filling plus ligation then creates a linear dimer, which can undergo additional cycles of replication. The eventual product is a linear concatemer in which 160-bp segments alternate with 40 kb of unique sequence DNA. The 160-bp segments and the 40-kb segments are in equimolar amounts, whereas in the packaged DNA they are in a 2:1 ratio. Before or during packaging, the 160-bp segment must be replicated to allow efficient packaging. This may be achieved by extra replication initiated from a foldback DNA near the terminus (14).

TABLE 9 Major genes of T7

Gene	Function
Class I (early)	
0.3	Anti-restriction
0.7	Protein kinase
1	RNA polymerase
1.2	Replication
1.3	DNA ligase
Class II (middle)	
2	Anti-RNA polymerase
2.5	DNA binding
3	Endonuclease
3.5	Lysozyme
4	Primase
5	DNA polymerase
6	Exonuclease
Class III (late)	
7, 7.3	Host range
8	Head-tail joining
9	Head assembly
10	Capsid
11, 12	Tail
13, 14, 15, 16	Core
17	Tail fibers
17.5	Lysis
18, 19	DNA maturation

Phages T1 and T5

The remaining two groups of T phages are represented by T1 and T5, both of which have double-stranded DNA genomes with terminal repeats.

T1 DNA is about 46 kb in length, with a terminal repetition of about 2.8 kb. It employs headful packaging, but (unlike T4) processive packaging from a concatemer is initiated from a unique *pac* site, so the circular permutation is highly nonrandom. T1 infection induces extensive degradation of host DNA. Replication of T1 requires the participation of both phage and host enzymes.

T5 DNA is about 125 kb in length, with a terminal repeat of about 10 kb. The DNA sequence is nonpermuted. Early in infection, T5 induces degradation of host DNA and hydrolysis of the deoxynucleotides to deoxynucleosides (which are not reutilized) and P_i. A feature unique to T5 and its relatives is that DNA injection occurs in two stages. After attachment, about 8 kb from one end of the virion DNA enters the cell. DNA transfer then stops and awaits transcription and translation of the very

early genes introduced on this first-step transfer DNA. Some of these gene products are needed for the subsequent entry of the rest of the T5 DNA into the cell.

TEMPERATE DOUBLE-STRANDED DNA PHAGES

All known temperate phages have double-stranded DNA in the virion. They can be classified according to their mode of association with the bacterial chromosome.

TABLE 10 Features of λ

Virion:	Icosahedral head, T=7; two major capsid proteins, several minor ones
Nucleic acid:	Linear ds, 48,502 bp; terminal complementary 12-nt 5′ overhangs
Receptor:	Maltose transport protein LamB; original isolate also used OmpF
Regulation:	In lysogens, phage genes turned off by repressor In lytic cycle, genes turned on by antitermination, early genes turned off through repression by Cro
Effects on host:	Inhibition of host ExoV by λ Gam protein.
Replication:	Bidirectional from unique origin on circularized DNA, using phage initiator protein and host primase, helicase, and replicase
Packaging substrate:	Rolling-circle replication at late times Linear concatemer (tail of rolling circle)
Lysis:	Holin and transglycosylase
Growth parameters:	Burst size 100 Latent period 35 min
Related phages:	21, ϕ80, ϕ81, ϕ82, 424, 434, HK022, HK97, Atlas, P22, L, LP-7

Phages That Insert Their DNA into Chromosomes by Site-Specific Recombination

The most thoroughly investigated temperate phage is λ (Tables 10, 11) (28). Both its natural biology and its experimental study are better appreciated in the context of the group of related phages (lambdoids) to which it belongs. That group is discussed first. Two unrelated phages (P2 and P4) illustrate important variations on the λ theme.

Lambdoid Phages as a Natural Gene Pool. Several collections made at widely different times in various geographical locations include an assortment of phages all of which can develop together with λ in mixed infection, have a common genetic map, and can recombine with λ to give viable hybrids. These include coliphages (such as λ, 21, 434, ϕ80, and HK022) and *Salmonella* phages (such as P22). Comparative sequence analysis shows that these phages are all related to one another by homologous recombination in nature. Any two of these lambdoid phages share some segments of similar sequence where recombination can occur. The natural recombinational pool includes not only active phages but complete and incomplete prophages as well.

As a group, the lambdoid phages are highly polymorphic, both in DNA sequence and in functional specificity. In some cases (as in their repressor genes), their elements have equivalent modes of action. Repression in phage 434 follows all the major rules established for λ; the differences between the two phages come to light only when they are tested against each other. A major use of heterologous systems has been in defining the determinants of specificity; for example, helix-swap experiments show that the entire difference in recognition specificity rests in the helix-turn-helix segment of the repressor protein (51). If that part of the protein is of λ origin, the repressor recognizes operator sites from λ rather than from 434, regardless of the origin of the rest of the protein.

In other cases, various phages use disparate mechanisms to achieve a common end. For example, P22 and other *Salmonella* lambdoids differ from λ in virion shape, mechanism of packaging, and repressor control.

Variation in specificity is seen for repression, integration, early gene regulation, late gene regulation, and attachment to the cell surface. Natural recombination creates a combinatorial array of phage types.

Bacteriophage λ Life Cycle and Major Variations Thereon.

Virion. The λ virion has an isometric T=7 head with two major capsid proteins, gpE and gpD. GpE can form functional capsids in the absence of gpD, but their diameter and packaging capacity are reduced. In addition, the capsid contains minor protein gpB, processed derivatives of gpB and gpC, and portal proteins gpW and gpFII. Virion DNA is linear and double stranded (48,502 nt) with 12-bp single-stranded complementary 5′ ends. The capsid is attached to a long flexible tail composed of several proteins, culminating in a single fiber of gpJ, which recognizes the cellular receptor, maltose transport protein LamB. Following contact between gpJ and LamB, phage DNA is released from the head through the tail core. The ends of the entering DNA find one another and are ligated together to form a covalently closed ring.

TABLE 11 Major genes of λ

Genes	Function
nul, A	DNA packaging
*W, B, C, nu3, D, E, F*I, *F*II	Head formation
Z, U, V, G, T, H, M, L, K, I, J	Tail formation
int, xis	Integration, excision
exo, bet, gam	Recombination
*c*III	Stabilization of gpcII
N	Early antitermination
*c*I	Repression (primary)
cro	Repression (secondary)
*c*II	Turn on *c*I, *int*
O, P	Replication
Q	Late antiterminator
S, R, Rz	Lysis

The virions of most lambdoid coliphages are constructed similarly to λ, but the lambdoid *Salmonella* phages follow a different plan. Whereas 11 λ genes are needed for tail formation, the short stubby tail of P22 consists of a single protein (gp9), probably a trimer, which recognizes lipopolysaccharides of the cell envelope. Gp9 has an endorhamnosidase activity apparently needed for DNA penetration following adsorption (7).

Pathway to lysis.

λ being a temperate phage, some of the infected cells follow a pathway leading to phage production and lysis, whereas others survive infection harboring the phage genome in latent form. Although the choice between the two pathways is not made immediately, it is convenient to outline first the entire course of events in those cells that eventually lyse.

Immediately after infection, short transcripts are initiated from three major promoters, p_L, p_R, and $p_{R'}$. The leftward transcript from p_L encodes a single protein, gpN; the rightward transcript from p_R encodes the Cro protein; and the 186-nt $p_{R'}$ transcript contains no genes. Cro and gpN are both regulatory proteins, Cro acting as a repressor and gpN as an antiterminator. Antitermination by gpN is effected by interaction between gpN and mRNA containing special sequences (*N*-utilization, or *nut* sites) that lie in the p_L and p_R transcripts. Several proteins (gpN and host proteins NusA, NusB, NusG, and S10) create an RNA loop that connects the *nut* site to the RNA polymerase downstream from it. This interaction reprograms the RNA polymerase so that it no longer recognizes the ordinary signals for transcription termination but rather continues through them. The p_L transcript is then extended past a terminator, t_{L1}, downstream from *N*, to include (among others) genes *redX* and β, which promote genetic recombination and serve an ancillary

function in DNA replication. Beyond the first rightward terminator t_{R1}, the p_R transcript includes replication genes *O* and *P* and the activator of late transcription, *Q*.

Most lambdoid phages have early antiterminator genes similar to *N*, although with different specificities; λ gpN does not act on ϕ80 *nut* sites or vice versa. An exception is phage HK022. In the same location as *N*, HK022 has a gene *nun* (slightly homologous to *N*). The *nun* gene plays no known role in HK022 infection, but it interferes with replication of λ. If λ infects an HK022 lysogen, Nun causes premature termination of λ early messages, competing with λ gpN for binding to *nut* sites. This exclusion of λ may be important in the natural competition between the two phages. Although Nun does not cause antitermination in HK022, *nut* site analogs are present and cause downstream antitermination, apparently effected entirely by host genes (38).

As time proceeds, Cro accumulates and binds to the operator sites controlling p_L and p_R, reducing the rate of transcription of the early gene. At the same time, gpQ antiterminates transcription from $p_{R'}$, which results in expression of the late genes controlling lysis and virion formation. By this time, replication is well under way, increasing the copy number of late genes available for activation. Thus the turn-on of late genes is controlled both by gpQ concentration and by gene dosage.

Bidirectional replication is initiated from an origin site within the *O* gene. gpO recognizes the specific origin sequence and also the second replication protein, gpP; gpP in turn binds to the host prepriming protein DnaB, recruiting it into the replication complex. (In P22, the gene analogous to *P* is homologous to host *dnaB*, which is unneeded for P22 replication.) Bidirectional replication is succeeded by rolling-circle replication, where multigenomic tails of double-stranded DNA are spun out from replicating circles.

These concatemers are the substrate for DNA packaging. Empty proheads form and later fill with DNA, which is cut at recognition sequences (cohesive end, or *cos* sites) at the termini of virion DNA, generating 12-nt protruding 5′ ends. The full *cos* sequence consists of three functional components (*cosQ*, *cosN*, and *cosB*), extending over about 150 bp. The cutting enzyme, terminase (not present in the mature virion but probably capsid associated at the time of cutting), specifically recognizes *cosB* (the binding site). The terminase-*cosB* interaction is phage specific; λ terminase does not bind to 21 *cosB* and vice versa. The actual cutting then takes place at the adjacent nicking site, *cosN*. To complete packaging, *cosQ* is needed to signal cutting of *cosN* at the other end of the packaged monomer. In *cosQ* mutants, DNA enters the head but fails to be cut properly (15). Packaging proceeds from left to right. Several monomers can be packaged processively, and *cosB* recognition is required only to initiate packaging of the leftmost monomer (18). From concatemers with λ *cosB* at the leftmost *cos* site and 21 *cosB* at a subsequent site, λ terminase can package the entire chain of monomers, whereas 21 terminase packages only rightward from 21 *cosB*.

Packaging in P22 differs qualitatively from packaging in λ and resembles more that of T1. Virion DNA does not have unique ends as in λ or T7 but rather represents a headful of DNA cut from a concatemer. P22 differs from T4 in using a specific interaction site (*pac*) to initiate packaging, with subsequent rounds of processive packaging coming progressively out of phase (48). The map location of the P22 *pac* site is very close to that of the λ *cos* site (between genes for lysis and virion formation), perhaps suggesting an ancient relationship between the two systems.

λ has three lysis genes, *S*, *R*, and *Rz* (54). The *R* gene encodes a transglycosylase that breaks bonds between *N*-acetylglucosamine residues (like a lysozyme), and the *S* gene encodes a holin that destroys the inner membrane and allows access of gpR to the murein layer. The action of gpRz is unknown, and it is needed only under special conditions, such as high concentrations of divalent cations. Unlike λ, all other lambdoid phages that have been tested make lysozymes rather than transglycosylases. The genes for these two enzymes are unrelated in sequence but allelic in map position. The timing of lysis is apparently controlled by the accumulation of gpS. Gene *S* is translated from two alternative sites. The larger peptide inhibits the holin activity of the smaller peptide, but this inhibition has no substantial effect on the time of lysis under laboratory conditions (12).

Pathway to lysogeny. Among cells infected by λ, some follow the lytic pathway whereas others survive to become lysogenic. The critical element determining which pathway is followed is the protein product of gene *c*II, which elicits expression of genes needed for two essential aspects of lysogeny, phage gene repression and prophage integration. Repression is caused by the product of gene *c*I (the repressor), which binds to the same operator sites as Cro. When these sites are occupied by repressor, transcription ceases from the early promoters p_L and p_R. Each of these promoters is controlled by a tripartite operator, whose individual binding sites have different orders of loading for repressor and Cro. The right operator is more interesting, because it controls not only rightward transcription of *cro, O, P,* and *Q* but also leftward transcription (from promoter p_M) of *c*I.

Repressor binds tightly and cooperatively to the o_{R1} and o_{R2} sites. The binding has two consequences: repression of rightward transcription from pR and stimulation of leftward transcription from p_M (Table 12). At higher repressor concentrations, repressor can also bind to o_{R3} and repress transcription from p_M. Cro binds noncooperatively to the three sites. Its greatest affinity is for o_{R3}, where it represses leftward transcription. At higher concentrations (such as those realized late in the lytic cycle), Cro binds to o_{R2} and o_{R1} and represses rightward transcription.

Repressor thus can stimulate its own synthesis from p_M, but the low rate of spontaneous transcription from p_M is inadequate to initiate repression, especially in the face of the competing activity of Cro. Establishment of repression is effected by leftward transcription from a promoter, p_E, that lies within gene *c*II. Gene *c*II is part of the rightward transcript immediately downstream from *cro* and t_{R1}. The p_E transcript proceeds leftward through *cro* (antisense direction) and into *c*I (sense direction). So if a sufficient high concentration of gpcII builds up in the cell, abundant repressor is produced, and its stimulation of *c*I transcription from p_M then makes repressor synthesis self-sustaining. Commitment to lysogeny can therefore be equated to achievement of a high concentration of gpcII.

TABLE 12 Regulation of λ p_R and λ p_M[a]

Leftward transcription of *c*I from p_M	Occupancy of operator sites[b]			Rightward transcription of *cro, O, P, Q* from p_R
	o_{R3}	o_{R2}	o_{R1}	
Low	—	—	—	ON
OFF	C	—	—	ON
OFF	C	C	C	OFF
ON	—	R	R	OFF
OFF	R	R	R	OFF

[a]Modified from reference 40. Only the most important states are shown.
[b]—, unoccupied; C, Cro; R, repressor.

Why does not every infected cell achieve such a high gpcII concentration? The gpcII protein is metabolically unstable (mainly due to proteolysis by a host protease encoded by gene *hflA*). Another λ gene, *c*III (to the left of *N*), makes a protein that stabilizes gpcII to some extent. Because of this instability, high gpcII concentration requires a high rate of synthesis. Because Cro accumulates with time and increasingly represses *c*II transcription, a high rate of gpcII synthesis must be attained early or not at all. Once the Cro concentration has reached a high level, the cell is committed to lysis. Thus in each cell there is a race between Cro synthesis and gpcII synthesis; because gpN is needed to synthesize gpcII but not Cro, gpN promotes lysogeny.

It is noteworthy that, under a wide range of ambient conditions, a substantial fraction of the population goes toward lysis. Some conditions, such as high multiplicity of infection or the presence of divalent cations or cyclic AMP, increase the proportion of cells that go toward lysogeny, but seldom to 100%. Uncertainty appears to be built into the system.

gpcII stimulates transcription from promoters with a characteristic –35 sequence ($TTGCN_6TTGC$) of which λ contains three: p_E, p_I, and p_{AQ}. Transcription from p_E makes repressor, whereas p_I controls the *int*

gene, whose product (integrase) inserts λ DNA into the host chromosome by site-specific recombination. Thus in those cells that survive infection by shutting off p_L and pR with repressor, phage DNA is efficiently inserted into the chromosome. The integrase gene can also be transcribed from p_L, but that part of the p_L transcript is rapidly degraded by RNase III cleavage of a sequence (*sib*) downstream from *int*, followed by exonucleolytic degradation. This degradation does not affect p_I transcripts, which terminate at *sib*, whereas the p_L transcript extends through *sib* because of the antiterminating effect of gpN. The p_{AQ} promoter lies within the *Q* gene and initiates an antisense transcript which probably interferes with *Q* expression and therefore with the turn-on of late genes (30).

Integration is effected by site-specific recombination between the covalently closed circular phage molecule and the bacterial chromosome. The sites on both partners are unique and share 15 bp of identical sequence. The phage site is immediately downstream from the *int* gene, and the bacterial site is at 17 min on the *E.* coli map (between *chlD* and *bio*). Within the 15-bp identity, two strands of DNA (one from each partner) first exchange to produce a Holliday structure, whose branch then migrates 7 bp to the right, followed by a resolving exchange between the other two DNA strands. All parental atoms are conserved in the product, and no external energy source is required.

Because the phage integration site (*attP*) is not at the same location as the cutting site for packaging (*cos*), the linear gene order of the prophage is a cyclic permutation of the gene order on the virion DNA.

Maintenance of and exit from the lysogenic state.

Lysogeny, once established, is relatively stable. Repressor continues to stimulate its own synthesis, and the inserted prophage is passively replicated as part of the bacterial chromosome. Almost all phage genes (including *c*II and *c*III) are turned off. In absence of gpcII, *int* is transcribed only at a very low rate. Once the prophage is inserted, integrase is not needed.

Rarely (about once every 10^4 cell divisions), repression breaks down, DNA is excised from the chromosome, and the lytic pathway is set in motion. The resulting spontaneous phage production is what led to the discovery of lysogeny. The rate can be increased by conditions (such as UV irradiation or thymine starvation) that turn on the host SOS repair system. Under such conditions, the host RecA protein is converted to a form in which it promotes cleavage of λ repressor into inactive fragments. RecA-promoted cleavage of the host LexA repressor initiates the SOS response, and λ repressor mimics LexA in this respect. A possible reason for the negative regulation of repressor synthesis by repressor binding at o_{R3} (Table 12) is to keep the repressor concentration low enough so that the phage can readily respond to conditions unfavorable for host growth by entering the lytic cycle and escaping to infect other cells. As in the establishment of lysogeny, λ seems designed to live at the edge.

Excision of λ prophage from the chromosome requires not only integrase but a second phage-specified protein, excisionase. Exactly how excisionase promotes the reversal of the integration reaction is not understood. The excisionase gene is located immediately upstream from *int* and is cotranscribed with *int* from the p_L promoter. This p_L transcript is not subject to $3' \rightarrow 5'$ digestion because the *sib* site is distal to *att*P and therefore lies at the other end of the prophage.

Because the *sib* site (unlike most regulatory sites) lies downstream from the gene it affects, its effect has been dubbed retroregulation. The term may be misleading, in that *sib* is not a regulatory site in the usual sense; i.e., its negative effect on *int* expression is independent of and unresponsive to any external signals. Its only differential effect is on the various *int* transcripts. It reduces *int* expression from the antiterminated p_L transcript of an infecting phage but not from the p_I transcript, which terminates at *sib*, or from the p_L transcript of a derepressed prophage, which does not contain *sib*. The status of retroregulation as a general regulatory strategy remains to be established.

Genes that affect host phenotype. At least two genes of λ (*lom* and *bor*) confer properties on lysogens that may contribute to the pathogenicity of some *E. coli* strains (5). Both these genes are

expressed from weak promoters in the prophage. The *rexA* and *rexB* genes of λ are cotranscribed with *c*I and make λ lysogens unable to support growth of *rII* mutants of phage λ. They may also confer a selective advantage to their bearers under certain growth conditions (13).

Phage P2. Phage P2 and its relatives resemble the lambdoid phages in their mode if integration but differ in several other properties. The P2 virion has an icosahedral head (T=9), with one major capsid protein (present mostly in processed form) and at least two minor proteins, and a tail with contractible sheath, baseplate, and tail fibers (Table 13) (9). The linear double-stranded DNA is nonpermuted, with 19-nt 5′-terminal overhangs. Although P2 and λ are not significantly related in DNA sequence, the P2 tail fiber gene is related to a λ gene for side tail fibers (present in the original λ isolate but inactivated by mutation in most laboratory strains [27]) and to genes from several other otherwise unrelated phages (T4, Mu-1, e14) (22). Apparently the advantages of new attachment specificities can sometimes be sufficiently strong that rare incorporations of determinants from heterologous sources are selected in nature.

TABLE 13 Features of P2

Virion: Icosahedral, T=9, sheathed tail with baseplate and fibers
Nucleic acid: Linear ds DNA, ~33 kb with terminal 19-nt complementary 5′ overhangs
Receptor: Lipopolysaccharide
Regulation: In lysogens, lytic cycle genes shut off by repression In lytic cycle, late gene expression activated by P2 Ogr protein
Effects on host: P2 Old protein lethal for *recB recC* hosts
Replication: Unidirectional from fixed origin
Packaging substrate: Monomer circles
Lysis: Phage lysin
Growth parameters: Burst size 120 Latent period 30 min
Related phages: 186, 299, 18, PK, W-ϕ, ϕ-D, HK239, PSP3

In the lytic pathway, phage genes are transcribed by host RNA polymerase. After replication has started, a phage protein (Ogr) activates late genes for virion structure and lysis (Table 14) (9). Replication is unidirectional from a unique origin and requires a phage initiation protein, gpA (which acts only in *cis* [32]), and host proteins DnaB, DnaE, DnaG, and RepA. The major replication products seem to be monomer circles, which are also the preferred packaging substrate.

In the lysogenic pathway, phage genes are repressed by P2 gpC. Integration takes place by site-specific recombination: as in λ, efficient excision requires, in addition to integrase, a second phage protein, Cox. As with λ, P2 lysogens spontaneously enter the lytic cycle at a low rate and produce some phage, but, unlike λ, P2 lysogens are not induced to lyse by activation of the SOS system.

Linkage mapping of P2 offers some interesting contrasts to that of λ. The rate of recombination per kilobase between coinfecting phages is much lower in P2 than in λ, even lower than for a λ *red* mutant, where only the host Rec system is used. Because general recombination is so low, site-specific recombination at the *att* site accounts for most of the observed recombination between P2 phages. This high rate of recombination creates an apparently linear map terminating at *att.* In *int* mutants of P2, the map becomes circular rather than linear. In λ (either int^+ or int^-), the linkage map is linear, terminating at *cos.*

These findings may result from the packaging mechanisms of the two phages. In many systems, most recombination is provoked by molecular ends; in λ, *cos* cutting of concatemers apparently is a major source of recombinogenic ends (43), hence the higher rate of recombination in λ. Where recombination generates mixed concatemers, *cos* cutting in λ has the same genetic consequences as a crossover at *cos;* i.e., markers that flank a *cos* site are always separated from one another at packaging, hence the linearity of the λ map. Packaging of monomer circles in P2 does not have that consequence, because the markers flanking *cos* end up in the same virion.

TABLE 14 Major genes of P2

Genes	Function
Q, P, O, N, M, L	Head synthesis
K	Lysis
R, S, V, J, H, G, F, E, T, U, D	Tail synthesis
int	Integration
C	Repression
cox	Excision
B, A	Replication

Phage P4. Phage P4 is a satellite virus which can carry out a complete lytic cycle only in the presence of P2 or its relatives. P2 can be present either as a coinfecting phage or as a prophage. Most experiments on P4 have used P2 lysogens both as infection hosts and as plating bacteria. In absence of P2, P4 can infect, replicate, and lysogenize, but does not produce virions or lyse the cell.

P4 DNA is linear and double stranded, 11,624 bp in length (Table 15) (33). This is about one-third the size of P2 DNA, and the head, though built of the major P2 capsid protein, is correspondingly smaller. Following infection, a P4 leftward transcript is made, which includes genes *c*I, ε, and α (Table 16) (33). This induces some replication (but not excision) of P2 prophage and expression of P2 late genes, activated by P2 Ogr. P4 replication is initiated by gpα (which performs primase and helicase functions as well), using host DnaE for chain elongation. Ogr also activates the P4 late (rightward) operon, which includes *sid*, δ, and *psu.* Although not itself a virion component, Sid reprograms the assembly of P2 capsid proteins to produce P4-size heads. Gpδ, like P2 Ogr, turns on the P2 late operons.

In those cells that become lysogenic, the *c*I product accumulates and shuts off early transcription. The *c*I product is not a protein repressor but rather a heterogeneous collection of short RNA transcripts of *c*I DNA. These products are thought to bind to RNA sequences downstream from *c*I, causing termination of transcription. Short RNAs from this part of the genome are present in lysogens but also predominate late in infection, suggesting that they function not only to silence lytic functions in P4 lysogens but also to reduce early gene expression late in the lytic cycle.

P4 can lysogenize nonlysogens as well as P2 lysogens. The prophage is generally inserted by site-specific recombination catalyzed by P4 integrase. In nonlysogens, P4 can also persist as a plasmid (whether or not there is also an integrated copy), and the plasmid state can be stabilized by mutations in P4.

TABLE 15 Features of P4

Virion:	Icosahedral capsid (T=4), made from P2 proteins plus P4 protein Psu
	P2-determined tail with contractile sheath, baseplate, tail fibers
Nucleic acid:	Linear ds DNA, 11,624 bp, with terminal 19-nt complementary 5′ overhangs
Receptor:	Lipopolysaccharide
Regulation:	In lysogens, genes turned off by premature termination
	In lytic cycle, P4 gpE turns on P2 early genes, gpδ turns on P2 genes, and P2 Ogr turns on late genes of both phages
Effects on host:	Lethal P4 protein, Kil
Replication:	Bidirectional from unique origin
	P4 gpα serves as initiator, primase, and helicase
Packaging substrate:	Monomer circles
Lysis:	P2 lysin
Growth parameters:	Burst size 300
	Latent period 60 min
Related phage:	ϕR73

P4-related prophages are common in natural strains of *E. coli.* One of these (ϕR73) is special in that it includes a retron element, which makes single-stranded DNA through a mechanism requiring retron-coded reverse transcriptase (47). The retron serves no known function in the biology of the phage, which may act simply as a vehicle for its dissemination.

Phages That Use Transposition To Replicate and Lysogenize

Like λ, phage Mu-1 and its relatives lysogenize by integrating their DNA into the chromosome. However, the mechanism of integration is very different.

Mu-1 has an icosahedral virion, with a tail, that packages a double-stranded genome of about 43 kb (Table 17) (23). The DNA at the molecular ends (about 100 bp at the left and 2 kb at the right) is host DNA, different from one virion to the next. When Mu-1 infects a cell, the replication genes *A* and *B* are transcribed early (Table 18) (23). Their products catalyze both the initial nonreplicative insertion into the chromosome and subsequent replicative transposition; gpA is the primary transposase, whose activity is enhanced by gpB. Mu-1 also has a repressor gene (*c*) whose product shuts off transcription of these and other phage genes in lysogens, a secondary repressor, Ner, which (like λ Cro) turns off transcription both of the repressor gene and of the replication genes, and an activator of late transcription (gpC). Unlike λ, Mu-1 integration is prerequisite to lytic pathway replication as well as to lysogeny.

TABLE 16 Genes of P4

Genes	Function
gop, β, *c*II	Nonessential
int	Integration
α	Replication
orf106, *orf151*	Open reading frames of unknown function
ε	Turn-on of P2 early genes
kil	Host killing
*c*I	Immunity
orf199, *orf88*	Open reading frames of unknown function
sid	Morphogenesis
δ	Turn-on of P2 late functions
psu	Antitermination

TABLE 17 Features of Mu-1

Virion: Icosahedral, tail with contractile sheath
Nucleic acid: Linear ds DNA, ~43 kb; host sequences at termini
Receptor: Lipopolysaccharide
Regulation: In lysogens, phage genes turned off by repressor In lytic cycle, early genes turned off through repression by Ner, late genes turned on by gpC
Effects on host synthesis: Growth inhibition by DNA modification
Replication: Replicative transposition of inserted phage DNA
Packaging substrate: Phage DNA inserted into host chromosome
Lysis: Phage lysin
Growth parameters: Burst size 200 Latent period 60 min
Related phages: D3112, B3, D108

Like some bacterial transposons, Mu-1 DNA can change location by either replicative or conservative transposition. In the first cycle following infection, nonreplicative transposition moves the Mu-1 DNA from the infecting molecule (where it is flanked by host DNA) into a random location on the chromosome. In the lytic pathway, subsequent cycles of transposition are replicative, so that one semiconserved copy remains at its original location and a second copy is inserted elsewhere. Repeated cycles of replicative transposition from one chromosomal site to another cause both inversion and fragmentation of chromosomal sequences. DNA is packaged from such chromosomal sites. The packaging system recognizes the left end of Mu-1 and cuts 50 to 150 bp to its left. The right cut site is determined by headful packaging, whose imprecision is evident in the length variation of host DNA from about 500 to 3,000 bp.

In the lysogenic pathway, repression is established, and the prophage remains inserted in its original location (with rare transpositions to new sites). Insertion can happen at any chromosomal site, sometimes resulting in gene disruption. The name Mu-1 (mutator) was based on the phage's ability to cause such disruptions.

TABLE 18 Genes of Mu-1

Genes	Function
c	Repression (primary)
ner	Repression (secondary)
A, B	Transposition, replication
C	Stimulation of late transcription
D, E, H, F, G, I, J	Head synthesis
K, L, M, Y, N, P, Q, V, W, R, S	Tail synthesis
gin	G segment inversion
mom	DNA modification

Mu-1 has two additional special features: a DNA modification activity (*mom*) that helps it to overcome host restriction systems and a segment (called the G segment) of about 3 kb (found also in P1 phage) that can invert by the action of a site-specific recombinase (encoded by the adjacent *gin* gene) acting on an oligonucleotide inverted repeat at the ends of the segment. The fiber protein of the Mu-1 tail is coded by a gene that crosses one boundary of the invertible segment, so inversion switches the C-terminus of this protein to one with a different host specificity. This strategy allows Mu-1 to infect a wider range of potential hosts than might otherwise be possible.

Phages That Lysogenize by Plasmid Formation

A final group of temperate phages do not insert their DNA into the chromosome, but instead can be stably maintained as plasmids. The prototype is coliphage P1 (Table 19) (53).

TABLE 19 Features of P1

Virion: Icosahedral head; tail with contractile sheath, baseplate, tail fibers
Nucleic acid: ds DNA, ~100 kb, circularly permuted, ~7-kb terminal repeat
Receptor: Lipopolysaccharide
Regulation: In lysogens, some phage genes turned off by repressor binding to numerous dispersed sites In lytic cycle, host polymerase transcribes both early and late genes
Effects on host: Introduces restriction-modification system
Replication: Infecting molecules circularize through specific (*lox*-Cre) or general recombination Prophage replicates as circular plasmid, bidirectionally from *oriR* In lytic cycle, bidirectional replication from *oriL*, rolling-circle replication to give ds concatemers
Packaging substrate: Linear concatemer (tail of rolling circle)
Lysis: Probably lysozyme plus holin
Growth parameters: Burst size 80 Latent period 45 min
Related or similar phages: P7, ϕW39, j2, p15B, D6

P1 DNA is double stranded and terminally repetitious, about 100 kb in length. Soon after infection, the terminal segments cross over to form a circular molecule, shorter than the linear form by the length of the terminal repeat (about 7 kb). This crossover is frequently effected by a site-specific recombinase (Cre) acting on its target sites (*lox*) in the terminal repeat. After circle formation, P1 DNA can replicate either as theta forms and later as rolling circles (lytic pathway) or (from a different replication origin) as theta forms, once per division cycle (lysogenic pathway).

In the lytic cycle, the packaging substrates are the linear concatemers formed by rolling-circle replication. As in P22 and T1, headful packaging is initiated at a specific *pac* site and continues processively for several rounds. Because the *pac* site is close to *lox, lox* is present in the terminal repeats of virions formed during the first rounds of packaging and can be used for circularization following infection.

In the lysogenic state, some lytic cycle genes are repressed. Cre continues to be made. Repression is effected by the product of a phage gene (*c*1). Like P22, P1 makes an antirepressor, which is itself repressed by a secondary repressor. Replication of the P1 plasmid DNA is determined by a 1.5-kb segment that includes a replication origin (*oriR*), the gene for a replication protein (*repA*), and a regulatory sequence (*incA*) that sequesters RepA. Replication requires host proteins, including DnaA (53). Replication control may be accomplished through plasmid-plasmid pairing, where the *incA* sequences of one plasmid bind to the *oriR* site of the other through RepA protein, so that the number of plasmid copies is limited to two per cell (1). Immediately adjacent to the replication segment is 2.7 kb of DNA that controls partitioning of these two plasmids at cell division. Partitioning requires both *cis*- and *trans*-acting factors and is facilitated by the *lox*-Cre system, which resolves into monomers any dimeric molecules generated during replication or general recombination. The resulting high fidelity of transmission is measured by a rate of plasmid loss of about 10^{-5} per cell per generation.

SUMMARY

Bacteriophages are important components of the natural environments of *E. coli* and *Salmonella typhimurium* and (through lysogeny) of their genomes as well. Phages are a highly diverse group, whose only common feature is the packaging of their genomes into protein coats to form infectious virions. In each phage's life cycle, there is a continual interplay between host products and phage products.

LITERATURE CITED

1. **Abeles, A. L., and S. J. Austin.** 1991. Antiparallel plasmid-plasmid pairing may control P1 plasmid replication. *Proc. Natl. Acad. Sci. USA* **88:**9011–9015.
2. **Aoyama, A., and M. Hayashi.** 1986. Synthesis of bacteriophage ϕX174 *in vitro:* mechanism of switch from DNA replication to DNA packaging. *Cell* **47:**99–106.
3. **Artz, S., D. Holzschu, P. Blum, and R. Shand.** 1983. Use of M13 mp phages to study gene regulation, structure and function: cloning and recombinational analysis of genes of the *Salmonella typhimurium* histidine operon. *Gene* **26:**147–158.
4. **Barbas, C. F., A. S. Kang, R. A. Lerner, and S. Benkovic.** 1991. Assembly of combinatorial antibody libraries on phage surfaces. The gene IV site. *Proc. Natl. Acad. Sci. USA* **88:**7978–7982.
5. **Barondess, J. J., and J. Beckwith.** 1990. A bacterial virulence determinant encoded by lysogenic coliphage lambda. *Nature* (London) **346:**871–874.
6. **Benzer, S.** 1961. On the topography of the genetic fine structure. *Proc. Natl. Acad. Sci. USA* **47:**403–415.
7. **Berget, P. T., and A. R. Poteete.** 1980. Structure and function of the bacteriophage P22 tail protein. *J. Virol.* **34:**234–243.
8. **Berkhout, B., B. F. Schmidt, A. Van Strien, J. H. Van Boom, J. Van Westrenen, and J. Van**

Duin. 1987. Lysis gene of bacteriophage MS2 is activated by translation termination at the overlapping coat gene. *J. Mol. Biol.* **195:**517–524.
9. **Bertani, L. E., and E. W. Six.** 1988. The P2-like phages and their parasite, P4, p. 73–144. *In* R. Calendar (ed.), *The Bacteriophages*, vol. 2. Plenum Publishing Corp., New York.
10. **Bhattacharyya, S. P., and V. B. Rao.** 1993. A novel terminase activity associated with the DNA packaging protein gp17 of bacteriophage T4. *Virology* **196:**34–44
11. **Black, L., and M. Showe.** 1983. Morphogenesis of the T4 head, p. 219–245. *In* C. Mathews, E. Kutter, G. Mosig, and P. Berget (ed.), *Bacteriophage T4.* American Society for Microbiology, Washington, D.C.
12. **Bonovitch, M. T., and R. Young.** 1991. Dual start motif in two lambdoid *S* genes unrelated to λ S. *J. Bacteriol.* **173:**2897–2905.
13. **Campbell, J. H., D. Dykhuizen, and B. G. Rolfe.** 1978. Effects of the *rex* gene on phage λ lysogeny. *Genet. Res.* **32:**257–263.
14. **Chung, Y.-B., C. Nardone, and D. Hinkle.** 1990. Bacteriophage T7 packaging. III. A "hairpin" end formed on T7 concatemers may be an intermediate in the primary reaction. *J. Mol. Biol.* **216:**939–948.
15. **Cue, D., and M. Feiss.** 1993. A site required for termination of packaging of the phage λ chromosome. *Proc. Natl. Acad. Sci. USA* **90:**9290–9294.
16. **d'Herelle, F.** 1917. Sur un microbe invisible antagoniste des bacilles dysentériques. *C. R. Acad. Sci.* **165:**373–375.
17. **Doermann, A. H., F. A. Eiserling, and L. Boehner.** 1973. Genetic control of capsid length in bacteriophage T4. I. Isolation and preliminary description of four new mutants. *J. Virol.* **12:**374–385.
18. **Feiss, M., J. Sippy, and G. Miller.** 1985. Processive action of terminase during sequential packaging of bacteriophage λ chromosomes. *J. Mol. Biol.* **186:**759–771.
19. **Flemming, M., B. Deumling, and B. Kemper.** 1993. Function of gene 49 of bacteriophage T4. III. Isolation of Holliday structures from very fast-sedimenting DNA. *Virology* **196:**910–913.
20. **Geiduschek, E. P.** 1991. Regulation of expression of the late genes of bacteriophage T3. *Annu. Rev. Genet.* **25:**437–460.
21. **Goulian, M., A. Kornberg, and R. L. Sinsheimer.** 1967. Enzymatic synthesis of DNA. XXIV. Synthesis of infectious phage ϕX174 DNA. *Proc. Natl. Acad. Sci. USA* **58:**2321–2328.
22. **Haggard-Ljungquist, E., C. Halling, and R. Calendar.** 1992. DNA sequences of the tail fiber genes of bacteriophage P2: evidence for horizontal transfer of tail fiber genes among unrelated bacteriophages. *J. Bacteriol.* **174:**1462–1467.
23. **Harshey, R. M.** 1988. Phage Mu, p. 193–234. *In* R. Calendar (ed.), *The Bacteriophages*, vol. 2. Plenum Publishing Corp., New York.
24. **Haruna, I., and S. Spiegelman.** 1965. Specific template requirements of RNA replicases. *Proc. Natl. Acad. Sci. USA* **54:**579–587.
25. **Hausmann, R.** 1988. The T7 group, p. 259–290. *In* R. Calendar (ed.), *The Bacteriophages*, vol. 1. Plenum Publishing Corp., New York.
26. **Hayashi, M., A. Aoyama, D. L. Richardson, Jr., and M. N. Hayashi.** 1988. Biology of the bacteriophage ϕX174, p. 1–72. *In* R. Calendar (ed.), *The Bacteriophages*, vol. 2. Plenum Publishing Corp., New York.
27. **Hendrix, R. W., and R. L. Duda.** 1992. λ Papa is not the mother of all λ phages. *Science* **258:**1154–1158.
28. **Hendrix, R. W., J. W. Roberts, F. W. Stahl, and R. A. Weisberg.** 1983. *Lambda II.* Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
29. **Hershey, A. D., and R. Rotman.** 1949. Genetic recombination between host-range and plaque-type mutants of bacteriophage in single bacterial cells. *Genetics* **34:**44–71.
30. **Ho, Y. S., and M. Rosenberg.** 1988. Structure and function of the activator protein cII and its regulatory signals, p. 725–756. *In* R. Calendar (ed.), *The Bacteriophages*, vol. 2. Plenum Publishing

Corp., New York.
31. **Kastelein, R. A., E. Remaut, W. Feiss, and J. van Duin.** 1982. Lysis gene expression of RNA phage MS2 depends on a frameshift during translation of the overlapping coat protein gene. *Nature* (London) **285:**35–41.
32. **Lindahl, G.** 1970. Bacteriophage P2: replication of the chromosome requires a protein which acts only on the genome that coded for it. *Virology* **42:**522–533.
33. **Lindquist, B. H., G. Dehó, and R. Calendar.** 1993. Mechanisms of genome propagation and helper exploitation by satellite phage P4. *Microbiol. Rev.* **57:**683–702.
34. **Mathews, C. K., E. M. Kutter, G. Mosig, and P. B. Berget (ed.).** 1983. *Bacteriophage T4.* American Society for Microbiology, Washington, D.C.
35. **Model, P., and M. Russell.** 1988. Filamentous bacteriophage, p. 375–456. *In* R. Calendar (ed.), *The Bacteriophages,* vol. 2. Plenum Publishing Corp., New York.
36. **Mosig, G.** 1988. A map of distances along the DNA molecule of phage T4. *Genetics* **59:**137–151.
37. **Mosig, G., and F. Eiserling.** 1988. Phage T4 structure and metabolism, p. 521–606. *In* R. Calendar (ed.), *The Bacteriophages,* vol. 2. Plenum Publishing Corp., New York.
38. **Oberto, J., R. A. Weisberg, and M. E. Gottesman.** 1989. Structure and function of the *nun* gene and the immunity region of the lambdoid phage HK022. *J. Mol. Biol.* **207:**675–693.
39. **Parma, D. H., M. Snyder, S. Sobolevski, M. Mowray, E. Brady, and L. Gold.** 1992. The Rex system of bacteriophage λ: tolerance and altruistic cell death. *Genes Dev.* **6:**497–510.
40. **Ptashne, M.** 1988. *A Genetic Switch.* Cell Press and Blackwell, Cambridge, Mass.
41. **Shcherbakov, V. P., L. A. Plugina, and M. L. Nesheva.** 1992. Genetic recombination in bacteriophage T4: single-burst analysis of cosegregants and evidence in favor of a splice/patch coupling model. *Genetics* **131:**769–781.
42. **Simon, E.** 1965. Recombination in bacteriophage T4: a mechanism. *Science* **150:**759–763.
43. **Stahl, F. W., I. Kobayashi, and M. M. Stahl.** 1985. In phage λ, *cos* is a recombinator in the Red pathway. *J. Mol. Biol.* **181:**199–209.
44. Streisinger, G., J. Emrich, and M. M. Stahl. 1967. Chromosome structure in phage T4. III. Terminal redundancy and length determination. *Proc. Natl. Acad. Sci. USA* **57:**292–295.
45. **Studier, F. W., and J. J. Dunn.** 1983. Organization and expression of bacteriophage T7 DNA. *Cold Spring Harbor Symp. Quant. Biol.* **47:**999–1007.
46. **Studier, F. W., A. H. Rosenberg, J. J. Dunn, and J. W. Dubendoff.** 1990. Use of T7 DNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185:**60–89.
47. **Sun, J., M. Inouye, and S. Inouye.** 1991. Association of a retroelement with a P4 like cryptic prophage integrated into the selenocysteinyl tRNA gene of *Escherichia coli. J. Bacteriol.* **173:**4171–4181.
48. **Susskind, M. M., and D. Botstein.** 1978. Molecular genetics of bacteriophage P22. *Microbiol. Rev.* **42:**385–413.
49. **Twort, F. W.** 1915. An investigation on the nature of ultramicroscopic viruses. *Lancet* **189:**1241–1243.
50. **van Duin, J.** 1988. The single-stranded RNA bacteriophages, p. 117–168. *In* R. Calendar (ed.), *The Bacteriophages,* vol. 1. Plenum Publishing Corp., New York.
51. **Wharton, R. P., E. L. Brown, and M. Ptashne.** 1984. Substituting an α-helix switches the sequence specific DNA interactions of a repressor. *Cell* **38:**361–369.
52. **Womack, F. C.** 1963. An analysis of single-burst progeny of bacteria singly infected with a bacteriophage heterozygote. *Virology* **21:**232–241.
53. **Yarmolinsky, M. R., and N. A. Sternberg.** 1988. Bacteriophage P1, p. 291–438. *In* R. Calendar (ed.), *The Bacteriophages,* vol. 1. Plenum Publishing Corp., New York.
54. **Young, R.** 1992. Bacteriophage lysis: mechanism and regulation. *Microbiol. Rev.* **56:**430–481.