Molecular Biology and Biotechnology of Bacteriophage

Kazukiyo Onodera

Abstract The development of the molecular biology of bacteriophage such as T4, lambda and filamentous phages was described and the process that the fundamental knowledge obtained in this field has subsequently led us to the technology of phage display was introduced.

Keywords Phage T4, Phage lambda, Filamentous phage, Phage display, Application of Phage Display.
1 A Brief History

At the beginning I want to clarify why I have chosen the bacteriophages as the representative of various viruses for the people who is not familiar with virus research. By the late 1930s it had been shown that specific filterable viruses were the causative agents of some diseases in plants, some cancerous growth in animals, and of the lysis of some bacterial species. One of these viruses, causing the tobacco plant mosaic (TMV) disease, had been isolated in crystalline form by W.M. Stanley in 1935 [1]. That a self-replicating object like a virus can be crystallized as if it were inorganic material gave momentum to the notion of viruses as “living molecules.” Moreover it was shown that TMV was dissociated in acetic acid into protein and RNA and it was reconstructed in vitro as an infectious virus particle by adjusting to pH 7.0. The concept of “self assembly” emerged from this experiment but it was found later that the morphogenesis of viruses was different from self-assembly.

All cells contain double-stranded DNA as their genetic material. By contrast, viruses are known that have single-stranded RNA, double-stranded RNA, single-stranded DNA, or double-stranded DNA as their genetic material. This makes for interesting schemes of replication and gene expression. Many viruses are known that infect Bacteria and increasing numbers are known that infect Archaea.

I want to focus on three bacteriophages – T4 phage (virulent virus), lambda phage (template virus) and filamentous phage (not harmful to host). The study of genetic materials of virus has led to tremendous progress so that we can transfer the foreign genetic information to the cells that we want to manipulate and open the field of genetic engineering for the results to be used in bioindustry.

Therefore I will not talk about the genome of bacteriophages but rather discuss the structural features of these bacteriophages in this article.

It was thought that Felix d’Herelle recognized the particulate nature of phages [2]. However the real breakthrough was made by Max Delbrück who came to biology from physics. He organized the Phage group with Salvador Luria and Alfred D. Hershey in the late 1930s. Their meeting in 1940 marked the origin of the Phage Group. [3]

The members of this group were united by a common goal, namely the desire to understand how, during the brief half-hour latent period, the simple bacteriophage particle achieves its 100-fold self reproduction within the bacterial host cells.

To see the structure of virus particles, we need an electron microscope. From the beginning, Thomas F. Anderson, one of the first American electron microscopists, was a member of this group. Therefore the structure of the bacteriophage was explored extensively [4].

He brought a few stereoscopic pictures of adsorbed T2 phage to the Virus: 1950 meeting at Caltech to show it to Delbrück. Delbrück finally agreed that the phages are adsorbed by the tips of their tails and that none of the particles seem to enter the bacteria or their ghosts Other critical evidence was shown by the so-called Hershey, Chase experiment [5]. He described the experiment as follows.

Anderson had found that stirring the cell suspension in a blender prevented attachment of phage particles to bacteria, and perhaps Case and I (Hershey) should
have thought of using that machine first. Instead, we tried various grinding arrangements, with results that were not very encouraging. But when Margaret McDonald loaned us her blender, the experiments quickly succeeded.

A chilled suspension of bacterial cells recently infected with phage T2 is spun for a few minutes in a blender and afterwards centrifuged briefly at a speed sufficient to throw the bacterial cells to the bottom of the tube. One thus obtains two fractions: a pellet containing the infected bacteria and a supernatant fluid containing any particles smaller than bacteria.

Each of these fractions is analyzed for the radiophosphorus in DNA or radiosulfur in protein with which the original phage particles have been labeled. The results are:

1. Most of the phage DNA remains with the bacterial cells
2. Most of the phage protein is found in the supernatant fluid
3. Most of the initially infected bacteria remain competent to produce phage
4. If the mechanical stirring is omitted, both protein and DNA a sediment with the bacteria
5. The phage protein removed from the cells by stirring consists of more-or-less intact, empty phage coats, which may therefore be thought of as passive vehicles for the transport of DNA from cell to cell which, having performed that task, play no further role in phage growth

At present it is fair to say that bacteriophage injected their DNA into the host cell and bacteria produced virus particles through the subtle interaction between bacteria and phage genome.

However, the mechanism of entry of animal virus into the host cell is quite different from that of bacteriophage. The entry of influenza virus is described as follows.

The influenza virus nucleocapsid is of helical symmetry, about 6–9 nm in diameter and about 60 nm long. This nucleocapsid is embedded in an envelope that has a number of virus-specific proteins as well as lipid derived from the host. Because of the way influenza virus buds as it leaves the cell, the virus has no defined shape and is said to be polymorphic. There are proteins on the outside of the envelope that interact with the host cell surface. One of these is called hemagglutinin, so named because it causes agglutination of red blood cells. The red blood cell is not the type of host cell the virus normally infects, but contains on its surface the same type of membrane component, sialic acid, that the mucous membrane cells of the respiratory tract contain. Thus, the red blood cell is merely a convenient cell type for assaying agglutination activity. An important feature of the influenza virus hemagglutinin is that antibodies directed against this hemagglutinin prevent the virus from infecting a cell. A second type of protein on the influenza virus surface is an enzyme called neuraminidase. This enzyme breaks down the sialic acid component of the cytoplasmic membrane, which is a derivative of neuraminic acid. Neuraminidase appears to function primarily in the virus assembly process, destroying host membrane sialic acid that would otherwise block assembly or become incorporated into the mature virus particle.

It is helpful to choose simple paradigms to represent more complex systems.
The relative simplicity of the bacteriophage and of its bacterial host played an important role in the development of molecular biology. Therefore I want to choose three bacteriophages – T4 phage, lambda (λ) phage and filamentous phage.

I also want to discuss the molecular biology of these phages from the viewpoint of their structure and application of the results obtained from study of the structure to the biotechnology.

2 One Step Growth of Phage T4

2.1 Sequence of Events in T4-Infected Bacteria

In the 1950s chemical analysis of highly purified T-even phage by means of modern techniques fully substantiated that DNA and protein, in roughly equal proportions, constitute more than 90% of the dry weight of the particles. The total DNA complement of a single T-even phage was found to amount to 175 kb pairs, or about 6% that of the chromosome of the E. coli host cell.

Chemical analyses of the total T-even phage protein did not bring to light any facts; its amino acid composition resembles more or less that of the total E. coli protein. The phage protein is composed of at least five different types of polypeptides, of which the head protein makes up by far the major part. Each of the tail components – sheath, tube, base plate, tail pins, and tail fibers – contains one or more specific polypeptides. Chemical analysis of the phage DNA did show the difference.

The T-even phage DNA, unlike all other types of DNA, was found to have no cytosine. Instead of cytosine, it contains the cytosine analog 5-hydroxymethylcytosine (HMC). Thus T-even phage DNA obeys the \([A]=[T]\) and \([G]=[C]\) equivalence relation demonstrated by the Watson–Crick structure, provided that HMC, which can form the same three hydrogen bonds with G as can C, replaces C in the double helix (Fig. 1).

![Fig. 1](image-url) The unusual deoxynucleotide containing 5-hydroxymethylcytosine (HMC) found in the DNA of T-even phages (left) and the manner in which glucose is attached to some of the HMC bases (right) (Stent GS, Calendar R (1978) Molecular genetics: an introductory narrative, 2nd edn. W.H. Freeman, San Francisco, p 300)
The low \([G]+[HMC]\) content of 35\% places the T-even phage DNA in a compositional range quite remote from that of the DNA of its \(E.\ coli\) host, which has a \([G]+[C]\) content of 52\%. The T-even phage DNA was found to contain glucose attached to some of the hydroxymethyl groups of HMC. The one-step growth experiment demonstrated that the progeny of the infecting phage particle appear after a period of constant phage titer. The one-step growth experiment is a basic procedure for studying phage multiplication.

A dense suspension of growing bacteria is infected with phages, incubated for a few minutes to allow most of the phage particles to attach themselves to the bacteria, and then diluted with nutrient medium to a concentration that may range from one ten-thousandth to one millionth that of the suspension. The diluted culture is incubated further and samples plated on sensitive bacteria from time to time for plaque assay of the instantaneous number of infective units in the culture. The protocol and the results of a typical one-step growth experiment are shown in Fig. 2. The number of plaque-forming units in the culture remains constant for the first 24 min after infection. This initial period is the latent period. After some 24 min have elapsed, the number of plaque-forming units in the culture
begins to rise rapidly, until a final plaque is attained 10 min later. The period of time during which the number of plaque-forming units increases is the rise period, and the ratio of the final titer of the plateau to the initial titer of phage-infected bacteria is the burst size. The latent period represents the time that elapses between the moment at which the bacterial culture is infected with a phage stock and the moment at which the first infected cells in the culture lyse, thereby liberating into a medium a litter of progeny phage particles. The rise period represents the time span during which more and more of the infected bacteria lyse, and the final plateau of infectivity is attained when all the infected bacteria that are going to lyse have done so; no further phage multiplication occurs after this stage, since progeny and residual uninfected bacteria on the culture have been separated from each other by the high dilution of the culture just after the initial infection.

The burst size corresponds to the average number of progeny phage particles produced per infected bacterium, which in the experiments presented here amounts to about 100 phages per infected cell. After the conclusion of the latent period, when the intracellular phages have escaped from the host cell into medium, each progeny phage can form its own focus of infection on the agar surface (Fig. 3).

The phage enters accompanied by several minor peptides and protein. At least one protein, the product of phage 2 (gp2), probably enters attached to the lead part of the DNA. Examination of T4-infected bacteria at various times after infection by electron microscopy and by a combination of analytical methods reveals the pattern illustrated in the Fig. 2. Phage DNA increases after a brief delay; virion-specific
proteins begin to appear somewhat later and their appearance is soon followed by appearance of organized capsid precursors and then by the formation of mature infectious capsid. Some mRNA of viral specification is made promptly after glucosylated derivatives in phage DNA also provide infection and throughout the latent period. Bacterial mRNA and bacterial proteins stop being synthesized within a few minutes after the entry of phage DNA. Bacterial DNA is rapidly degraded to acid-soluble fragments and the “nuclear bodies” or DNA-containing areas of the bacterium become dispersed.

2.2 DNA synthesis

T-even phages have hydroxymethylcytosine (HMC) instead of cytosine in their DNA. In mature DNA, HMC has a pattern of glucosylation specific for each T-even phage: phage-coded glucosylating enzymes glucosylate HMC groups in DNA. HMC is a good chemical marker for T-even DNA. Its glucosylated derivatives in phage DNA also provide antigenic specificities recognizable by specific antibodies. By tracing DNA from the infecting phage it has been established that replication of T-even phage DNA takes place. The DNA molecules that enter from T-even virions into the bacterium are linear, circularly permuted, and have a terminal redundancy of 2–4% corresponding to 2–4 × 10^6 Da of DNA.

2.3 Phage-Coded Enzyme

The early transcription of newly entered phage DNA is done by the bacterial RNA polymerase. A number of phage proteins are made from such transcripts. Then the RNA becomes modified by the action of early phage products, its specificity changes, transcription of host genes stops even before the host DNA is broken up and other phage genes are transcribed.

Among the phage-coded proteins are enzymes belonging to several classes: (1) enzymes that produce unique phage DNA constituents such as HMC deoxynucleotides, or glucosylate these nucleotides, or destroy precursors of cytosine deoxynucleotides; (2) enzymes that play specific roles in DNA replication and recombination (Fig. 4); (3) enzymes that destroy host-cell DNA and make its nucleotides available for phage DNA synthesis; (4) enzymes that take part in the processing of virion proteins.

A critically important protein is gp32, the product of gene 32. This is a DNA unwinding protein that associates with single-stranded DNA stretches in stoichiometric amounts, one molecule per ten nucleotides. This protein can actually promote some unwinding of double-stranded DNA to allow initiation of replication to occur at sites where single-strand nicks are present. A peculiarity of gp32 is its ability to regulate its own rate of synthesis by a feedback mechanism: when more single-stranded regions are present in DNA, more gp32 is synthesized.
Fig. 4 The morphogenetic pathway of phage maturation has three principal branches leading independently to formation of heads, tails, and tail fibers, which then combine to form complete phage particles. The numbers refer to the T-even phage genes whose products are involved at each step. The solid arrows indicate the steps that have been shown to occur in extracts. Infection of *E. coli* with a phage carrying a mutation in any one of these genes leads to accumulation of the electron-microscopically visible structure shown immediately before the step in which that gene is involved, as well as of the last structure(s) of the other converging pathway branches (Stent GS, Calendar R (1978) Molecular genetics: an introductory narrative, 2nd edn. W.H. Freeman, San Francisco, p 319)
The six proteins needed for DNA synthesis may form a multienzyme complex. Thus T-even phage produce in their host bacterium a complex of enzymes and subsidiary proteins that make their replication independent of the DNA synthesizing systems of the bacterium.

2.4 Program of Phage Gene Expression

Based on the time of appearance of their messengers or gene products one can distinguish immediate early, delayed early, quasi-late, and late genes. The immediate early, whose mRNA is made even in the presence of an inhibitor of protein synthesis such as chloramphenicol, are transcribed from phage DNA by *E. coli* RNA polymerase either in vivo or in vitro. Another early class is that of delayed early genes, whose mRNA is initiated early but whose products are made only after a delay because they are coded by genes located distally from their promoters on the phage operons. The quasi-late genes are transcribed by RNA polymerase that has been altered by the action of certain products of immediate early phage genes. Particularly important are the loss or modifications of polypeptide σ and alterations in other subunits of polymerase, which thereby acquire new promoter-recognition specificities. Especially interesting is the alteration of the α-polypeptides of the enzymes, which acquire an adenosine diphosphoribose residue donated by nicotinamide adenosine dinucleotide. Host promoters are no longer recognized, and no more host specific RNA is transcribed. A second shift in RNA polymerase specificity occurs later (about 6–8 min in bacteria infected at 30 °C) and is due to the association of *E. coli* polymerase with three polypeptides including gp33 and gp55. At this point in the program two events occur: a more or less complete shut-off of the synthesis of the phage enzymes under control of phage gene 62 and the start of synthesis of all the late gene products, including the virion constituent proteins and phage lysozyme.

The switch from bacterial to phage program caused by changes in RNA polymerase has one important consequence. Since no more ribosomal RNA or proteins are made after T-even infection and since the phage does not code for new ribosomes, all phage protein synthesis must take place on bacterial ribosomes made before the first shift on program. This circumstance made possible one of the classical experiments that demonstrated the existence of mRNA as distinct from ribosomal RNA, confirming an earlier surmise also based on experiments with T-even phage. It is believed that this result is clear evidence to demonstrate the existence of messenger RNA. Not all the synthesizing apparatus for phage proteins is purely bacterial; several tRNAs different from the bacterial ones are coded by phage T4.

2.5 Assembly of Virions

In contrast to the early phases of phage development, the assembly of capsids and complete virions is not programmed in time by successive expression of phage genes.
All virion proteins and other late proteins such as phage lysozyme appear to be synthesized more or less simultaneously and accumulate in “precursor pools.” From the pool they are then withdrawn by direct specific interactions with other protein molecules to construct subassemblies that are then assembled into complete virions. Essentially, virion assembly consists of five major processes interacting with each other only at some critical points (Fig. 4):

1. The baseplate of the phage tail consists of 15 gene products and its synthesis involves also several other genes. The baseplate appears to contain some molecules of two phage-coded enzymes; dihydrofolate reductase and thymidylate synthetase.
2. The complete baseplate, specifically after the addition of gp54, provides a primer for the assembly of the tail tube. Around the tube the sheath assembles as a polymer 114 copies of gp18. The products of two other genes stabilize this assembly.
3. The shell of the phage head is the product of many genes and consists of more than ten proteins, one of them, gp23, representing the bulk of it. In complete heads gp23 is present in a partially cleaved form due to removal of a 10,000 molecular weight piece from the original polypeptide. Protein gp22 is an internal protein that digests itself to small peptides.
4. Once the tail and head of the phage are assembled separately they combine spontaneously in vitro as well as in vivo.
5. The tail fibers are made up of four gene products. Their assembly is independent of the rest of the virion, but they attach to the baseplate only after head and tail have been joined. The gp63 plays a role in this reaction, which also involves an interaction with whiskers attached to the collar between head and tail.

The specific shape of the phage head is determined by gp23 and by other proteins. The normal T4 head consists of a distorted icosahedron with an inserted extra band of subunits in the long axis with about 840 copies of gp23; the gp20 is located at vertices. This shape must reflect precise constraints resulting from protein–protein interaction.

### 2.6 Packaging the DNA

In the T-even virions the phage DNA is tightly packed. The topography of packing is unknown; there may be a central hollow. It is believed that the DNA collapses into a compact within the capsid when the gp22 protein inside the capsid is cleaved. How DNA is “sucked” into the capsid head is not clear; it may simply be brought in by a displacement reaction caused by exit of internal proteins. The gp49 is needed to fill the head normally with DNA.

### 2.7 Lysis

One of the late gene products of T-even phage is a lysozyme (gpe) that cleaves bacterial peptidoglycans. Lysozyme is made within infected cells well before the onset of lysis. Lysis takes place, however, only if the lysozyme gains access to the peptidoglycan layer.
One other gene, sp, is involved in the lysis control program. This gene may normally act to maintain the integrity of the cytoplasmic membrane, making continuing aerobic metabolism possible and preventing the rigid layer from being attacked prematurely by bacterial enzymes.

T-even phages have been a major model system in the development of modern genetics and molecular biology since 1940 and bacteriophages T2 and T4 were instrumental in the first formulation of many fundamental biological concepts. The advantages of T4 as a model system came from the fact that virus infection inhibited host gene expression, which allows investigators to differentiate between host and phage macromolecular synthesis.

Quite recently the sequence of its total genome was completed [6].

Therefore I will confine my story to T4 phage. First of all let us look at the structure of T4 phage (Fig. 5). Everyone will be impressed with the complexity and the beauty of its morphology.

Historically, Edward Kellenberger wrote an article in the same book as mentioned before entitled “Electron Microscopy of Developing Bacteriophage” and introduced the concept of morphopoiesis [7].

![Fig. 5 Structure components of the T4 particle. Features of the particle have been resolved to about 3 nm. The positions of several head, tail, baseplate, and tail fiber proteins are indicated [6]](image-url)
He has defined morphopoiesis as the process leading from a pool of subunits to a morphologically characterized biological entity (like an organelle or a virus). Within one system, only a determined number of different types of subunits participate. Substances contributing specifically to the specific morphology are called “morphopoietic factors or principle.” They can – but need not – be a part of the final product.

A morphopoiesis is of the first order if only one type of subunit is used to build a structure by self-assembly.

In morphopoiesis of higher order, supplementary bits of morphopoietic information are contributed through morphological factors. Morphopoietic regulation appears to act at the level of the gene products through the phenomenon of sequential triggering.

About 40 years later, since the Phage group started, the genome of the host cell, *E. coli* was also determined [8] and the morphopoiesis of phage can be explained through the interaction of viral genes and host genes.

T4 phage has about 300 genes packed into 168,903 bp genome and a total of 289 probable protein-encoding genes, 8 tRNA genes, and at least 2 other genes that encode small, stable RNAs of unknown function [6]. T4 phage builds the most complex virus particles. It devotes more than 40% of its genetic information to the synthesis and assembly of the prolate icosahedral heads, tail with contracted sheaths, and six tail fibers that contribute to its high efficiency of infection.

Extensive genetic and biochemical analysis revealed the complex assembly pathway of the T4 particle [9–11]. Twenty-four proteins are involved in head morphogenesis, and there are 22 for tail morphogenesis, and seven for tail fibers, including 1 for tail fiber attachment. Five of the 54 proteins are catalysts for assembly rather than components of the final virion. In the assembly pathway, the head, tail and tail fibers are assembled independently. A head and tail are associated, and then the six tail fibers attach to the baseplate [12].

Of the 24 proteins assigned to head morphogenesis, 16 are involved in prohead formation and maturation, 5 are responsible for DNA packaging, and 3 complete and stabilize the head. Only 10 of the 16 genes for prohead formation are essential; these include GroEL, the one host-coded protein involved in head assembly. This contrasts with phage lambda, where GroES, DnaK, DnaJ, GrpE, and GroEL are all essential [13–15]. It was shown that T4 gp31 can substitute for the function of GroES [16]. T4 gp31 has an extra loop that makes the inner cavity of the GroEL–gp31 complex larger so that it can accommodate for the folding intermediate of gp23, the major capsid protein [17].

The head is assembled on the initiator complex, which is a 12-mer of gp20 arranged in a ring. The scaffold, made of gp22–gp21 and the capsid protein, gp23, are assembled onto the initiator, which eventually becomes the portal vertex.

Pentamers of gp24, which is 28% identical in amino acids residues to gp23, are placed at the other 11 vertices. After the scaffold is completely surrounded by gp23 and gp24, the T4 prohead protease, gp21, degrades the scaffold and cleaves most of the other head proteins, including gp23 and gp24.

This creates space in the cavity of prohead. The prohead thus formed is then detached from the membrane. It was demonstrated by pulse-chase experiment that ESP (empty small particle) then initiates DNA packing and forms the ISP (initiated
small particle), which contains about 10 kb of DNA. The resulting mature head is much more resistant to any kind of perturbation [18, 19].

The tail and tail fibers are responsible for the high efficiency of T4 infection. The tail is made of a baseplate and two cocylinders. The inner cylinder consists of 144 subunits of gp19 arranged in 24 stacked hexameric rings. The inner space of the tail tube allows for the passage of phage DNA. The same number of gp18 molecules form the outer tail sheath, with the subunits arranged in the same manner as gp19.

Each stacked sheath ring is offset 17° to the right of the one below it, which gives an apparent right-handed helix [20]. While the noncontracted tail sheath is 98.4 nm long, the contracted tail sheath is only 36 nm long and the twist angle of sheath proteins is increased to 32°.

The baseplate consists of a hub surrounded by six wedges, which are assembled independently. Hub assembly is very complex. The six products of genes 5, 27, 29, 26, 28 and 51 have been reported to be involved in the assembly. gp51 is a catalytic protein, and gp26 and gp28 have not been proven to be components of the hub or baseplate. gp5 and gp27 associate first. The hub is completely by binding of gp29 to the gp5–gp27 complex. It appears that some structural modification of gp29 is necessary before associating with the gp5–gp27 complex. Wedge assembly is initiated by the association of gp10 and gp11, followed by addition of gp7, gp8, gp6, gp53 and gp25, in that order. In the absence of any of the other components, the assembly stops at that point and the remaining components are left free in “assembly-naïve states” [21].

Baseplate morphogenesis appears to occur in association with the cell membrane. The baseplates remain attached to the membrane by 300-Å fibers from the six corners of the baseplate during the remainder of phage assembly until the time of cell lysis, as shown by electron microscopy [22]. The finding that gp7 has a predicted membrane-spanning domain near its C-terminus suggests a possible mechanism for this attachment. The gp5–gp27 heterohexameric complex is attached at the tip of the tube. When the tail sheath contracts and the tail tube protrudes from the bottom of the baseplate, the triple-stranded β-helix is considered to play a role like that of a needle to puncture the cell. The short tail fiber is a trimer of gp12. It consists of three domains called the pin (N terminus), shaft, and head (C terminus).

The shaft is mainly β-helix and β-spiral. gp11 is located at the tip of the tail pin and bound to the middle part of the P12 trimer, at a site where the P12 shaft is bent about 94°. gp9 forms the socket of a long tail fiber [23] consisting of four gene products, gp34, gp35, gp36, and gp37, where gp34 and gp37 are the proximal and distal long tail fibers, respectively. gp35 and gp36 attach to the distal fiber, forming the junction between the half-fibers.

Structural analysis of the C-terminal portion of the whisker [24] revealed a three-stranded coiled-coil structure with a beta structure “propeller” at the C terminus. This beta structure is thought to bind the bend or “knee” in long tail fibers to facilitate tail attachment to the baseplate. The assembly of the tail fibers requires two molecular chaperone-like proteins, gp57A and gp38.

About 100–150 phage particles have accumulated in the cell by the lysis that occurs. Two proteins are involved in the lysis process: gpe and gpt. gpe is the so-called T4
lysozyme [25]. The gp5 lysozyme can substitute for gpe [26]. gpt is the T4 polin, which creates a hole in the inner membrane so that lysozyme can reach the peptidoglycan layer form the cytosol: the timing of polin assembly determines the time of lysis [27, 28]. In the absence of either lysozyme or the polin, lysis does not occur.

The T-even phages display a unique phenomenon, lysis inhibition, which allows them to sense when there are numerous phages around and respond appropriately to delay lysis, maximize the use of their host, and potentially await the accumulation of additional host [29]. Lysis is delayed if more phages attack the infected cell at any time after 5 min into infection. This signal is somehow mediated by the rI protein, which regulates assembly of the t holing [30].

After we have got the whole DNA sequence of T4, we can list the proteins whose functions are not clear at this moment. It is conceivable that the proteins which are related to the membrane seem to be involved somehow in the process of T4 morphogenesis. We are still far from completely understanding T4 morphogenesis.

3 Molecular Biology of Morphogenesis of Bacteriophage Lambda

Bacteriophage lambda was an important tool in the study of E. coli. And lambda is the major prototype for a template phage, which has two types of life cycle, lytic cycle and lysogenic cycle, and has played a central role in studies defining current concepts of gene regulation and genetic regulatory circuitry. The lysogenic life cycle, in which its DNA is inserted at a specific site in the host chromosome such that phage DNA passably replicates with the host DNA. A bacteriophage that can follow either a lytic or a lysogenic life style is known as a temperate phage, whereas those that have only a lytic mode are said to be virulent.

More recently, lambda strains were used to refine concepts of how cloning and expression vectors should work, and lambda vectors are still commonly used in the construction of genomic libraries. The genome size is 48,502 base pairs and the complete genome sequence is available online (accession number JO2459).

It has a linear genetic and physical map, sometimes presented in circular representation since the molecule circularizes at the cohesive ends during some stages of virus activity. In general, dsDNA phages are assembled by putting together separately tails and empty head shells, packing DNA in the head shells, and attaching tails to the filled heads. The structure of the lambda phage is shown in Fig. 6.

**Fig. 6** A sketch of bacteriophage λ indicating the locations of its components. The letters refer to specific proteins. The bar represents 50 nm [31]
The form is simpler compared with that of T4 phage. The size of DNA is about one third of T4 phage DNA. It has a 55-nm-diameter icosahedral head and a flexible 15- to 135-nm long tail that bears a single thin fiber at its end. Bacteriophage lambda (λ) adsorbs to *E. coli* through a specific interaction between the viral tail fiber and maltoporin which is coded by *E. coli* gene. This is a component of bacterium’s outer membrane.

The mature λ phage head contains two major proteins: gpE, which forms its polyhedral shell, and gpD, which “decorates” its surface. The λ head contains four major proteins, gpB, gpC, gpFII, and gpW, which form a cylindrical structure that attaches the tail to the head. The tail is a tubular entity that consists of 32 stacked hexagonal rings of gpV for a total 192 subunits. Phage assembly is studied through a procedure developed by Robert Edgar and William Wood that combines genetics, biochemistry, and electron microscopy. Conditionally lethal mutations are generated that, under non-permissive conditions, block phage assembly at various stages.

The assembly of bacteriophage λ occurs through a branched pathway in which the phage heads and tails are formed separately and then joined to yield mature virions.

Phage head assembly occurs in four stages [31]:

1. gpB and gpNu3, with two host-supplied chaperonin proteins, GroEL and GroES, interact to form an “initiator” that consists of 12 copies of gpB arranged in a ring with a central orifice. This precursor of the mature phage head-tail connector apparently organizes the phage head’s subsequent formation. GroEL and GroES, provide a protected environment that facilitates the proper folding and assembly of proteins and protein complexes such as the connector precursor. gpNu3 functions as a molecular chaperone in that it has but a transient role in phage head assembly.

2. gpE and gpNu3 associate to form a structure called an immature prohead. If gpB, GroEL, or GroES is defective or absent, some gpE assembles into spiral or tubular structures, which indicates that the missing proteins guide the formation of a proper shell.

3. In the formation of mature prohead, the N-terminal 22-residue segment of ~75% of the gpB is excised to form gpB*; the gpNu3 is degraded and lost from the structure; and ten copies of gpC participate in a fusion-cleavage reaction with ten additional copies of gpE to yield the hybrid proteins pX1 and pX2, which form the collar that apparently holds the connector in place. The enzymes that catalyzes this process has not been identified.

4. In the final stage of phage head assembly, gpW and gpFII add in that order to stabilize the head and form the tail-binding site.

Tail assembly occurs independently of head assembly from the 200-Å-long tail fiber toward the head-binding end. This strictly ordered series of reactions can be considered to have three stages:

1. The formation of the “initiator,” which ultimately becomes the adsorption organelle, requires the sequential actions on gpJ (the tail fiber protein) of the products of
phage genes I, L, K, H, and M. Of these, only gpI and gpK are not components of the mature tail.

2. The initiator forms the nucleus for the polymerization of gpV, the major tail protein, to form a stack of 32 hexameric ring. The length of this stack is thought to be regulated by gpH, which becomes extended along the length of the growing tail and somehow limits its growth.

3. In the termination and maturation stage of tail assembly, gpU attaches to the growing tail, thereby preventing its further elongation. The resultant immature tail has the same shape as the mature tail and can attach to the head. In order to form an infectious phage particle, the immature tail must be activated by the action of gpZ before joining the head.

The completed tail spontaneously attaches to a mature phage head to form an infectious λ phage particle.

4 Morphogenesis of Filamentous Phage

The filamentous bacteriophage are a group of viruses that contain a circular single-stranded DNA genome encased in a long capsid cylinder. Many use some type of bacterial pilus to facilitate the infection process. Ff class of the filamentous phages (f1, fd, and M13) have been the most extensively studied. These bacteriophages use the tips of the F conjugative pilus as a receptor and thus are specific for E. coli containing the F plasmid. The DNA sequences of these three phages shows them to be 98% homologous; consequently, the protein sequences of the gene products are practically the same. Phage M13 is the model filamentous bacteriophage [32]. The phage has been of extensive use as a cloning vector and DNA-sequencing vehicle in genetic engineering. The virion of phage M13 is only 6 nm in diameter but is 860 nm long and is semi-flexible (worm-like). These filamentous DNA phage have the additional interesting property of being released from the cell without lysing the host cell. Thus, a cell infected with phage M13 can continue to grow, all the while releasing virions. Virus infection causes a slowing of cell growth, but otherwise a cell is able to coexist with its virus. Typical plaques are not observed; only areas of reduced turbidity occur within a bacterial lawn. Many aspects of DNA replication in filamentous phages are similar. The property of release without cell killing occurs by budding. With phage M13 there is no accumulation of intracellular virions as with typical bacteriophages. Instead, the assembly of mature M13 virions occur on the inner surface of the cytoplasmic membrane and virus assembly is coupled with the budding process.

Several features of phage M13 make it useful as a cloning and DNA sequencing vehicle. First, it has single-stranded DNA, which means that sequencing can easily be carried out by the Sanger dideoxynucleotide method. Second, a double stranded form of genomic DNA essential for cloning purposes is produced naturally when the phage produces the replicative form. Third, as long as infected cells are kept in the growing state, they can be maintained indefinitely with cloned DNA, so a continuous
source of the cloned DNA is available. And there is an intergenic space in the genome of phage M13 that does not encode proteins and can be replaced by variable amounts of foreign DNA. So phage M13 is an important part of the biotechnologist’s toolbox.

The mass of the particle is approximately 16.3 MD, of which 87% is contributed by protein. The genome is a single-stranded, covalently close DNA molecule of about 6,400 nucleotides encased in a somewhat flexible protein cylinder. The length of cylinder consists of approximately 2,700 molecules of the 50-amino-acid major coat protein, also called gene VIII protein (pVIII). At one end of the particle, there are about 5 molecules each of the 33-residue gene VII (pVII) and the 32-residue gene IX protein (pIX). The other end contains approximately 5 molecules each of the 406-residue gene III and 112-residue gene VI proteins (pIII and pVI). The DNA is oriented within the virion such that a 78-nucleotide hairpin region called the packaging signal (PS) is always located at the end of the particle containing the pVII and pIX proteins.

Filamentous phage assembly is a non-lytic, membrane-associated process. This is quite different from that of T4 or lambda phage. Phage particles are extruded from the infected host, which continues to grow and divide, albeit at a reduced rate. Prior to assembly, newly synthesized coat proteins are anchored in the inner membrane (IM) by single transmembrane domains. Filamentous phage proteins pI and pIV are morphogenic proteins required for phage assembly not part of virion. Neither pI nor pIV from phage f1 can substitute for its equivalent in other related phages. When the two proteins are supplied as pairs, partial restoration of heterologous phage assembly occurs. This observation suggests that the two proteins interact. A selection for revertants of a temperature sensitive mutants of f1 gene IV resulted in the isolation of a suppressor mutation in gene I. Another similar evidences suggest that it is pI that interacts with both pIV and pVIII. Thus the process by which filamentous phage are concomitantly assembled and secreted across the cell membranes is likely to involve a series of protein–protein interactions that are accessible to genetic analysis [33].

It has been proposed that f1 is secreted through the outer membrane (OM) via a phage-encoded channel protein, pIV. A functional pIV mutant was isolated that allowed E. coli to grow on large maltodextrins and rendered E. coli sensitive to large hydrophilic antibiotics that normally do not penetrate the OM. In planar lipid bilayers, both mutant and wild-type pV formed highly conductive channels with similar permeability characteristics but different gating properties: the probability of the wt channel being open much less than that of the mutant channel. The high conductivity of the pIV channel suggests a large-diameter pore, thus implicating pIV as the OM phage-conducting channel [34].

Thus it is concluded that the phage genome is encoded for proteins that are not found in the phage particle but are necessary for phage assembly and export. Multimers of protein IV in the outer membrane (OM) interact with multimers of protein-I and protein-II in the IM to form an assembly site that may function as a gate pore through which assembling phage are extruded.

The M13 filamentous bacteriophage coat protein is a symmetric assembly of several thousand α-helical major coat proteins (pVIII) that surround the DNA core.
pVIII molecules initially reside in the host membrane and subsequently transit into their role as coat proteins during the phage assembly process. A comprehensive mutational analysis of the 50-residue pVIII sequence revealed that only a small subset of the side-chains were necessary for efficient incorporation into a wild-type (wt) coat. In the three-dimensional structure of pVIII, these side-chains cluster into three functional epitopes: a hydrophobic epitope located near the N terminus and two epitopes (one hydrophobic and the other basic) located near the C terminus on the opposite faces of the helix. These interactions could facilitate the transition of pVIII from the membrane into the assembling phages, and the incorporation of a single pVIII would be completed by the docking of additional pVIII molecules with the second hydrophobic epitope at the C terminus. In this chapter the authors have constructed a minimized pVIII that contained only nine non-Ala side-chains yet retained all three functional epitopes. The minimized pVIII assembled into the wt coat almost as efficiently as wt pVIII, thus defining the minimum requirements for protein incorporation into the filamentous phage coat. The complex mechanism is shown in Fig. 8 [35].
The Ff genome encodes only ten proteins, a simplicity which is achieved by encoding a variety of functions within individual proteins. For example, during the phage life cycle, the major coat protein (gene VIII) must be stably inserted into *E. coli* host inner membrane prior to interacting with phage DNA in the lipid-free virion. A key characteristic of this 50-residue coat protein is that “small” residues on one face of the effective transmembrane (TM) helical segment (gly34, Ala35 and Gly38), have been shown to participate in in vitro helix-helix dimerization and oligomerization of micelle-solubilized protein; this process may facilitate the interaction of protein chains in preparation for extrusion from the membrane during phage assembly. Once the coat protein leaves the host membrane, it encapsulates the Ff circular single-stranded DNA genome to produce an elongated virion about 100 times longer than it is wide [32]. The coat protein N-terminal segment consists of an amphipathic helix with an Ala-rich face (Ala7, Ala10, Ala18, Ala25) upstream from the residues which comprise the TM helix-helix interface [37–40]. It is possible that these “small” residues enable close helix-helix packing between successive layers of coat protein in the virion. To approach this question experimentally, it was reported that they used randomized mutagenesis techniques established for preparation and analysis of M13 viable mutants [41]. About 100 viable M13 mutants of the entire gene VIII may be used to assess the susceptibility of each position to mutation. In the resulting library, “small” residues (Ala, Gly, Ser), which constitute the non-polar
face of the N-terminal amphipathic helical segment, and a face of the hydrophobic helical segment, were found to be highly conserved. They propose a model in which coat protein packing is stabilized by the presence within each protein subunit of the two “oligomerization segments,” i.e. specific helical regions with faces rich in small residues which function to promote the close approach of helices.

5 The Genes of Host Cell (E. coli) Governing Phage Morphogenesis

Virus particles are aggregates of protein and nucleic acid, synthesized with the aid of small molecules, enzymes and ribosomes of the host cell. Many viruses depend on host proteins for the synthesis of viral nucleic acid and for the transcription of viral genes. Therefore it is surprising that all structural proteins of virus particles seem to be specified by viral genes as described above. It is reasonable to ask whether host proteins or other host structures might be required for particle assembly but not incorporated into the finished structure.

An attempt to isolate bacterial mutants which fail to support phage growth was taken in the 1970s. It was shown with certain bacterial mutants, called groE, that T4 phage head assembly was blocked specifically, implying that the host plays a direct role in head assembly. The block occurred early in the assembly process at the level of action of T4 gene 31 [42]. And it was also shown that bacterial strains of E. coli which have groE mutants were defective for λ head formation [43]. Mutants of λ, designated λEP, which were able to grow in the three groE strains, had been isolated. An analysis of these mutants indicated that at least some carried a mutation in λ head gene E and these made reduced levels of active gene E protein in groE hosts. The results indicated that it was the interaction between the gene E protein and the proteins specified by genes B and C that was adversely affected by the groE mutation. It is conceivable that the relative level of gene E protein is too high in groE strains for proper head formation. The λEP mutation compensates for this effect by reducing the level of this protein, and so restoring a balance.

It was demonstrated that during normal head assembly the protein encoded by phage head gene B or C appeared to be converted to a lower molecular weight form, h3, which was found in phage. The appearance of h3 protein in fast sedimenting head related structures required the host groE function. It was then suggested that the proteins encoded by phage gene E, B and C, and the bacterial component defined by groE mutations, acted together at an early stage in head assembly [44].

Many groE mutations exerted pleiotropic effects, such as the inability to propagate phage T4 and T5 and inability to form colonies at 43 °C. The groE bacterial gene product as a protein of 65,000 molecular weight was identified [45]. Lately one groE gene, groEL, has been shown to encode the synthesis of a 65,000 molecular weight polypeptide, whereas the second, groES, codes for the synthesis of a 15,000 M polypeptide [46].
6 Phage Display

Phage display began with the paper by George Smith in 1985 and it became a core biotechnology [47]. This technology is based on two concepts. The first is that phage can be used to link protein recognition and DNA replication. The protein is displayed on the surface of the phage particle and the genes encoding it are contained within the particle. The second is that large libraries of the DNA sequences encoding these molecules can be cloned into phage.

6.1 The Phage Genome

The genomes of the Ff phage (M13, f1 and fd) have been completely sequenced [48–50]. Each genome encodes 11 genes, whose products are listed in Table 1.

Two of the gene products, pX and pXI, are the result of a translational start at the internal methionine codon in genes II and I, respectively. These internal methionine codons are in-frame, so the smaller proteins have the same sequence as the carboxy-terminal portions of their larger counterparts. The genes are grouped in the genome according to their function in the life cycle of phage. One group (gene II, V, X) encodes the proteins required for the replication of the phage genome. Other encodes capsid proteins (pI, pX, and PIV) are involved in the membrane-associated assembly of the bacteriophage. There is a short sequence called the intergenic region that does not code for protein. It contains the sites of origin for synthesis of viral (+) and complementary (−) DNA. The PS is in the intergenic region near the end of gene IV.

There are very few regions in the phage genome that do not code for protein. Cassettest encoding antibiotic resistance are generally inserted in the intergenic region or in the space between the end of gene VIII and the beginning of gene III. In the latter case, some alterations of the positions of the terminator and promoter in this region must be made, and care must be taken not to interfere with the origins of replication or other control areas. There appears to be a delicate balance in the

<table>
<thead>
<tr>
<th>Gene</th>
<th>Function</th>
<th>Protein MW</th>
</tr>
</thead>
<tbody>
<tr>
<td>II</td>
<td>DNA replication</td>
<td>46,137</td>
</tr>
<tr>
<td>X</td>
<td>DNA replication</td>
<td>12,672</td>
</tr>
<tr>
<td>V</td>
<td>Binding ss DNA</td>
<td>9,682</td>
</tr>
<tr>
<td>VIII</td>
<td>Major capsid protein</td>
<td>5,235</td>
</tr>
<tr>
<td>III</td>
<td>Minor capsid protein</td>
<td>42,522</td>
</tr>
<tr>
<td>VI</td>
<td>Minor capsid protein</td>
<td>12,342</td>
</tr>
<tr>
<td>VII</td>
<td>Minor capsid protein</td>
<td>3,599</td>
</tr>
<tr>
<td>IX</td>
<td>Minor capsid protein</td>
<td>3,650</td>
</tr>
<tr>
<td>I</td>
<td>Assembly</td>
<td>39,502</td>
</tr>
<tr>
<td>IV</td>
<td>Assembly</td>
<td>43,476</td>
</tr>
<tr>
<td>XI</td>
<td>Assembly</td>
<td>12,424</td>
</tr>
</tbody>
</table>
synthesis of the phage proteins that allows phage production without seriously affecting bacterial cell growth.

Needless to say, the technology which is developed using filamentous phage can be applicable to other phage such as T4 or lambda phage as mentioned earlier. However, I want to discuss here only filamentous phage.

### 6.2 Display of Peptides and Proteins on Phage Particles

Large libraries of peptides and proteins have been made using pIII as the display vehicle [51, 52], leading to the development of a number of techniques for selecting the molecules desired from such libraries [53, 54]. Peptides and proteins have been fused to the amino-terminal portion of the major capsid protein pVIII [55]. There has been a report of phage particles displaying proteins that are fused to the carboxy-terminal portion of pVI, although the efficiency of display appears to be lower in this case [56]. Recently, antibody heavy- and light-chain variable regions have been fused to the amino terminus of pVII and pIX and displayed on phage, showing that these two minor coat proteins can also be used for display [57].

Proteins can be displayed on smaller filamentous particles called phagemids [58, 59]. The phagemid genome contains the filamentous phage intergenic region with its origin of replication for viral and complementary strand synthesis as well as the hairpin packing signal. The genome contains a plasmid origin of replication and a gene coding resistance to a specific antibiotic. Chimeric genes encoding peptide–phage protein fusion can be placed under control of a specific promoter in these phagemid genomes. The phagemid can maintain itself as a plasmid, directing the expression of the protein in bacteria. Infection of bacteria with a filamentous helper phage activates the phage origin of replication, resulting in single-stranded phagemid DNA being encapsulated into filamentous phage-like particles using helper phage proteins. A helper phage containing a defective packaging signal can be used so that the majority of particles produced contain the phagemid single-stranded DNA [60]. Bacteria can be infected with phagemid-phage mixture and colonies selected that are resistant to the antibiotic. The resistant colonies will contain only the phagemid DNA, which can be propagated again by infection with helper phage. Because the phagemid particles can transmit antibiotic resistance, they are referred to as “transducing particles.”

To access technical details readers should consult Phage Display [61].

### 6.3 Application of Phage Display in Biotechnology

Phage display, largely developed in the 1990s, has begun to make critical contribution to major endeavors of biological scientists. In 1985, George Smith first showed that the linkage between phenotype and genotype could be established in filamentous bacteriophage and gave birth to the new technology of phage display.
First of all I want to show you G. Smith’s work and subsequent works using other phage.

6.4 Filamentous Phage

Gene III of filamentous phage encodes a minor coat protein, pIII, located at one end of the virion. The amino terminal half of pIII binds to the F pilus during infection, while the carboxy-terminal half is buried in the virion and participates in morphogenesis. It was reported that foreign DNA fragments could be inserted into filamentous phage gene III to create a fusion protein with the foreign sequence in the middle. The fusion protein was incorporated into the virion, which retained infectivity and displayed the foreign amino acid in immunological accessible form. These “fusion phages” can be enriched more than 1000-fold over ordinary phages by affinity for antibody directed against the foreign sequence. Fusion phages may provide a simple way of cloning a gene when an antibody against the product of that gene is available [47].

6.5 Phage Lambda

The tolerance of bacteriophage lambda morphogenesis for C-terminal additions to the tail tube major protein subunits (the V gene product; gpV) was shown to be useful for the application of phage display for the biotechnology (Fig. 8). A second modified copy of the lambda V gene, either within a novel phage vector itself or plasmid-borne, was expressed during phage growth. High-level substitution of wild-type gpV by modified gpV bearing a basic C-terminal peptide sequence (RRASV; a target site for cAMP-dependent protein kinase) was possible using multiple repeats of a serine-glycine (SGGG) linker sequence. Highly purified phage bearing copies of gpV-RRASV could be efficiently phosphorylated by the appropriate protein kinase, and the incorporated label was shown to migrate exclusively at the expected size in protein gels. A large tetrameric protein (β-galactosidase) could be incorporated into active virions in at least one copy, again using a Ser-Gly linker. This result suggested that with a suitable spacing linker and controlled levels of expression, it is likely that a wide range of protein or peptide substitute can be fused with gpV at its terminus and assembled as component subunits of the tail tube [62].

6.6 Phage T7

The ideology of “rational drug design” embraces the notion that (1) drugs should be targeted against specific proteins known to malfunctioning within cells, (2) the candidacy of these proteins as attractive targets for therapeutic intervention should
be further determined by their predicted drug ability, and (3) the detailed molecular structures of such target proteins should inform the design of the chemical structures of the drugs that are to be developed.

In the case of tyrosine kinase inhibitors, attempts at identifying all of the kinases that might be affected by a drug have involved assays of only a small proportion of the large cohort of protein kinases known to be present in human cells. Consequently, certain off-target effects are likely to have eluded drug developers. This has begun to change with the advent of more systematic screening of a far larger portion of the kinases that might be affected by these inhibitors. Discovery of off-target activities of a drug is actually useful in two ways; (1) it may explain toxicities of a drug and (2) it may reveal new clinical application for the drug.

A promising example of application of phage display using T7 phage for this purpose was reported [63]. Kinase inhibitors are important cancer drugs, demonstrating powerful clinical activity in tumors in which the target kinase is activated by mutation. The success of imatinib (Gleevec, Novartis Pharma AG), an inhibitor of the mutant kinase, Bcr-Abl, in treating chronic myelogenous leukemia, has stimulated many scientists’ interest in finding new kinase inhibitors. The new in vitro assay for determining kinase inhibitor specificity and using it to create interaction map of 20 different kinase inhibitors against 119 different protein kinases was reported. The majority of kinase inhibitors target the kinase ATP site, and because all of the more than 500 protein kinases identified in human genome have an ATP site, there is great potential for cross-reactivity. Compounds must be tested experimentally against many kinases to assess molecular specificity and to identify off-target interactions. Binding specificity and affinity are not readily predicted based on available sequence or structural information, and conventional profiling methods based on in vitro activity are limited by the difficulty of building and running large numbers of kinase activity assays. A quantitative assay is shown in Fig. 9.

Kinase domains are expressed as fusions to T7 bacteriophage, and test drugs in solution compete with tethered “bait” compounds for binding to the ATP-binding site of each kinase. The amount of phage bound to the tethered ligand is then quantified to determine the affinity of the drug. In addition to simply expanding the kinase profile of some of the 20 drugs tested, the authors reported a potpourri of unanticipated interactions. These include inhibition of an imatinib-resistant mutant of ABL by the p38 inhibitor BIRB-766 and inhibition of SRC family kinase LKC by imatinib. The key assay compounds are human kinases expressed as fusions to T7 bacteriophage and a small set of immobilized probe ligands that bind to the ATP site of one or more kinases. The results are read out by quantifying the amount of fusion protein bound to the solid support, which is accomplished with extraordinary sensitivity by either traditional phage plaque assay or by quantitative PCR using the phage DNA as a template.

T7 phage replication leads to lysis of the bacterial host, and lysates containing properly folded, tagged kinase are used directly in the assay with no need for conventional protein purification.
Fig. 9a–c Competition binding assay for measuring the interaction between unlinked, unmodified (free) small molecules and kinases. a Schematic overview of the assay. The phage-tagged kinase is shown in blue, “free” test compound in green and immobilized “bait” ligand in red. b Binding assay for p38MAP kinase. The immobilized ligand was biotinated SB202190. The final concentration of the test compounds during the binding reaction was 10 μM. c Determination of quantitative binding constants. Binding of tagged p38 (to immobilized SB202190 was measured as a function of unlinked test compound concentration. Tagged p38 kinase was quantified by real-time quantitative PCR and results normalized.

References

43. Sternberg N (1973) J Mol Biol 76:25
44. Georgopoulos CP, Hendrix RW, Casjens SR, Kaiser AD (1973) J Mol Biol 76:45
47. Smith GP (1985) Science 228:1315
Nano/Micro Biotechnology
Endo, I.; Nagamune, T. (Eds.)
2010, XVI, 272 p. 129 illus., 28 illus. in color., Hardcover
ISBN: 978-3-642-14946-7