

T4 Tail Morphogenesis

PETER B. BERGET¹ AND JONATHAN KING²

Department of Biochemistry and Molecular Biology, The University of Texas Medical School, and Graduate School of Biomedical Sciences, Houston, Texas 77225,¹ and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139²

The most distinctive feature of the familiar T4 image is the complex contractile tail apparatus, responsible for delivering the phage chromosome into the host cell. Composed of more than 20 different protein species, the tail has been instructive in understanding the genetic control of organelle assembly, the regulation of protein-protein interactions during morphogenesis, and the initiation and termination of repeated structures (Fig. 1). Important features of the assembly process include (i) regulation of the assembly process at the level of the interactions between proteins in the pathway, rather than at the level of genome expression; (ii) critical regulatory interactions in the assembly process between the growing structure and soluble subunits, rather than interactions between soluble subunits themselves, and (iii) regulation by switching subunits from a nonreactive precursor conformation into a reactive conformation upon incorporation into the growing structure (Caspar, 1980; King, 1980).

TAIL GENES

Table 1 lists the genes and gene products which are required for tail assembly. Nearly half of the genes known to be required for T4 particle formation are involved in the assembly of the tail structure. Twenty-two genes code for proteins which are incorporated into the tail structure, and two code for catalytic factors required for complete assembly. Nine of these genes map in one cluster, genes 53, 5, 6, 7, 8, 9, 10, 11, and 12. Eight of the genes map in a second cluster, 25-26-51-27-28-29-48-54, separated from the first cluster by the major head gene cluster. Genes 18 and 19, which code for the major tail sheath and tail tube subunits (King and Laemmli, 1973), map as a separate pair next to the head cluster, and two more genes, 3 and 15, are flanked by head genes. The *frd* and *td* genes map in one of the early regions. We discuss the possible significance of these gene clusters later on.

Table 1 also lists the functions or structural locations of these gene products, where known, and the "small" molecules known to be components of the T4 tail. In addition, the stoichiometries in terms of copies per phage particle and polypeptide molecular weights of these proteins are given. The variation in these stoichiometries points out that the assembly of this structure involves the interaction of diverse oligomers of gene products and differs from the assembly of ribosomes, where most structural components are found in single copies per structural unit. Tail polypeptide chains cover a large molecular weight range from 15,000 to 140,000 g/mol (King and Mykajewycz, 1973).

STRUCTURE OF THE TAIL

The complete tail is defined as the structure accumulating in cells blocked in head morphogenesis. An electron micrograph of sucrose gradient-purified T4 tails is shown in Fig. 2. These tails do not have tail fibers associated with them; they have the extended sheath covering the tail tube, and a short extension of the tail tube projects from the end of the sheath (King, 1968). A few baseplate-tail tube structures are also seen in Fig. 2. At late times in mutant-infected cells, additional proteins add to the proximal end of the complete tail, forming a neck structure normally found only in the mature phage (Coombs and Eiserling, 1977; see Eiserling, this volume).

Contractile Sheath

The sheath is composed of 24 annuli spaced 40 Å (4 nm) apart with six subunits per annulus. The spatial disposition of these subunits is known from three-dimensional reconstructions from electron micrographs (DeRosier and Klug, 1968). These subunits are bonded to each other radially within the same annulus and also between annuli. There is probably some weak interaction between sheath subunits and tube subunits since during assembly the sheath subunits do not polymerize onto the baseplate unless the tail tube has initiated. The sheath subunits are much less tightly bound to each other in the extended sheath than in the contracted sheath (To et al., 1969). The polymerized sheath subunits on partially assembled tails are in equilibrium with unassembled subunits, with the partially assembled structure being somewhat cold sensitive (King, 1968). The stabilization of the mature sheath depends on the stabilization of the last annulus when the tail is completed (King, 1971).

Sheath contraction. Considerable effort has been put into defining the mechanism of sheath contraction and the nature of the transition from the extended to the contracted state (Moody, 1973). Two points are important for purposes of understanding the assembly process: the sheath appears to be built in a high-energy state such that once triggered, contraction is spontaneous; and almost all of the sheath annuli must slide past the tube annuli during contraction (see Fig. 7 of Eiserling, this volume). However, for mechanical force to be generated the proximal sheath annuli must stay bound to the proximal end of the tube, and the distal sheath annuli must remain bound to the baseplate. The contracted sheath is very tightly bonded; it remains intact when the entire rest of the phage, including the tube and capsid, has been dissociated under denaturing conditions (To et al., 1969). The propagation of the contraction proceeds from the

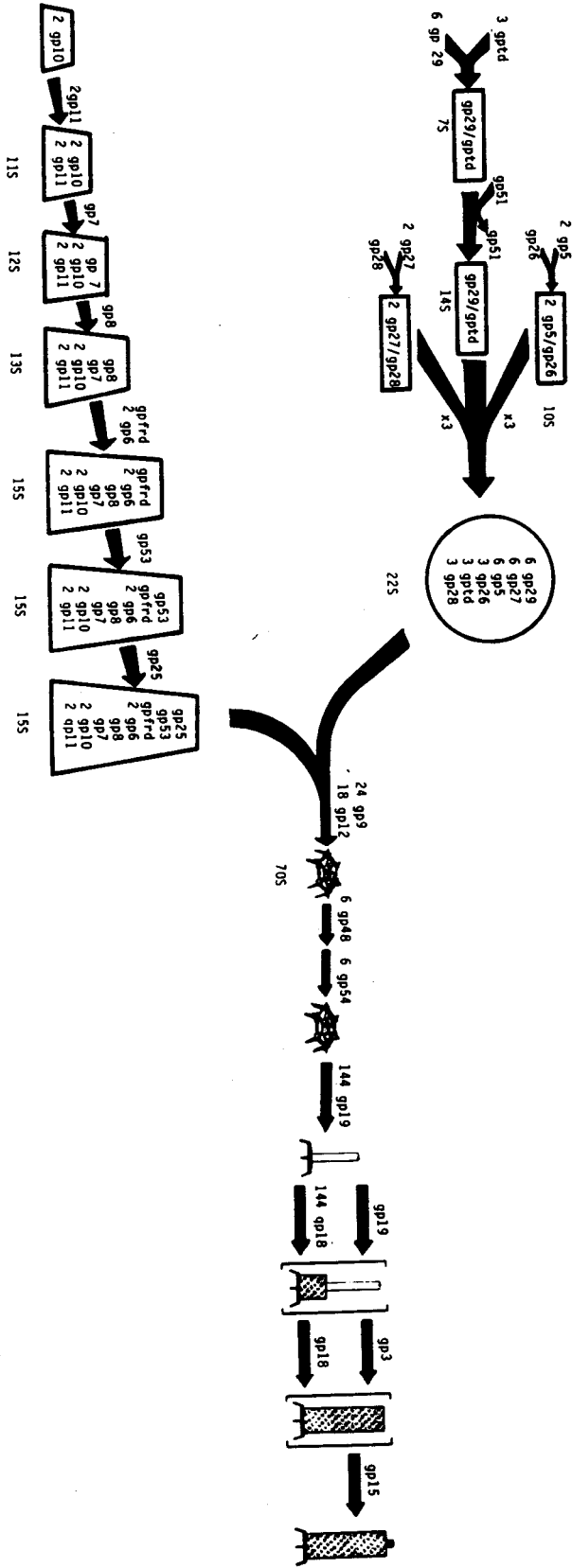


FIG. 1. Morphogenetic pathway of the T4 tail. The data used to construct this pathway come from Kikuchi and King (1975a, 1975b, 1975c), Plishker, Chidambaram, and Berget (unpublished data), and Kozloff (1981; personal communication). The abbreviation "gp" indicates gene product, the protein product of an identified T4 gene, i.e., gp10 refers to the protein product of T4 gene 10. Where known, the stoichiometries of the various protein components are indicated. Note that the exact point of addition of gp10 to the 16th arm subassembly is not known; we have shown its point of addition from the data of Mosher and Mathews (1979).

TABLE 1. Genes, gene products and other factors required for tail assembly

Gene	Gene product mol wt ($\times 10^3$)	Copies per phage	Function/location
3	29	?	Tail tube proximal tip
5	44	6	Center hub, lysozyme
6	85	12	1/6th arm
7	140	6	1/6th arm
8	46	6	1/6th arm
9	34	24	Baseplate, long tail fiber attachment
10	88	12	1/6th arm
11	24	12	1/6th arm
12	55	18	Baseplate short tail fibers
15	35	?	Proximal tail sheath stabilizer
18	80	144	Sheath monomer
19	21	144	Tail tube monomer
25	15	6	1/6th arm
26	41	3	Central hub
27	49	6	Central hub
28	24	3	Central hub, gamma-glutamyl hydrolase
29	77	6	Central hub, folyl polyglutamyl synthetase
48	44	6	Baseplate tail tube length "determiner"?
51	?	Nonstructural	Catalytic factor for central hub formation
53	23	6	1/6th arm
54	36	6	Baseplate, tail tube polymerization initiator?
57	6	Nonstructural	Catalytic factor for gp12, gp34, and gp37 maturation
<i>td</i>	29	3	Central hub, thymidylate synthetase
<i>frd</i>	20	6	1/6th arm, dihydrofolate reductase
—	1.1	6	Dihydropteridine hexa-glutamate
—	0.065	6	Zinc, in gene 12 protein

baseplate up the sheath and resembles a class of crystal dislocations known as Martensitic transitions (Olson and Hyman, 1983).

Tail Tube

Cells infected with mutants defective in the sheath protein accumulate baseplates with tail tubes (see Fig. 2). The tail tubes are 95 nm in length and are quite smooth. However, their underlying structure is annular, and the ratio of the annular spacing of the tube to the sheath is 1:1 (Moody, 1971). Tail tubes can also be isolated as free structures. These are very stable and often have their central channel penetrated by negative stain (Duda and Eiserling, 1982; To et al., 1969). Isolated tail tubes are composed almost entirely of gp19 and lack the small projection composed of gp15 present on the sheathed tail (King and Laemmli, 1973). Subunits isolated by dissociation of assembled tubes behave differently from precursor unassembled tube subunits; the former polymerize spontaneously,

whereas the latter require a baseplate template (King, 1971; Poglazov and Nicolskaya, 1969; To et al., 1969).

Baseplate

Baseplates accumulate in cells infected with mutants defective in the gene 19 tail tube subunit. These structures have the tail spikes projecting from one face, the distal face, and a small platform with a diameter of about 20 nm on the other face, the proximal face. In the assembly of the baseplate the last two proteins to add are the products of gene 48 and 54 (see Fig. 1). If the gp48 protein is removed by mutation, the baseplates dimerize between their proximal faces. These structures sit on electron microscope grids on their edges (Fig. 3), revealing more clearly the baseplate profile (King, 1971). The morphology and conformational rearrangement of the baseplate have been studied in considerable detail by rotational filtering of electron micrographs (Crowther et al., 1977; Fig. 4). The baseplate has an outer rim representing the joining of the six vertices, a central hub which strongly excludes stain, and a complex inner ring interconnecting the hub and outer rim (Fig. 5). The completed baseplates only rarely transform to stars. However, if the gene 12 protein is removed by mutation, the baseplates transform spontaneously to stars (Fig. 6).

TAIL FUNCTION

As originally described by Simon and Anderson (1967a; 1967b), the injection process involves (i) attachment of the tail fibers and baseplate to the cell surface, (ii) transformation of the baseplate from a hexagonal to a star-shaped structure, simultaneously releasing the tip of the tail tube and initiating the contraction of the sheath, and (iii) contraction of the sheath, driving the released tail tube tip through the cell envelope and delivering the phage chromosome into the host cytoplasm.

Other chapters in this volume describe the initial contact with the host cell through the distal tips of the long, slender tail fibers. The proximal ends of these tail fibers are attached to the vertices of the baseplate. Upon interaction of the long fibers with their receptors, the short fibers of the baseplate interact with a second set of host cell receptors. Since particles lacking tail fibers are not infectious, both sets of interactions are probably needed to activate the baseplate and trigger the contraction process (see Goldberg, this volume).

The hub of the baseplate presumably covers and closes the tip of the tail tube in the mature phage. However, during injection the tube penetrates the cell envelope and must open to allow DNA exit. Thus the hub can be thought of as a diaphragm which opens either during sheath contraction or after penetration of the cell envelope. It is not known which proteins of the hub enter the cell and which remain with the expanded star-shaped baseplate.

The mature form of the tail which has to be assembled is clearly a metastable structure, which must remain in the metastable uncontracted form through sewer and stomach and still maintain the ability to be efficiently triggered upon interaction with the bacterial receptors. Some of the proteins of the baseplate are

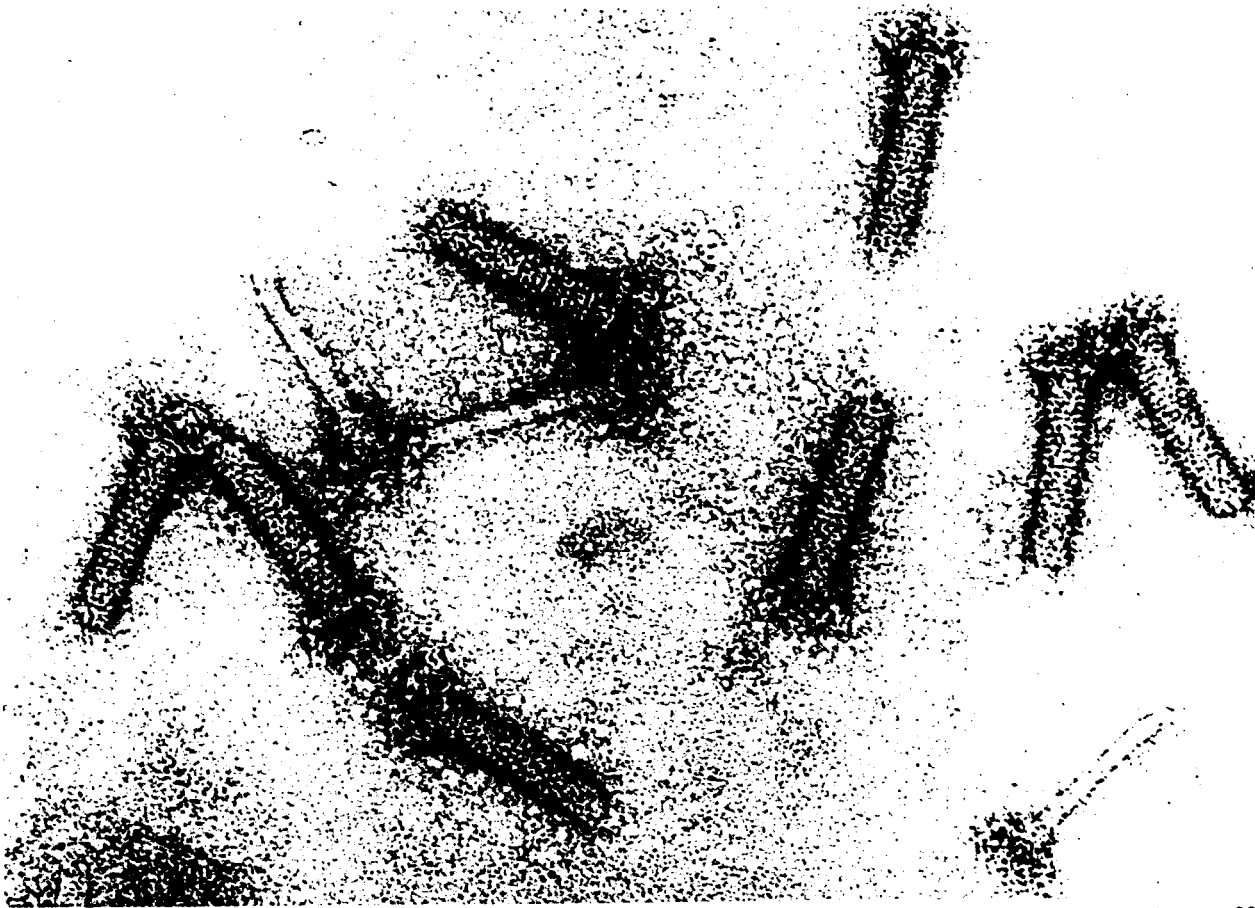


FIG. 2. Precursor T4 tails. Tails were isolated from su^- cells infected with phage carrying an amber mutation in gene 23, defective in capsid assembly. The tails were isolated by sucrose gradient centrifugation and negatively stained with uranyl acetate for electron microscopy. Some tails have lost their sheath, revealing the tail tube.

probably involved in this balance between stability and contractility (Crowther, 1980).

It is not clear whether the end of the DNA that first gets injected resides in the head or is already threaded down the tail of the mature particle. If the end of the DNA is still completely in the head in the mature phage, the injection process also requires the opening of a valve or channel between the tail and the head. Alternatively, this channel may open upon the joining of the head to the tail during assembly so that the DNA is prethreaded through the tail tube (see Goldberg, this volume).

PROTEIN LOCATION AND FUNCTION

Though we focus below on the assembly pathway for the tail, some of the tail gene products have been identified with functional activities during the infection process. Others have been shown to have unusual or unexpected properties which deserve special mention. We briefly summarize this information.

The gene 9 protein is located at the baseplate vertices and is required for attachment of the tail fiber (Crowther et al., 1977; King, 1968). It plays a role on the triggering of baseplate expansion (Crowther, 1980).

The gene 11 protein forms the tips of the six spikes

that project from the baseplate and plays a role in the attachment and injection process (Crowther et al., 1977; King, 1968).

The gene 12 protein is located on the distal face of the baseplate facing the cell and may be the morphological short fibers (Crowther et al., 1977; Kells and Haselkorn, 1974). Complete phage particles assemble in its absence. These particles can adsorb to cells but do not kill them (King, 1968). The gene 12 protein is a Zn^{2+} -containing metalloprotein (Zorzopulos and Kozloff, 1978). The maturation of the gene 12 protein requires the catalytic activity of the gene 57 protein for its maturation into its trimeric structure (Kells and Haselkorn, 1974; King and Laemmli, 1973). The requirement for gene 57 activity is shared with two of the tail fiber proteins, gp34 and gp37 (King and Laemmli, 1971).

The gene 5 protein has recently been shown to possess a "lysozyme" activity (Kao and McClain, 1980b). This protein is a structural component of the central hub of the baseplate, and it may function in the cell wall penetration of the tail tube during DNA injection.

The gene 28 protein has been shown recently to be a structural component of the T4 baseplate (Kozloff and Zorzopulos, 1981) and to possess a gamma-glutamyl carboxypeptidase enzymatic activity (Kozloff and

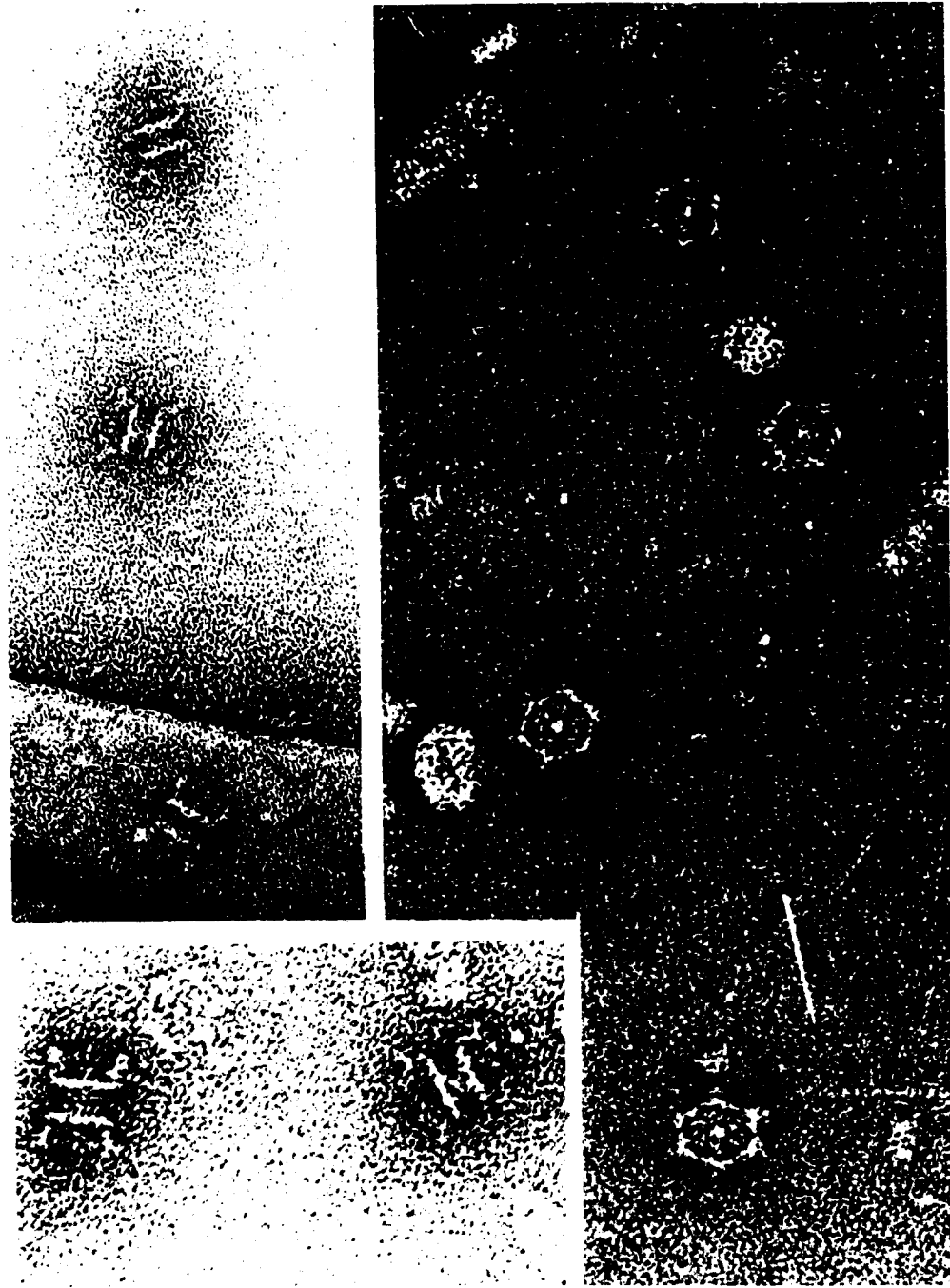


FIG. 3. Baseplates from 48^- -infected cells. Baseplates were isolated by sucrose gradient centrifugation from cells infected with phage carrying an amber mutation in gene 48. The dimers sediment faster than the monomers, at about 105S (King, 1971).

Lute, 1981). This enzymatic activity seems to play a role in the formation of the unusual folate molecules found in T4-infected cells (Kozloff and Lute, 1965; Kozloff, Verses, Luter, and Crosby, 1970; Nakamura and Kozloff, 1978; Kozloff, this volume). Furthermore, it has been suggested that this protein, in cooperation with other folate binding proteins found in the T4 baseplate, may provide the binding force required to hold the baseplate subassemblies together (Kozloff, 1981).

The gene 29 protein has recently been reported to be a folyl polyglutamyl synthetase (Kozloff, this volume) which is responsible for part of the alteration in folate metabolism after infection by T4. It is also one of the major polypeptides of the T4 baseplate central hub (Kikuchi and King, 1975b).

The products of genes *td* and *frd* are the T4-induced enzymes thymidylate synthetase and dihydrofolate reductase, respectively. Both of these gene products are found as structural components of the T4 base-

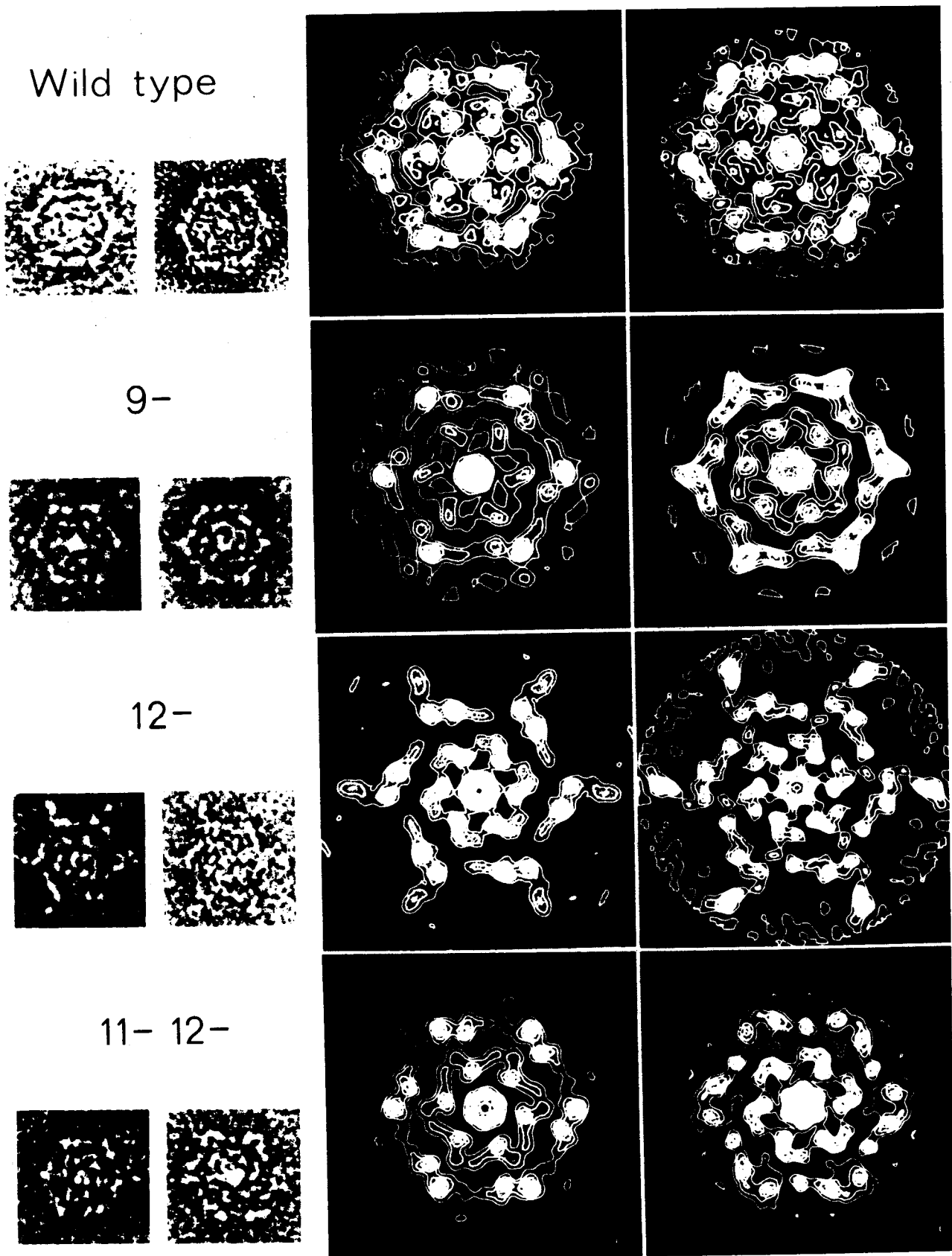


FIG. 4. Rotationally filtered baseplate images. Negatively stained electron micrograph images of various classes of baseplates were rotationally filtered by the procedures of Crowther and Amos (1971). The proteins missing from the baseplates are shown above the micrographs. The images of the 12^- baseplates suggest structures in transition from the hexagon to the star form. From Crowther et al. (1977).

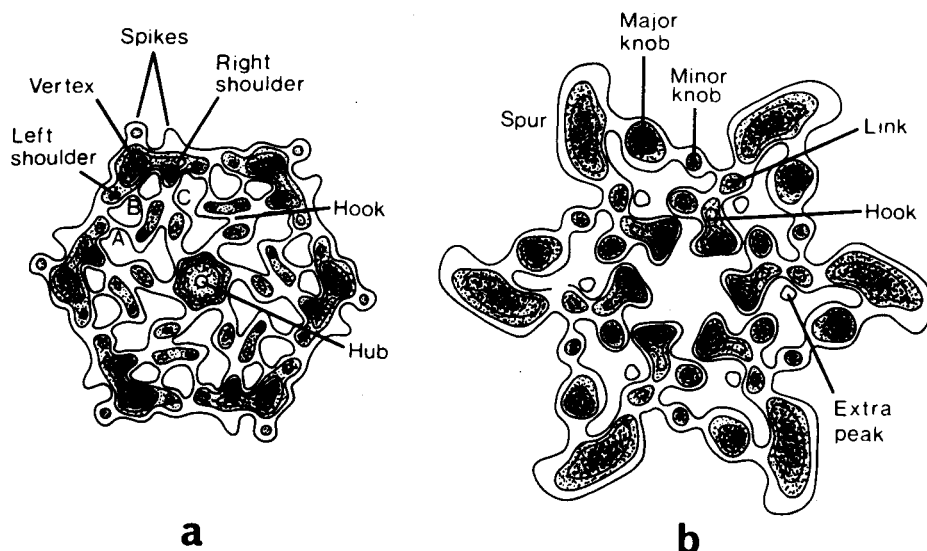


FIG. 5. Schematic diagram of baseplate hexagon and star forms. The transformation from hexagon to star form is complex and involves rearrangement of most of the mass of the baseplate. It is not known whether the parts of the hub move to the outer radius or whether some proteins are actually lost. The star drawing does not include the thin fibrils often seen projecting from the vertices of stars derived from complete baseplates. These fibrils may be the gene 12 protein, which is folded up beneath the baseplate in the hexagonal state.

plate. Recently substantial evidence has been presented that these proteins reside in the central hub and 1/6th arm subassemblies of the baseplate, respectively (Kozloff, 1981; Kozloff and Zorzopulos, 1981; Mosher and Mathews, 1979).

MORPHOGENETIC PATHWAY OF THE T4 TAIL

The morphogenesis of the tail can be broken down into two stages, the formation of the baseplate and the polymerization of the tail tube and sheath on the completed baseplate (Fig. 1). Of the 22 proteins found in the T4 tail, 18 are structural components of the baseplate (King and Mykolajewycz, 1973). The remaining four structural proteins are found in the tail tube and sheath.

Much of the work on T4 tail assembly involved *in vitro* complementation analysis (Edgar and Lielausis, 1968; Edgar and Wood, 1966). The complementation reactions in the tail pathway proceed relatively efficiently, but certain features must be kept in mind. Since 144 tube subunits and 144 sheath subunits are required to assemble one tail, the reactions are quite sensitive to the phage protein concentration in the extracts. These experiments require very concentrated extracts made from mutant phage strains also carrying a gene *t* amber mutation, delaying cell lysis, and extending phage maturation past the normal lysis time (Josselin, 1970). In addition, the sheath subunits aggregate at very high concentration into polysheath, in which form they are not available for sheath assembly. The tube subunits are also somewhat unstable in the crude extract during the incubation (King, unpublished data). Thus very concentrated infected-cell extracts are required, and these have to be kept cold and used quickly after thawing.

Baseplate Morphogenesis

The T4 baseplate (sedimentation coefficient of ca. 70S) is formed from two subassemblies. These are the central hub and the 1/6th arm. The original work reported by Kikuchi and King (1975a, 1975b, 1975c) elucidated the general outline of the assembly of these two structures from their precursor proteins. With the discovery that gp26 and gp28 are components of the baseplate (Kozloff, 1981; Kozloff and Zorzopulos, 1981) and the localization of *gpfrd* and *gptd* in the assembly scheme (Mosher and Mathews, 1979), we have added these proteins to the assembly pathway.

1/6th arm assembly. The baseplate 1/6th arm is constructed from eight structural proteins (gp6, gp7, gp8, gp10, gp11, gp25, gp53, and *gpfrd*) without the requirement of any known catalytic factors. A tracing of a sodium dodecyl sulfate (SDS) gel of the isolated 1/6th arms is shown in Fig. 7. The elucidation of the assembly pathway involved two approaches: identifying subassembly complexes by their unique sedimentation behavior in sucrose gradients, and characterizing these complexes with respect to their *in vitro* complementation activities. By incorporating amber mutations in the major head proteins, the head assembly pathway was blocked. This made it possible to identify phage-specific complexes sedimenting in the 10 to 20S region of sucrose gradients. Many of the 1/6th arm proteins are of relatively high molecular weight, such as gp7 (140,000 g/mol), gp6 (85,000 g/mol), and gp10 (88,000 g/mol), so that they could be unambiguously identified in SDS-gel analysis of gradient fractions.

In addition, Kikuchi and King (1975a) were able to obtain *in vitro* complementation between all pairs of extracts infected with mutants defective in the 1/6th arm genes. The positive complementation results es-

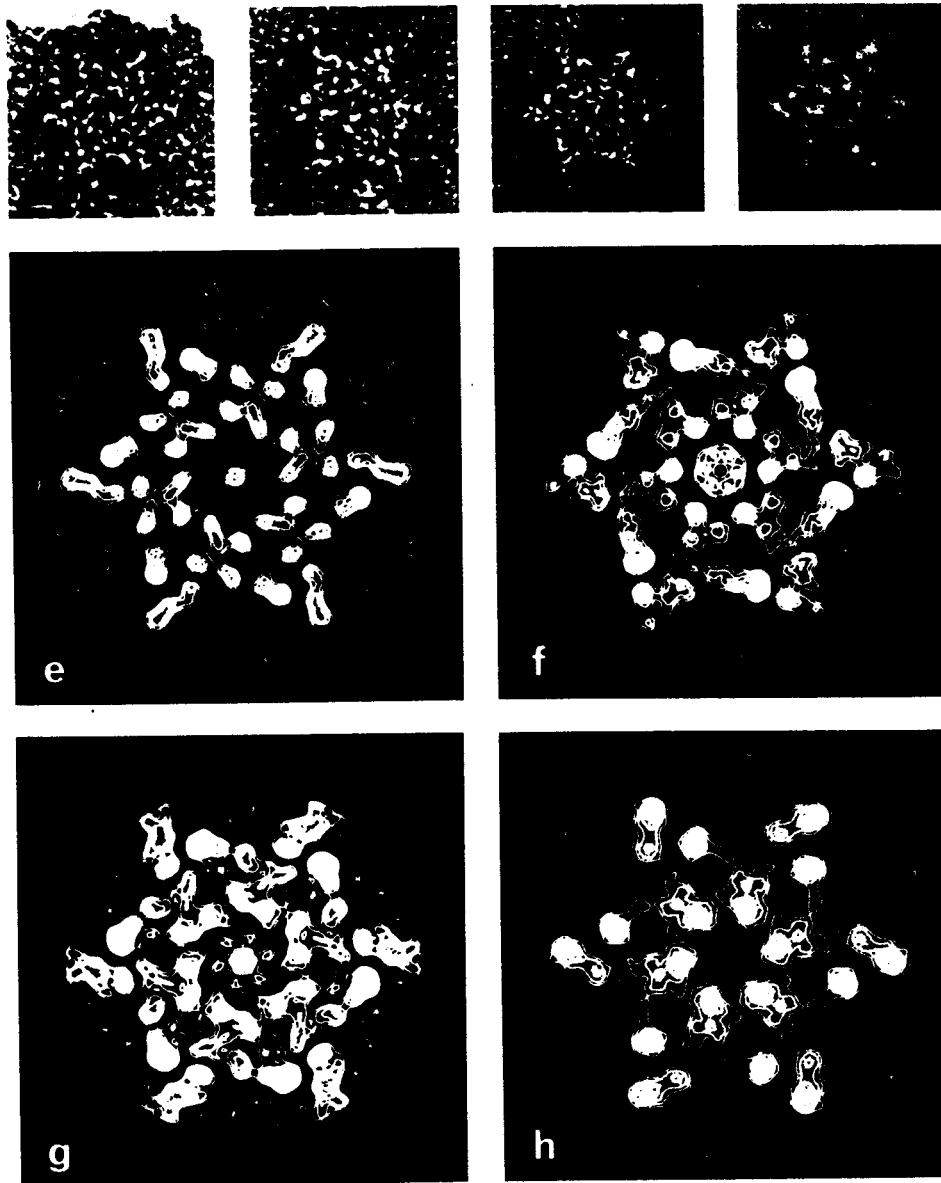


FIG. 6. Baseplate stars from 12^- -infected cells. The electron micrographs were taken of the peak fraction of baseplates isolated from 12^- -infected cells. These sedimented at 63S and were probably hexagons. The transformation to star shape presumably takes place during preparation of the electron microscope grid. These baseplates lack the morphological fibrils seen on stars from complete baseplates, which probably represent the gene 12 protein (Kells and Haselkorn, 1974). a-d, Electron microscope images; e-h, the respective rotationally filtered images. From Crowther et al. (1977).

tablished that the baseplate proteins accumulated in precursor form in the extracts. Kikuchi and King (1975a) were able to recover these activities from sucrose gradients and to characterize directly the precursor complexes accumulating in the mutant-infected cells.

Baseplate 1/6th arm structures accumulate as a 15S complex in cells infected with mutants defective in hub assembly (e.g., 5^- , 27^- , or 51^- mutants). These structures slowly aggregate into aberrant sixfold symmetrical assemblies closely resembling complete baseplates, but lacking the hub proteins and unable to serve as a substrate for baseplate completion or tail tube polymerization (Fig. 8). These structures are further discussed below. The next to last step in the

formation of this arm complex is the addition of gp53. If this protein is removed by mutation, a 15S complex accumulates, containing all of the 1/6th arm proteins except gp53 and gp25 (Fig. 7). These do not polymerize, indicating that gp53 plays a critical role in the radial interactions.

The 1/6th arm proteins assemble in a strictly sequential fashion, as shown in Fig. 1. The assembly of each structural protein, except gp11, into the growing structure is a prerequisite for the next protein in the pathway. Thus each protein seems to modify the growing 1/6th arm structure so that it is then a substrate for the next protein addition reaction. In this pathway the first two proteins to interact are the products of genes 10 and 11. The gene 11 protein

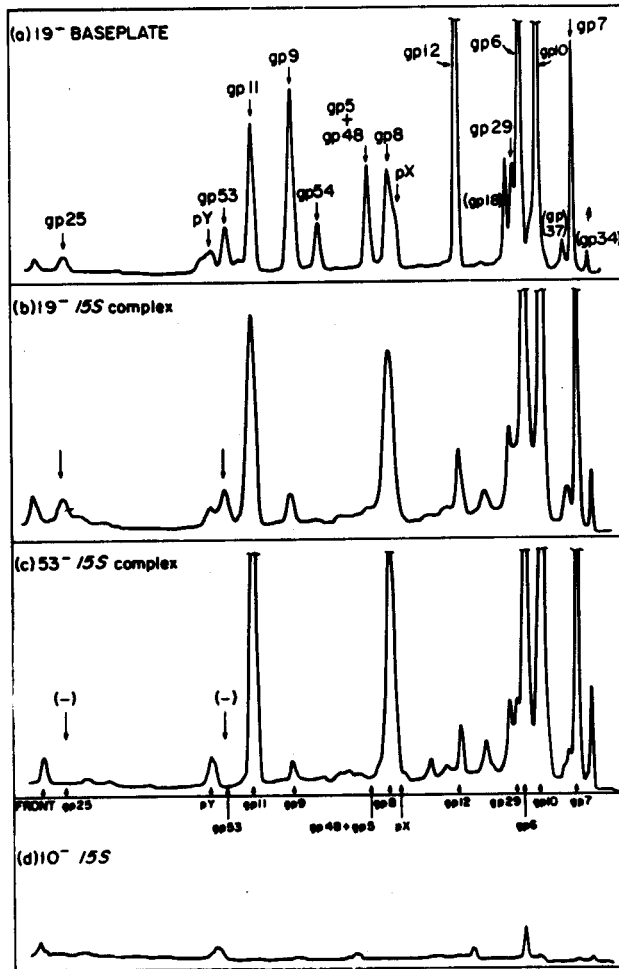


FIG. 7. Protein composition of baseplates and baseplate precursors. These tracings of SDS-gel patterns show the protein compositions of complete baseplates (a), the 15S precursor from the same lysate (b), and the 15S complex from 53⁻ infected cells (c). The infected cells were labeled with ¹⁴C-mixed amino acids, and lysates were fractionated by sucrose gradient centrifugation. The peak fractions of the relevant distributions were analyzed. The bottom panel (d) shows a control from a 10⁻ lysate, unable to form the 15S complex. pX is now known to be gp27 (Berget and Warner, 1975), and pY is probably *gpfrd* (Mosher and Mathews, 1979).

forms the tips of the morphological spikes on the distal face of the baseplate (Crowther et al., 1977), indicating that the arm assembly pathway proceeds from the outer vertex towards the hub.

Although the addition of gp11 to gp10 is not an obligatory step in the assembly of the 1/6th arms (Edgar and Lielausis, 1968), in the absence of gp7 all of the gp10 and gp11 is found in a 10/11 complex (Berget and King, 1978). The product of gene 10 can be thought of as the keystone protein in the assembly of the 1/6th arms. In the absence of gp10, no 1/6th arm-related assemblies are formed (Kikuchi and King, 1975a). The gene 7 protein adds to the 10/11 complex to make the 7/10/11 complex. Then, in sequential order, the products of genes 8, 6, 53, and 25 add to the growing 1/6th arm structure.

The product of T4 gene *frd* (dihydrofolate reductase)

is also a component of the 1/6th arm structures. Mathews and co-workers have determined that it is present in the 1/6th arm intermediate which accumulates in cells infected with a gene 53 amber mutant (Mosher and Mathews, 1979). However, the exact point of assembly into the 1/6th arm structure is unknown.

Assembly of the central hub. The assembly of the central hub of the baseplate is more complicated than that of the 1/6th arm. In addition, the much smaller mass of the hub, the smaller size of the hub proteins, and the presence of aberrant side reactions have rendered its assembly somewhat more difficult to dissect (Kikuchi and King, 1975b). The central hub is comprised of six structural proteins (gp5, gp26, gp27, gp28, gp29, and *gptd*). These structures accumulate as a 22S complex in cells infected with mutants defective in 1/6th arm assembly (i.e., 10⁻ mutants). They have not been directly visualized by electron microscopy. In addition, because of the much smaller mass of the hub, it has only been identified through *in vitro* complementation assays.

The *in vitro* complementation between different pairs of mutants in the hub complex was much less efficient than the complementation reactions in the 1/6th arm pathway (Kikuchi and King, 1975c). A number of pairs of mutant-infected extracts complemented very weakly, if at all. This might be due to polarity effects (Stahl et al., 1970) or to the nature of the pathway itself. A further feature limiting the efficiency of the complementation reactions is the ability of the complete 15S arm complexes to polymerize in the absence of the arm into aberrant baseplate-like structures such as those shown in Fig. 8. These are less stable than the complete baseplates, and many of them are recovered as 40S structures which have transformed to a starlike configuration (Kikuchi and King, 1975b). The 40S structures are missing some of the proteins found in the hexagonal 70S forms, namely, gp9 and gp12 (Fig. 9). It seems that the addition of gp9 and gp12 to the aberrant hexagons is inefficient, resulting in baseplates lacking these proteins which transform to stars.

The 22S complex appears to form through the interaction of two smaller complexes, a 7S gene 29 complex and a 10S complex composed of gp5 and gp27. It is not clear whether the formation of the 22S complex is a radial polymerization process or the interaction of two already cyclical complexes.

Recent results from Kozloff's laboratory have extended the original work of Kikuchi and King (1975c) by showing that the products of genes 26, 28, and *td* are structural components of the baseplate central hub and by determining where these proteins are involved in the assembly pathway of the central hub (Kozloff, 1981; Kozloff, personal communication). To locate these structural proteins they have used an *in vitro* "chase" technique which allows the detection of proteins which occur in a few copies per phage particle. Through the use of this technique, protein complexes which were not detected in the original study by Kikuchi and King were identified. The central hub assembly pathway currently suggested by Kozloff is incorporated in Fig. 1 (Kozloff, personal communication). In this pathway gp29 and *gptd* interact to form a 29/*td* complex which sediments at 7S. Gp51 catalyzes

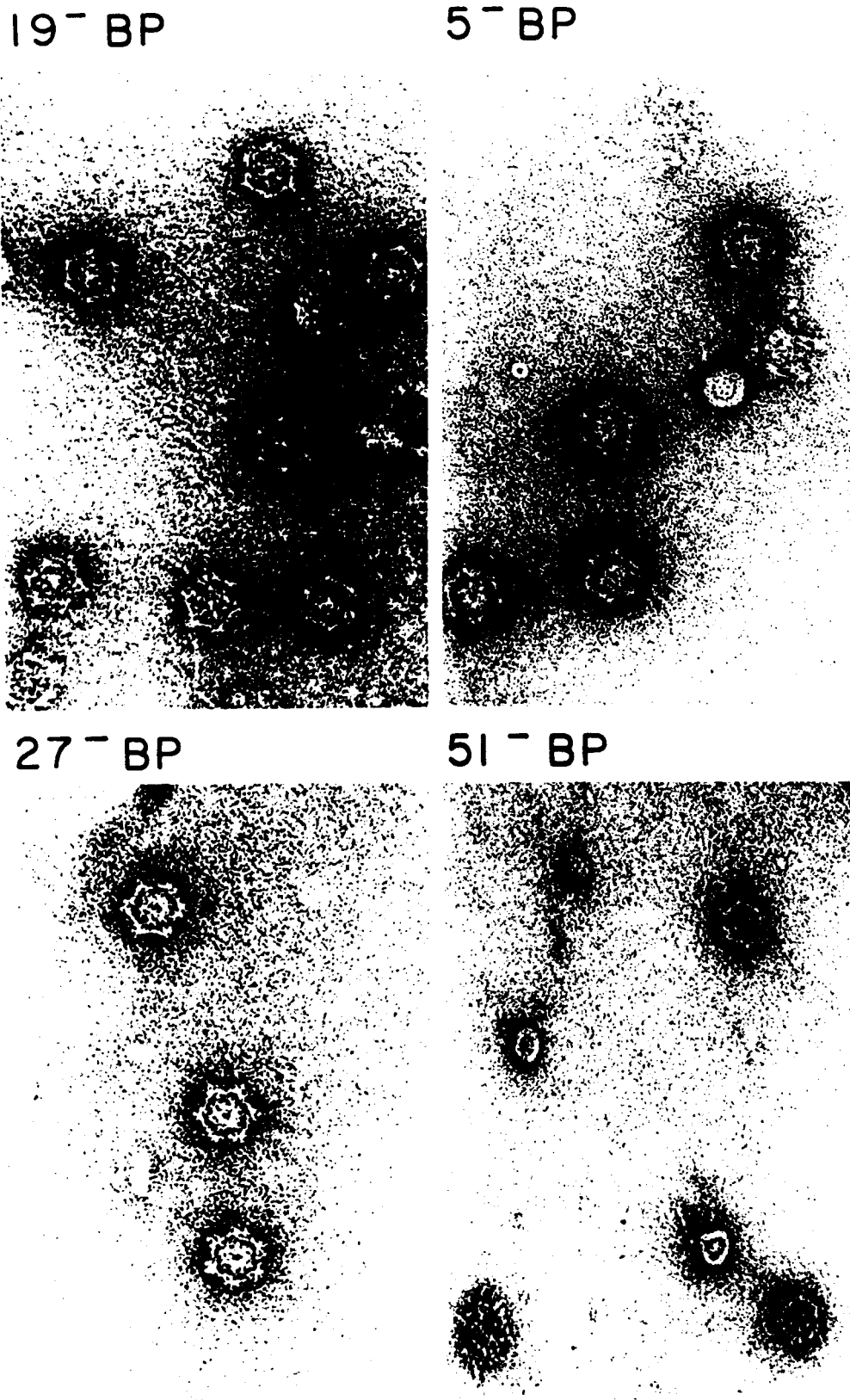


FIG. 8. Inactive baseplates. The top left panel shows complete baseplates from 19⁻-infected cells. Note the exclusion of negative stain from the very center of the structures. The other three panels show structures accumulating in cells defective in various steps in hub assembly. The centers of these structures are penetrated by the stain. These baseplates are inactive as phage precursors when tested by *in vitro* complementation (Kikuchi and King, 1975b).

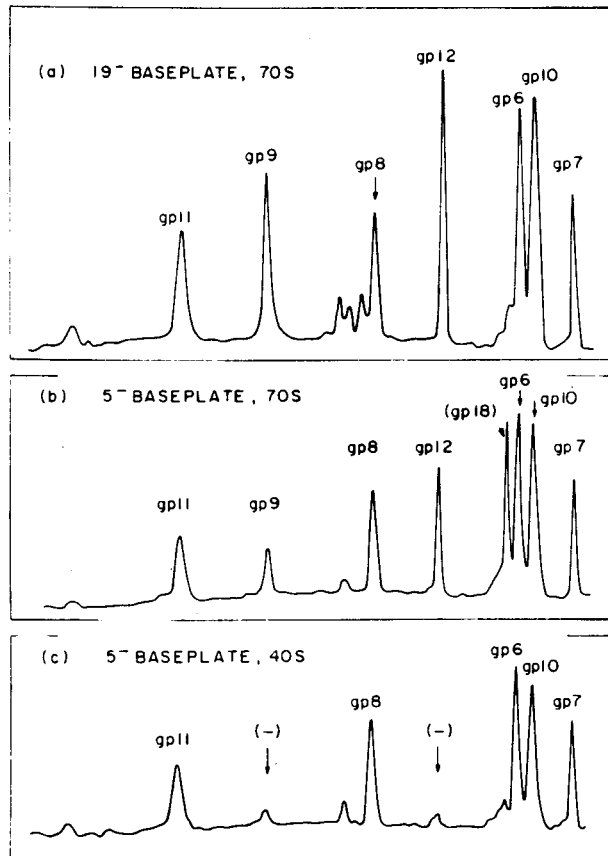


FIG. 9. Protein composition of inactive baseplates. The inactive baseplates found in hub-defective infected cells are isolated as two populations, 70S hexagons (b) and 40S stars (c). The 70S structures from 5⁻-infected cells (shown in Fig. 8) do not incorporate gp48 or gp54 (which are present in the lysates) and are therefore not substrates for tube assembly. The stars lack gp9 and gp12 in addition. From Kikuchi and King (1975b).

the conversion of this complex to a 14S 29/td complex. The molecular nature of this conversion is unknown; however, this form of the 29/td complex is the substrate for the ensuing assembly reactions. Through separate assembly reactions gp5 and gp26 form a 5/26 complex and gp27 and gp28 form a 27/28 complex. These three complexes, 29/td, 5/26, and 27/28, then coassemble in an unknown order to form the central hub structure.

The central hub pathway proposed by Kozloff and co-workers is not completely in accord with that proposed by Kikuchi and King (1975c). For example, the *in vitro* complementation data of Kikuchi and King suggest that gp27 and gp5 interact to form a complex (presumably a 5/26/27/28 complex) before the interaction with a gp29-containing complex. This type of interaction is not ruled out by the pathway shown here, but no additional data are available to confirm this proposal. Clearly, many proteins are interacting in complex ways in this pathway. Some of these proteins are being utilized in metabolic pathways as enzymes at the same time as they are being withdrawn as structural components of virus particle. The experiments of Kikuchi and King relied on the ability

of assembly complexes to remain associated through sucrose density gradients. Although it is clear that such was the case for the analysis of the 1/6th arm pathway, it may not have been true for the analysis of the central hub pathway. The experiments using the *in vitro* "chase" technique reported by Kozloff (1981) examined only a subset of the possible pairwise mixtures of detective extracts to derive the information summarized in Fig. 1. A serious effort needs to be made, using a combination of the techniques now available, to further characterize the central hub pathway.

Assembly of the sixfold symmetrical baseplate structure. Once the 1/6th arm and central hub subassemblies are completed, they spontaneously assemble to form the characteristic hexagonally shaped baseplate. This intriguing assembly reaction involves the interaction of one central hub structure with six 1/6th arm structures in what must be a radial polymerization reaction: six 1/6th arm assemblies polymerizing around one central hub subassembly. Kikuchi and King (1975b) demonstrated that in the absence of central hub structures, 1/6th arms can assemble slowly into baseplate-like structures which are missing the central density characteristic of baseplates containing central hubs (Fig. 8). These aberrant baseplate assemblies can also be formed in 25⁻ lysates which have incomplete arms but complete hubs. This suggests that gp25 links the arms to the hubs.

During or upon completion of the coassembly of the 1/6th arms and central hub, the products of genes 9 and 12 associate with the baseplate. These proteins are not associated with the arm before their radial polymerization. There is no obligatory order to the addition of these structural components, as both 9'/12⁻ and 9'/12⁺ baseplates can be isolated (Fig. 10). Note that the baseplate can be assembled from arms that lack gp11, the tips of the spikes, but that gp12 cannot add to these structures.

The products of genes 48 and 54 then add to the baseplate in a sequential fashion to complete the assembly of the baseplate (Berget and Warner, 1975; King, 1971; Meezan and Wood, 1971). These last two structural proteins of the baseplate prepare this structure for the polymerization of both the tail tube and sheath. These proteins may represent the length determiner and initiator proteins for the tail tube polymerization reaction (Berget and Warner, 1975; Duda and Eiserling, 1982).

Tail Tube and Sheath Assembly

The assembly of the distinctive tail tube and sheath structures occurs only after the baseplate structure has been completed via incorporation of gp48 and gp54. One or both of these species probably form the sites that bind the first annulus of tube subunits. The polymerization of gp19 utilizes 144 monomers of gp19 and terminates at a total tail length of exactly 100 nm. The proximal end of the tail tube is modified by the incorporation of gp3, which may act as a "glue" protein between the tail tube and the sheath, connecting these two tubular structures only at this end of the tail (King, 1971). During tail tube polymerization, sheath monomers, gp18, start polymerizing around the tail tube and continue to polymerize until the

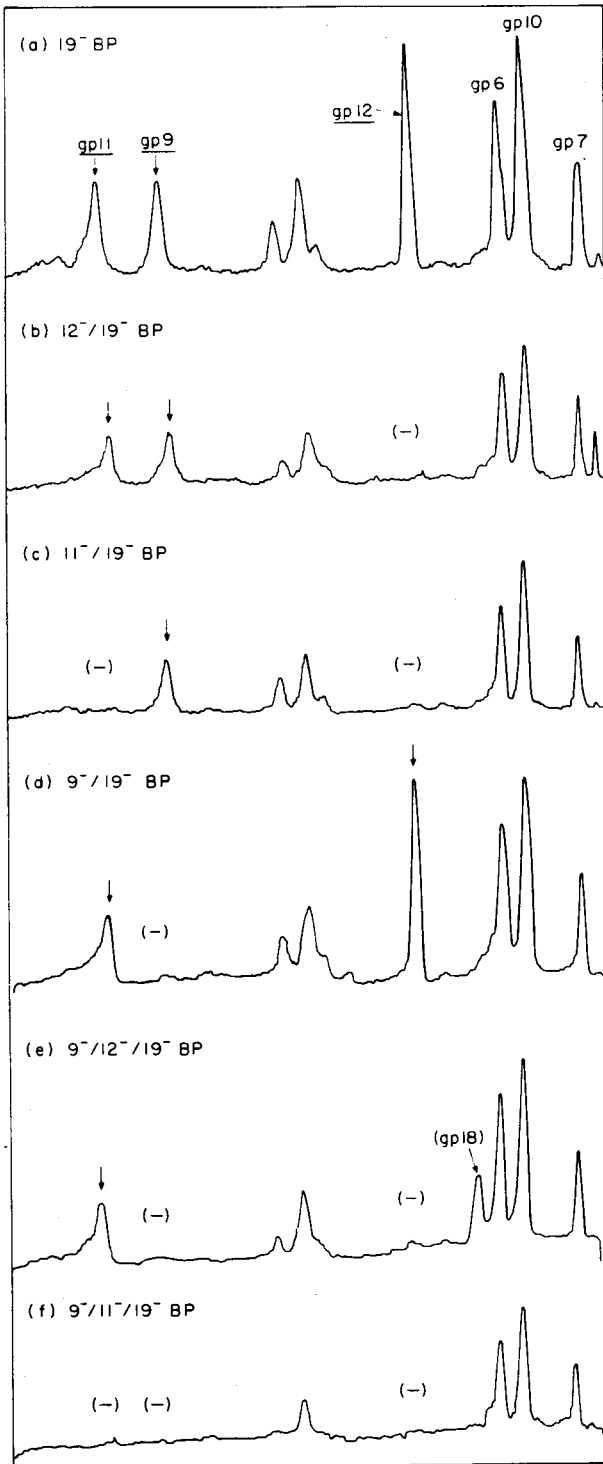


FIG. 10. Independence of gp9 addition from gp11/gp12 addition. The tracings are of SDS-gel patterns of radioactively labeled baseplates isolated from mutant-infected cells.

sheath reaches the proximal end of the tail tube. At this point the tail structure is completed by the addition of gp15, which stabilizes the sheath structure (King and Mykolajewycz, 1973) and produces the "connector" structure required for T4 head attachment.

The polymerization of the tail tube is very tightly regulated at the assembly level. No case has been reported of the polymerization of precursor tail tube subunits into tubes in the absence of the baseplate, either in vivo (King, 1971) or in vitro with purified tube subunits and baseplates (Wagenknecht and Bloomfield, 1977). The first set of tube monomers is clearly bound by the baseplate. However, each newly formed annulus must constitute the active site for binding of additional monomers from solution. This is analogous to the growth of *Salmonella* flagella by polar addition of monomeric flagellin subunits to the distal tip of the growing flagella (Asakura et al., 1968). In this case the subunits undergo a substantial conformational change upon binding (Uratani et al., 1972). We assume that the tail tube utilizes the same mechanism.

Tail sheath polymerization follows the same pattern but is less tightly controlled. In the absence of a proper substrate for tail sheath polymerization, an aberrant form of polymerized sheath monomers called polysheath forms. Moody (1973) proposed that these structures are analogous to the packing of the sheath subunits in the contracted sheath, but polysheath is relatively easily dissociated into subunits, unlike the contracted sheath.

Tail Length Determination

The assembly experiments reveal that the length of the sheath is determined by the length of the tail tube. If tail tube length is limited due to a shortage in supply of tube subunits [e.g., *ts19* mutants at intermediate temperatures], the sheath subunits polymerize only as far as the partially formed tube (King, 1971). Two mechanisms have been proposed for the regulation of the length of the T4 tail. Kellenberger (1972b) suggests that the tail tube monomers polymerize because the total free energy of a monomer is lowered through intersubunit bond formation. However, to achieve the most favorable bonding, each subunit may have to become slightly deformed. As more subunits are added, the deformation strain per subunit increases until finally the positive free energy accumulated equals the negative free energy associated with intersubunit bonding. Thus it then becomes unfavorable to add any more subunits, and the polymerization reaction stops.

The feasibility of such an "accumulated strain" hypothesis has been investigated by Wagenknecht and Bloomfield (1975). Their theoretical analysis shows that such a model is thermodynamically feasible for producing very narrow length distributions. In a further study Wagenknecht and Bloomfield (1977) have shown that gp19 monomers can be polymerized onto baseplates in a partially purified in vitro system yielding tail tubes with a length distribution narrowly centered around 100 nm, the same value found in vivo. Thus, depending on the purity of their system, few intracellular factors other than baseplates and gp19 are required to produce tail tubes of the correct length.

A second hypothesis is that there may be "tape measure" molecules along which the gp19 monomers polymerize (King, 1968; King, 1971). The length of this molecule then determines the length of the tail

tube. The natural candidate would be either gp48 or gp54. Recently it has been demonstrated that the product of gene 48 can be localized to the tail tubes which have been "broken off" tail tube-baseplate complexes (Duda and Eiserling, 1982). By electron microscopy, such tail tubes have been shown to have their central cavity filled with material that excludes stain. Occasionally tail tubes can be seen which have material protruding from one end and stain penetrating into the other. This suggests that these tail tubes may be filled with gp48, and it seems likely, given its position in baseplate assembly, gp48 may indeed be a length-determining molecule for the T4 tail (see Eiserling, this volume). The sensitivity of baseplates to proteases (Kikuchi and King, unpublished data; Bloomfield, personal communication) also suggests that a protein may be involved in tube initiation and length determination. Work on phage lambda suggests that gpH, a protein product found in the mature tail, acts as a tape measure for tail length determination (P. Youderian, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, 1978; Youderian and King, manuscript in preparation; R. Hendrix, personal communication).

PROTEIN-DETERMINED ORDERED ASSEMBLY REACTIONS

The assembly of the 1/6th arm of the baseplate is a straightforward example of a pathway controlled by protein-protein interactions. Four of the six assembly reactions in this pathway are ordered by the previous interaction of a structural protein with the growing complex (the additions of gp8, gp6, gp53, and gp25). Nothing is known about the molecular mechanisms governing the assembly of this structure; however, several formal possibilities exist for the protein-determined ordering of this pathway.

Modifications in the primary structure of a component in this pathway by proteolysis or some other covalent modification could change the nature of an incorporated protein so as to provide an otherwise obscured binding site for the next protein in the pathway. Proteolytic cleavage of the proteins involved in T4 tail formation has not been observed in pulse-chase experiments, but subtle changes cannot be ruled out. Second, two different structural components of the 1/6th arm may each contribute one half of a binding site for the next protein in the pathway. The interaction of each of these proteins with the third

may not be strong enough until the first two are combined into a complex. Finally, two proteins may interact with each other in a fashion which causes a conformational change or refolding of one or both of the proteins, thereby uncovering the binding site for the third component in the pathway.

All of these possibilities are potential regulatory mechanisms which may be involved in the control of the assembly of the T4 tail. In higher organisms these mechanisms, overlaid with temporal regulation of synthesis and compartmentalization, probably form a subset of regulatory devices used in developmental programming. Yet the simplest problem of sorting any three proteins into an assembly sequence has not been solved at the molecular level.

GENE CLUSTERING

All of the proteins of the baseplate 1/6th arm, except gp25, map in the 53-12 gene cluster. Similarly, all the hub proteins except gp5 map in the 25-54 cluster. The protein products of these genes are clear examples of tightly interacting proteins whose structural genes have remained closely linked. There are numerous hypotheses for such linkage phenomena. Since it is possible that these genes evolved from ancestral genes by duplication, the clustering may reflect their evolutionary history. On the other hand, it seems reasonable that natural selection may act to keep the genes for tightly interacting proteins closely linked since they may evolve as a unit. In this regard it is interesting that the two gene clusters are quite distant from each other, and one can imagine recombination between T4-like phages in which the clusters would be exchanged. Note that one of the first genes in the arm cluster, gene 5, codes for a protein of the hub and maps adjacent to gene 53, which may link the arm to the hub. Similarly, the first gene in the hub cluster, gene 25, codes for a protein of the 1/6th arm which is the best candidate for binding the arm to the hub. These additional features of the clustering may preserve particularly critical structural interactions between the hub and 1/6th arms in the overall baseplate structure.

This paper was prepared with the support of National Institutes of Health grant GM-17980 and National Science Foundation grant PCM 80-11661 to J.K., and National Science Foundation grant PCM 81-04523 to P.B.B.