

**4**  
*Pathways  
in Viral  
Morphogenesis*

*Peter B. Berget*  
University of Texas  
Health Science Center

## INTRODUCTION

Over the last 30 years, the study of viral morphogenesis has provided the paradigm for the study of the genetic control of structural development. Early in these studies the observation was made that bacteriophage development is similar to the pathways of intermediary metabolism in that bacteriophage morphogenesis occurs through controlled pathways. It has been recently stated that the rapid success of investigators in identifying the existence of and elucidating the individual steps in assembly pathways in phage morphogenesis has led to their incorporation into textbooks in an oversimplified or misleading form (King, 1980). It is the purpose of this chapter to clarify some of the thoughts concerning the roles of pathways in biological construction, and to provide a detailed description of a controlled-assembly pathway in bacteriophage morphogenesis.

## REQUIREMENTS OF BIOLOGICAL CONSTRUCTION

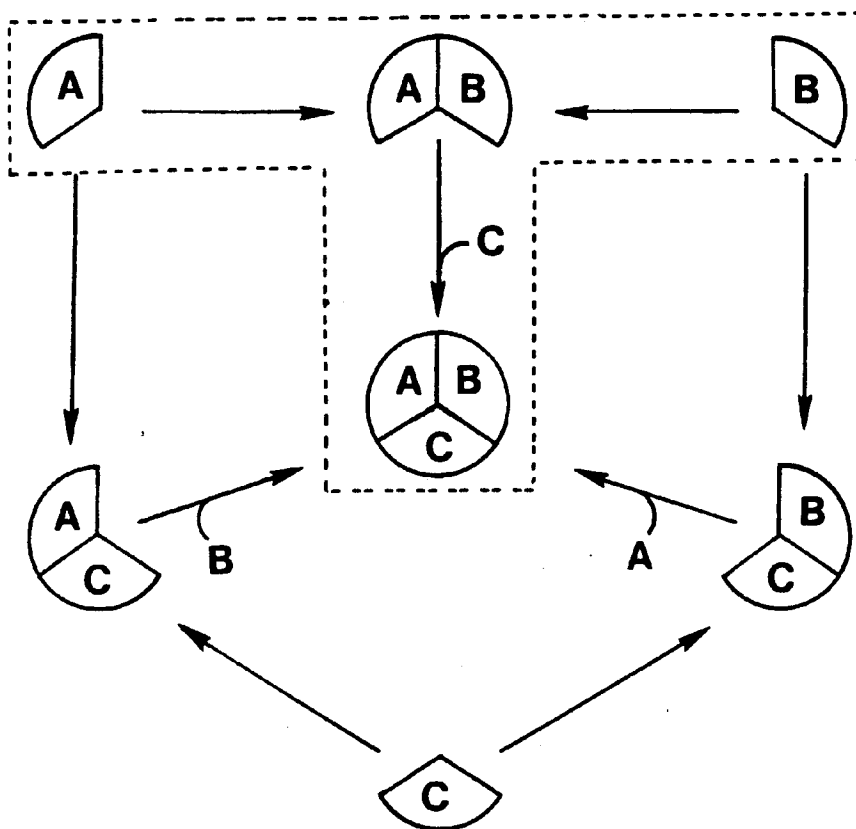
Supramolecular assemblies or subcellular organelles are built from a variety of biological components. For most bacteriophages only two types of biological molecules, protein and nucleic acid, are used in any great amount. The relative mass proportions of protein and nucleic acid can vary over a substantial range. For example, T4 phage is approximately 50% protein and the filamentous phages, such as fd or M13, are approximately 80% protein by weight. Furthermore, the number of protein molecules involved can vary from 60 in the simplest icosahedral viruses to several thousand in the helical and large icosahedral viruses. We now know that nucleic acid is the genetic component of viruses and that in the simplest of interpretations the protein component provides the packaging and delivery system for this information into the next host. Although the building blocks for phage structures are limited in their types, the variation that can exist in different protein molecules makes the actual number of kinds of building blocks effectively infinite. In this chapter I shall mainly address the problem of biological construction from protein molecules.

The primary requirement of biological construction is that the need of the organism for the structure should be met by the final product. This may seem like an obvious statement but all other requirements stem from it; clearly, if a bacteriophage constructs a DNA-injection organelle as part of its overall structure, it must in fact work as a DNA-injection organelle. The secondary requirements of biological construction are that the structure must be constructed in an appropriate time frame for the organism and that the construction should be carried out efficiently. It is my contention that the use of pathways and subassemblies allow these requirements of accuracy and efficiency to be met.

## THEORETICAL CONSIDERATIONS OF PATHWAYS

Consider a very simple assembly system in which only three components—A, B, and C—interact to form an ABC complex. As shown

in Figure 1, there are three possible orders in which ABC can be built: joining either A and B, A and C, or B and C first, and then adding the final component. Each of these is a potential pathway for the assembly of ABC. As I shall describe below, bacteriophages are invariably assembled from their protein and nucleic acid components according to very specific pathways; that is, the building blocks assemble in a strict order. In this simple example there are three possible assembly pathways. In Figure 1, a particular pathway,  $A + B \rightarrow AB + C \rightarrow ABC$ , is boxed. If only the boxed pathway is functional, A and B will bind one another, but C will bind neither A nor B alone. In a real structure, as the number of components increases, the number of possible pathways becomes very large. One



**FIGURE 1** Potential assembly pathways for a three-component system. The subunits A, B, and C of the ABC complex may interact in three possible ordered pathways of assembly by initiating with unique heterodimer formation, as described in the text. The outlined pathway may be selected as the unique pathway by "nucleation control" of the formation of the AB heterodimer.

great advantage of assembly by specific pathways is the ease by which control can be exerted over the assembly pathway. One such control is the regulation of the number of structures made. Again referring to the example in Figure 1, in order to regulate the entire assembly process control need only be exerted on the A + B step, rather than all three heterodimerization steps. This has often been referred to as control of the "nucleation" of an assembly process, and such regulation requires that assembly occur by a specific pathway. A second advantage of assembly by a specific pathway is that "spatial" control can be exerted. Imagine that a closed protein shell (as many viral capsids are thought to be), which has internal and external components, were being constructed. If the shell were built and closed before the internal components were added, the resulting structure would no doubt be nonfunctional, since the internal components could not be added after the outer shell were closed. Thus, if assembly were required to occur by a specific pathway in which the internal components were inserted before the shell were closed, the assembly of nonfunctional structures would be completely avoided. This type of control is no doubt very important in the assembly of structures, such as viruses, that contain a large number of protein molecules. (See Chapter 5, this volume, for details.)

Inherent in the concept of pathways is that specific "intermediates" or subassemblies are generated. It is through the consideration of subassemblies and "assembly-line" construction that useful thoughts on this concept have been generated. One of the first discussions on the benefits of subassemblies and assembly-lines in biology was put forth in "Principles and Problems of Biological Growth" (Crane, 1950). Several of Crane's thoughts are expanded upon in the remainder of this section, in which it will be seen that accuracy and efficiency are direct results of the use of subassemblies.

A useful approach to understanding the roles pathways and subassemblies play in the accuracy of structure formation is to consider the probability that a given structure will be formed successfully under a defined set of rules. For example, let us consider the assembly of a supramolecular structure having 1000 subunits that may or may not all be the same. In this example we make the assumptions: (1) the probability of making a mistake in the assembly of one subunit into a structure or in the interaction of two subassemblies is 0.1%; and (2) one incorrect assembly reaction either makes the structure unusable or causes the substructure to become assembly-incompetent for further reactions. The probability  $P(f)$  of successfully finishing the structure by adding one subunit at a time is the product of the probabilities of a correct reaction at each step, or  $P(f) = (1 - x)^{n - 1}$ , in which  $n$  = the number of subunits and  $x$  = the probability of incorrect association at each step. Thus,  $P(f) = (0.999)^{999} = 0.368$ , which means that there is only a 36.8% chance of correctly completing such a structure by using this mode of assembly. Let us explore alternate means of assembling this structure by determining the efficiency of subunit utilization.

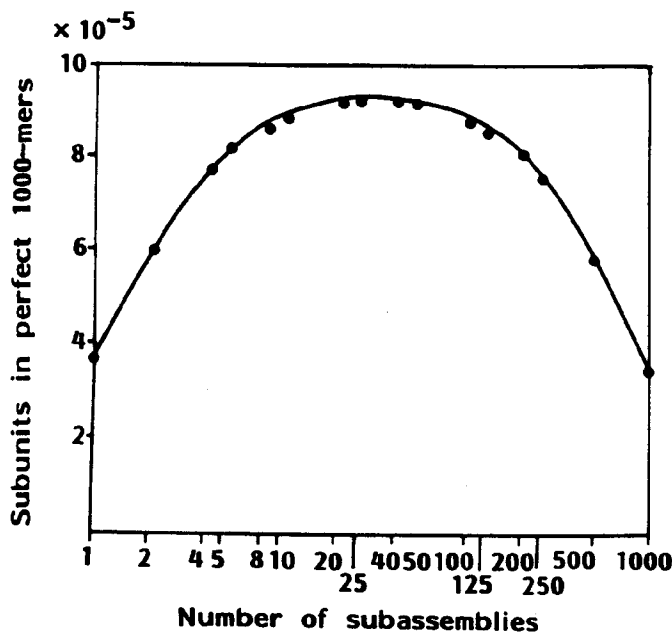
Assume now that we are given 1,000,000 subunits to start with, that any structure consumes its normal complement of subunits even if a mistake is made during its assembly, and that the error frequency

is 0.1%. Furthermore, assume that if a mistake is made in the assembly of a component that is an integral part of the whole, it will be incapable of entering into succeeding rounds of assembly. How many subunits will reside in perfect structures using a given assembly pathway? If we use an assembly scheme in which each 1000-mer is made by 999 successive additions of subunits, the probability of successful assembly of each 1000-mer is 0.368; thus, 368 out of a possible thousand 1000-mers will be perfect, or 368,000 of the 1,000,000 subunits will reside in perfect structures. If we make the structure by forming 500-mers and then assembling two of them into a 1000-mer as the last step, a different result is obtained. The probability of successfully assembling a 500-mer is  $(0.999)^{499}$  or 0.607; thus, 606,985 monomers will exist in perfect 500-mers (using only integer values, 1,213 perfect 500-mers will be formed consuming 606,500 monomers). Thus, 606 1000-mers could be assembled from the 1,213 500-mers, but only  $(606)(0.999) = 605$  perfect 1000-mers will form. Using two 500-mer subassemblies to form 1000-mers has increased the efficiency of usage of monomers from 368,000 to 605,000. In Figure 2 the results of calculating efficiencies by this method are presented for forming 1000-mers by different combinations of subunits and subassemblies.

According to this logic, the most successful way of assembling this structure is from 25 subassemblies with 40 subunits, or or 40 subassemblies with 25 subunits, in which almost every start results in a perfect structure ( $P(f) = 0.939$ ). In general, the most successful way of assembling a multisubunit supramolecular structure using one type of subassembly is to maximize both the total number of subassemblies per final structure and the number of subunits per subassembly. The "ideal" number of subunits per subassembly is the integer factor nearest the square root of the total number of subunits in the structure.

The probability of successful assembly at each step has been termed the "degree of selectivity" (Crane, 1950). We can investigate the theoretical "degree of selectivity" that may be required in such a biological process. By varying the "degree of selectivity", DS, one can calculate the probability of successful assembly of a 1000-mer using the best pathway of 25 40-mers with the following equation:  $P(f) = (DS)^{24}(DS)^{39}$ . Figure 3 shows the plot of the probability of successful assembly of a 1000-mer using this best pathway as a function of degree of selectivity. Clearly the success of assembly depends quite strongly on the "degree of selectivity" at each step in the assembly pathway. If this model assembly system is to create structures with at least a 50% chance of the final structure being correct or a 50% chance of successful completion, then each step in the assembly pathway needs to be carried out with an accuracy of about 99%. This demonstrates that an acceptable degree of successful assembly can be achieved without impossibly stringent requirements at each step of this assembly reaction. Compare this to the acceptability of a genetic replication system operating with a 0.1% error rate at each step in the "assembly" of the nucleic acid polymer!

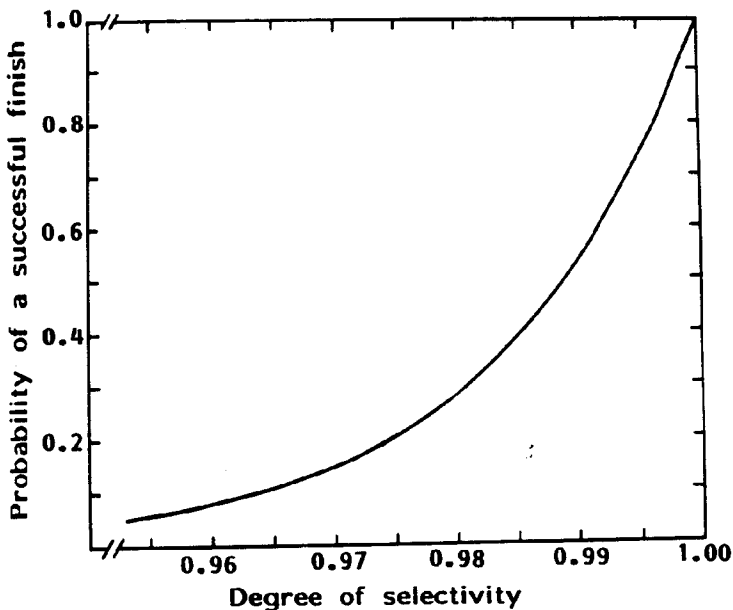
How accurate is the process of assembly for bacteriophages? One possible measure of the total accuracy of the process is the relative



**FIGURE 2** Efficiencies of alternate modes of assembly. The efficiency of utilizing subunits is plotted against the number of subassemblies used to form the final structure, as described in the text. Starting with 1,000,000 subunits, the number of subunits found in perfectly assembled 1000-mers is plotted against the number of subassemblies used to make the 1000-mer. For example, 368,000 subunits reside in perfect 1000-mers if the 1000-mers are made by 999 successive monomer additions (i.e., 1 or 1000 subassemblies). If two 500-mers (or 500 dimers) are used to form 1000-mers, then 605,000 of the 1,000,000 starting subunits end up in perfect 1000-mers, etc.

infectivity of the final particles, that is, what percent of the final structures are capable of initiating a subsequent round of infection in a new host? For *Escherichia coli* bacteriophage T4 and *Salmonella typhimurium* bacteriophage P22 this number is between 50% and 100%; however, for many animal viruses the number is less than 10%. Surely these numbers reflect both the accuracy of the final structure and the efficiency of the process of viral infection. However, it is clear that for bacteriophages such as T4, which contain more than 1000 protein molecules and have an infectious particle to physical particle ratio near 1, the overall assembly process must indeed be very accurate.

A further result of the subassembly principle is the increased speed with which a complicated structure can be made. If each assembly reaction, which may occur between subunits or subassemblies, takes a certain amount of time, the subassembly scheme can be shown to be more efficient. If each reaction in the previous example were to consume 1 second, the assembly of a 1000-mer by



**FIGURE 3** The effect of the degree of selectivity on the probability of successful assembly. The probability of successfully assembling a 1000-mer from 25 40-mers (or 40 25-mers) is plotted against the probability of a correct assembly reaction at each step of the assembly process.

successive monomer additions would take 999 seconds to complete. However, assembling the same structure from 40 25-mers would take  $39 + 24 = 63$  seconds to complete, which is a direct result of the simultaneous construction of subassemblies.

The reader should dwell for a moment on the thought of assembling a 1000-mer from two different types of intermediate subassemblies (a logical extension of the discussion above). Suppose that 10-mers are initially formed by monomer additions and that ten of these 10-mers subsequently assemble to form 100-mers, then ten of these 100-mers assemble to form the final 1000-mer. This can be shown by the same arguments made above to be an even more efficient pathway of assembling the desired structure. Nevertheless, the important points about accuracy and efficiency have been made. Using only probabilistic considerations, the use of subassemblies and "rules" or pathways by which they are brought together make the process of assembly possible using relatively relaxed "degrees of selectivity". Again, in contrast, consider the "degree of selectivity" required to replicate genetic information faithfully; this is on the order of 0.9999999999 (Drake, 1969).

H. R. Crane's discussion of "degree of selectivity" closes with a story too good to resist re-telling some 34 years later. "...During the early part of the war I had occasion to go through a vacuum

tube factory, where the miniature tubes for proximity fuses were being made. There I saw rows of women assembling the most minute parts under magnifying glasses, and giving the every appearance of being highly skilled jewelers. I asked if the training of these women did not present a great problem. To my surprise my guide told me that a completely green person could be put on the production line four hours after walking in the door. "Of course," he said, "they make mistakes, but each worker does only a small thing, and the pieces he spoils are, of course, not used in the next subassembly operation." Here a highly complex and precise product was being made entirely by unskilled hands, not only by virtue of the fact that each worker had only a small job to learn to do, but also by virtue of the fact that the system provided for casting aside the imperfect units at many levels of assembly. Therefore no individual was required to make a high score on accuracy. It was at this time that I was struck by the apparent parallels that existed between this method of factory production and many processes in nature, from evolution and natural selection down to the relations of the smallest molecules. The problem of how the fantastically complex and precise 'products' could be manufactured in nature without the requirement of an excessively high degree of selectivity or accuracy at any one stage of development seemed slightly more understandable."

Now, in addition to the rules of assembly put forth for the example system described above, we impose an additional rule: namely, as soon as a mistake is made in the assembly process of any component, the component is discarded and we start again. With that rule, the total number of monomers available for incorporation into perfect final structures increases. Casting aside imperfect structures at any point in assembly provides increased accuracy and efficiency.

The advantage of assembling supramolecular structures from subassemblies should now seem obvious. However, in the biological realm, where is the "blueprint" for the final structure, where are the skilled or unskilled laborers that assemble the desired structure, i.e., what determines the pathways through which subassemblies are formed and subsequently brought together? These questions impinge on the central problems in developmental biology, and they are the questions that investigators studying viral morphogenesis ask.

The specificity or degree of selectivity that gives rise to the interaction of subassemblies to form the final structure and to form the subassemblies themselves must somehow reside in both the protein molecules that comprise the final structure and any other proteins that are utilized as "catalytic" factors in the overall assembly scheme. All would agree that proteins seem ideally suited for these roles; however, the fact that some proteins must simultaneously serve as structural components and catalytic factors for assembly and must perform some highly specific function in the final structure is indeed remarkable.

The analysis of morphogenetic assembly pathways in bacteriophages has proceeded through two basic approaches. The first is a genetic approach through which the developmental phenomena are observed. The second is a biochemical approach through which the molecular mechanisms of the phenomena are elucidated. The success of the former

approach in identifying many pathways of morphogenesis and our current lack of understanding of the underlying molecular mechanisms indicate both the power of genetics and the complex nature of the problem.

As alluded to in the introduction, pathways of morphogenesis in viral development were observed early in the general studies of bacteriophages, especially T4. These observations were made possible by the development of the powerful genetic system of temperature-sensitive and amber conditionally lethal mutations in genes necessary for T4 growth (Epstein et al., 1963; Edgar, 1969). A virus, carrying a conditionally lethal mutation in an essential gene, grows under one set of permissive conditions but fails to grow under another defined set of restrictive conditions. At high temperatures or in the nonpermissive host, bacteriophage conditionally lethal mutants make all the active proteins normally found in the wildtype-infected cell, except for one. Using physical and immunological tools and a large collection of these conditionally lethal mutations in almost all of the essential T4 genes, investigators cataloged the defects in viral development that resulted from the absence of various gene functions (Epstein et al., 1963). It was clear from this study that gene products interacted with each other through specific pathways rather than in "random" fashion (Edgar and Lielausis, 1968; King and Wood, 1969). For example, defects in any of a set of seven genes resulted in the formation of T4 virus particles that lacked the characteristic T4 tail fibers. Defects in a larger different set of genes (ca. 20) resulted in the formation of free T4-head structures and tail fibers but no tail structures. Defects in a third set of genes (ca. 20) resulted in the formation of free tail structures and tail fibers, but no head structures. Such observations allowed these investigators to propose a model for the final steps in the assembly of the T4 phage particle. Three major subassemblies, the T4 head, the T4 tail, and the T4 tail fibers are each assembled through independent pathways and then joined together in two distinct assembly reactions. The T4 head and tail combine to form a fiberless particle; then and only then, the T4 tail fibers attach to finish the assembly of the virus particle. Thus, it is clear that the assembly of bacteriophage T4 utilizes the principles of subassembly and assembly-line structure development.

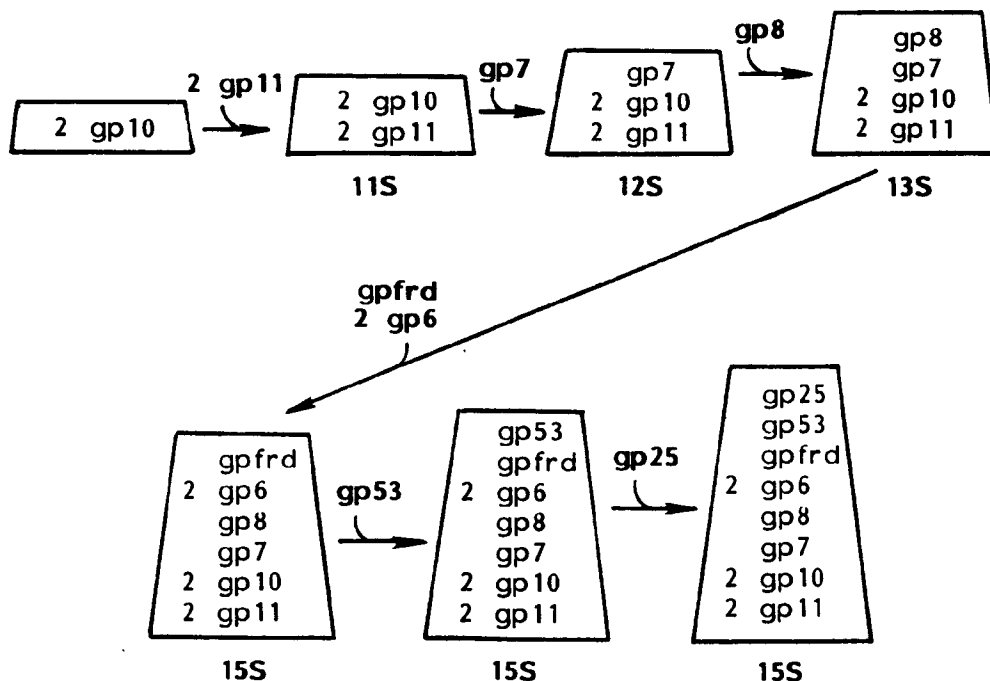
Further research has provided detailed descriptions of the morphogenetic pathways that operate in the assembly of these phage T4 subassemblies. A recent review that appears in the book *Bacteriophage T4* (Berget and King, 1983) provides details that supplements the material presented in the following section.

#### **ASSEMBLY OF THE MAJOR T4 BASEPLATE SUBASSEMBLY, A PROTEIN-DETERMINED, ORDERED ASSEMBLY PATHWAY**

The baseplate of T4 is a structure located at the distal end of the tail. It is the adsorption organelle of the phage and a complex molecular machine that triggers tail contraction and DNA ejection. This structure is composed of 18 different protein gene products, which are present in specific stoichiometries. The baseplate itself is assembled from two unique subassemblies: the "wedge", which contains

eight types of structural proteins and the "hub", which contains 5 different structural proteins (Kikuchi and King, 1975a,b,c; Mosher and Mathews, 1979; Kozloff, 1981). The hub and six wedge subassemblies co-assemble to form the hexagonally symmetrical baseplate. Then, five additional species of structural proteins then assemble onto this structure to finish the baseplate and to provide the nucleation sites for the polymerization of the T4 tail tube and sheath (King and Mykolajewycz, 1973; Berget and Warner, 1975; Meezan and Wood, 1971; King, 1971).

The assembly of the "wedge" structure provides an excellent example of the ordered, specific interaction of protein subunits to form a biological structure that is itself an intermediate in the assembly of a more complex structure. As depicted in Figure 4, the wedge is formed from the protein products of genes 11, 10, 7, 8, 6, *frd*, 53, and 25. No catalytic protein factors are required for the association of these proteins. Thus, the information or "blueprint" for their



**FIGURE 4** The morphogenetic pathway of the "wedge," the major subassembly of the T4 baseplate. The data used to construct this pathway come from Kikuchi and King (1975a, b, c); Plishker, Chidambaram, and Berget (1983; unpublished data); and Kozloff (1981; personal communication). The abbreviation "gp" indicates the protein product of an identified T4 gene; i.e., gp10 refers to the protein product of T4 gene 10. When known, the stoichiometries of the various protein components are indicated. Note that the point of addition of gpfrd is not exactly known; its point of addition is shown according to the data of Mosher and Mathews (1979).

interaction must be contained in the structural proteins themselves, and the final structure must be entirely determined by specific protein-protein interactions.

The assembly pathway of the wedge is unique and linear. The pathway was elucidated by Kikuchi and King (1975a,b, c) in a set of experiments utilizing a genetic approach. These investigators showed by sucrose-gradient sedimentation analysis and *in vitro* complementation tests (Edgar and Wood, 1966; Wood et al., 1968) that in infected cells, missing only the product of gene 10 (gp10), all of the other wedge components are found free in the cell in an unassociated, unassembled, active form. This observation pointed out the "keystone" role of gp10 in the assembly pathway. In infected cells missing only gp7, the structural components gp10 and gp11 are found associated with each other in a 10/11 complex, whereas the rest of the wedge components remain unassembled. In similar fashion, in infected cells lacking only gp8, an assembly complex consisting of gp10, gp11, and gp7 (10/11/7 complex) is found, while the rest of the wedge components remain unassembled. The complete set of experiments examined the state of assembly of all of the wedge components in infected cells lacking each structural element one at a time. Edgar and Lielausis (1968) had previously shown that in infected cells missing only gp11, defective phage particles accumulate that lack only gp11 and gp12, a component added later in the assembly of the baseplate. Thus, gp11 is not required for the assembly pathway depicted in Figure 4 to operate; in fact, gp11 can be added much later after the tail is completed, showing that in not all cases must there be an obligate order of addition. Nevertheless, in infected cells missing gp7, all gp11 is found associated with gp10 in a 10/11 complex (Berget and King, 1978). In T4 assembly the vast majority of assembly steps are obligately ordered.

Thus, the assembly pathway of the wedge shown in Figure 4 is initiated by the interaction of gp7 with the 10/11 complex to form the 10/11/7 complex. This complex is the only molecule in the infected cell with which gp8 can form a stable complex, resulting in the 10/11/7/8 complex. Likewise, this latter complex is the only molecule in the infected cell that can form a stable complex with gp6, etc. To put it another way, the only reactive site for accepting each structural protein into the growing wedge structure is a complex of proteins made from all the previous proteins in the wedge pathway. The linear assembly pathway of the wedge has two consequences: first, the reactive sites for protein-protein interaction are limited to the growing structure; second, all the proteins that comprise the structural elements of the wedge (except gp10, gp11, and gp7) are synthesized in forms unreactive towards each other, and these require incorporation into a growing wedge structure to become reactive. The overall effect is to limit the number of possible assembly reactions occurring in the cell and to generate only the desired final structure.

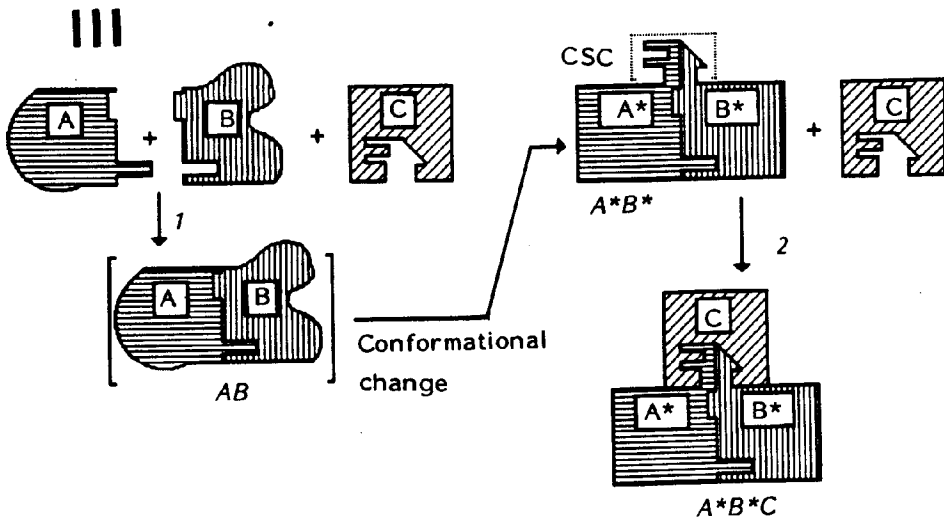
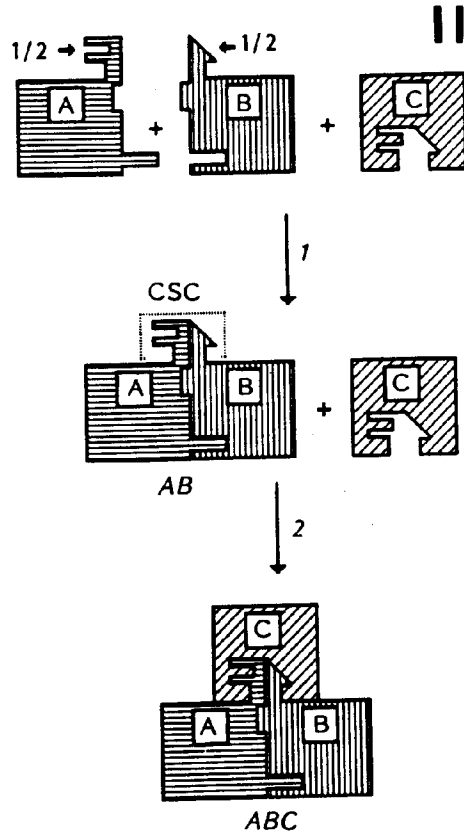
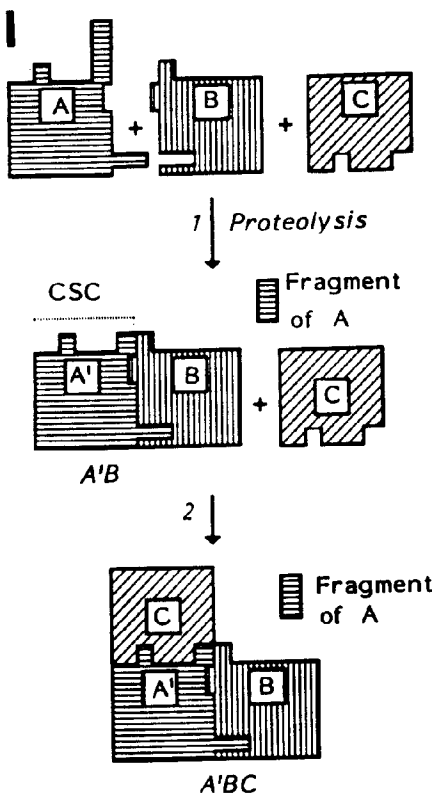
#### ***Possible Schemes for Control of Assembly***

In molecular terms, how is this linear assembly pathway generated? One possible control mechanism which is most certainly used in

eukaryotic development, is temporal control of synthesis of component parts. In T4 development, as in most virus systems, all components of the virion are synthesized simultaneously during the late stage of viral development. Thus, temporal control is not applied to the ordering of the assembly pathways in T4 morphogenesis. In analytical terms, the problem of ordering biological assembly reactions in the absence of temporal control can be simplified to one in which there are three protein molecules A, B, and C that combine to form an ABC complex via the pathway  $A + B \rightarrow AB$ ,  $AB + C \rightarrow ABC$ . That is, two proteins spontaneously interact to form the reactive site for a third protein.

The molecular mechanisms that may control such a reaction should be thought of as hypotheses to be experimentally tested. Three possible mechanisms are illustrated in Figure 5. The first scheme incorporates covalent modification of one or both of the proteins A and B after their interaction, thereby making available the complementary site for C. As depicted, proteolysis of A catalyzed by the association of A and B could uncover the reactive site for C. Any other type of covalent modification of either structural protein would serve as well as the one depicted. In the second scheme, the complementary reactive site for C is found separated as two half-sites, each half on A and B, respectively. By their interaction to form the AB complex, A and B bring together these half-sites to generate the complementary region with which C reacts. In this scheme, C cannot form a stable complex with either A or B as separate molecules, though C may interact transiently with A or B in a nonproductive manner. In the third scheme, neither A nor B display any portion of their structure as a complementary site for C; however, after or during their assembly they interact in such a way that a conformational rearrangement or refolding of one or both proteins occurs, generating or exposing the complementary site for C. Wood (1980) coined the name "heterocooperativity" for this mechanism. In contrast to the second scheme, C would not be capable of transient, nonproductive interactions with either A or B, because neither protein displays a part of the complementary site for C until after their assembly.

**FIGURE 5** Three schematic models (I, II, III) for molecular sorting of three proteins, A, B, and C, into an ordered reaction sequence. All models depict the assembly of an ABC complex through the ordered pathway  $A + B \rightarrow AB$ ,  $AB + C \rightarrow ABC$ , where the complementary assembly site for subunit C (CSC) is formed after the AB complex. Model I: formation of CSC by covalent modification (proteolysis) of subunit A catalyzed by its assembly with subunit B. Model II: formation of the CSC from two half-sites found on A and B. Model III: formation of the complementary site for C by conformational rearrangement of subunits A and B after they have formed the AB complex, i.e., "heterocooperative" interaction (Wood, 1980). 1 and 2 refer to the first and second assembly steps; italicized letters label the complexes; dotted lines surround the CSC; 1/2 denotes a region that contributes to the CSC.



Each one of these hypothetical assembly schemes accomplishes the ordering of the interactions of A, B, and C. In addition, if the assembly of C into the ABC complex also utilizes one of these proposed schemes, then the reaction of a fourth molecule could be ordered, and so on.

### ***Experimental Approaches for Distinguishing Assembly Schemes***

Experimental verification of any of these schemes will most certainly require the purification of assembly-competent proteins and the careful analysis of assembly reactions carried out *in vitro* using these purified components. Considerable effort has gone into such an endeavor in my laboratory, as we have focused on the initial assembly reactions in the T4 wedge pathway. We have purified the first four structural proteins that interact with each other in the wedge pathway (Plishker et al., 1983; Plishker and Berget, 1984; unpublished results). The protein products of genes 11, 10, 7, and 8 have been isolated either from cells infected with T4 carrying mutations that prevent wedge assembly or from cells carrying cloned genes that encode these proteins. Thus, these proteins are isolated in a form which I call "assembly-naive"; that is, they have never been assembled into a complex with other wedge proteins. Each of these proteins, which may exist in some multimeric form, is designated as P11, P10, P7, and P8 to differentiate them from the primary gene product (gp11, gp10, etc.). The polypeptide chains in these "assembly-naive" proteins have the same molecular weight (within the confidence limits of SDS gel electrophoresis) as the assembled forms found in mature phage or completed baseplates. This result is in accordance with the observations made by Laemmli (1970) and seemingly rules out all but the most subtle involvement of the covalent modification scheme described above as a possible control mechanism for the ordered interaction of these structural proteins. Although substantial proteolytic modification of structural proteins occurs in the morphogenesis of the T4 head subassembly, none seems to occur in the morphogenesis of the wedge.

We have developed a native gel-electrophoresis assay that detects the formation of complexes between the purified wedge proteins (Plishker et al., 1983). Our preliminary *in vitro* assembly experiments have shown that the order of interaction predicted from the assembly pathway proposed by Kikuchi and King (1975) and shown in Figure 4 can be verified using purified proteins. Using this assay system, we have demonstrated the *in vitro* formation of the following complexes: (10/11), (10/7), (10/11/7), (10/7/8), and (10/11/7/8). No interactions have been detected between P11 and any other protein other than P10. P10 does not interact to form a complex with P8 except in the presence of P7, again as predicted from the assembly pathway. However, P8 does interact inefficiently with P7 to form a complex that can be seen in our gel assay system. This interesting result could be taken as support for the second scheme for ordering shown in Figure 5.

By measuring both the affinities of the wedge structural proteins

for each other and the kinetics of the *in vitro* assembly reactions we should be able to distinguish between the second and third schemes described above. Although no definitive confirmation of any of the proposed ordering schemes has yet resulted, the reagents for the analysis of this part of an ordered pathway of assembly determined by protein-protein interaction are at hand.

### CATALYSIS OF PROTEIN-DETERMINED ORDERED ASSEMBLY STEPS

The possibility exists that protein-determined ordered assembly steps might be catalyzed by other proteins. There are numerous examples of proteins that are bound to particular intermediates of an assembly pathway, but are not found in the final product. In at least one well-documented case, that of the scaffolding protein of phage P22, such a protein is released from the structure in an active form and is reused in subsequent rounds of assembly (King and Casjens, 1974; Chapter 5, this volume). Thus, the P22 scaffolding protein, which is required for proper head assembly, can be considered to catalyze P22 head assembly. Because the scaffolding protein is found as a stably bound component of an intermediate in the head-assembly pathway, it is possible to view its addition and release from head-assembly intermediate structures as two separate protein-determined ordered steps of the kind discussed above (fig. 5). In this sense, scaffolding protein does not catalyze an individual assembly step.

Two T4 tail-fiber assembly proteins, gp38 and gp63, are also candidates for such "assembly enzymes". Gp38 is required for proper assembly of the distal portion of the tail fiber, but it is not found stably bound to the components of that structure. This distal fiber is a dimer of gp37 in which the two molecules of gp37 are likely to be interwound (Earnshaw et al., 1978). It has been speculated that gp38 forms a template for the initiation of the proper interaction between the two (possibly unfolded) gp37 polypeptides (Wood, 1979; Wood, 1980). In the reaction promoted by gp63, the completed tail fiber is joined to the otherwise completed phage particle. In this case, the components of the reaction (fibers and fiberless phage) are likely to be fully folded. The reaction is somewhat more complex than might have been expected in that the fiber appears to have to interact with two separated sites on the fiber-less phage, its attachment site on the edge of the baseplate and the whisker (a fiber extending from the tail-proximal portion of the head) (Coombs and Eiserling, 1977; Wood and Conely, 1979; Terzaghi et al., 1979). The mechanism by which gp63 catalyzes tail-fiber attachment is unknown. However, Wood (1979) has speculated that it may help one of the components to occupy transiently an energetically unfavorable conformation needed for assembly. On the other hand, it is possible that gp63 forms a "template" that transiently binds the two components in relative orientations that allow their rapid association.

Thus, both gp38 and gp63 are excellent candidates for "enzymes" that catalyze the noncovalent association of other protein molecules; however, we do not yet understand the mechanism by which they operate or the reason why these particular steps require catalysis.

**FURTHER COMMENTS ON PROTEIN-DETERMINED ORDERED ASSEMBLY MECHANISMS'**

The building of the T4 tail also shows that ordered assembly occurs during polymerization of a large number of identical protein subunits into a supramolecular structure. After the baseplate is completed, 144 molecules of gp19 are added to form the tail tube, a long hollow structure around which the tail sheath assembles (see Berget and King, 1983). During the assembly of the tail tube, after the first annulus of six gp19 subunits is bound to the baseplate, subsequent subunits use previously incorporated subunits as their active site of polymerization. This process repeats until 144 molecules of gp19 are bound. The addition of gp19 monomers stops at this point. It is likely that this stop is controlled by the T4 baseplate protein gp48, which may act as a "tape measure" for determination of the tail tube length (Duda and Eiserling, 1982; Chapter 5, this volume). Assembly-naive P19 has been purified, and it shows no tendency to self-aggregate in the absence of baseplates (Wagenknecht and Bloomfield, 1977). Thus, it is clear that the assembly of multimers of identical subunits can be controlled in a manner analogous to that discussed above for nonidentical subunits, in that after each subunit binds to the growing structure, it creates a binding site for another similar subunit (in Figure 5, subunits A, B, and C would be identical). This type of controlled assembly of multimers from identical subunits has been discussed by King (1980) and Caspar (1980) and called "self-regulated" and "autosterically controlled" assembly, respectively, by these authors.

Assembly of tobacco mosaic virus (see Chapter 3, this volume) also exhibits "self-regulated" assembly of its single species of coat protein. In this case, atomic resolution structures have been determined for the helical virus and for the "free" coat protein (in a nonhelical assembly without RNA), so a detailed model for changes in the coat protein during the assembly process can be constructed. In this case, the coat-protein monomer by itself does show some self-affinity and is found in fairly small oligomers or double ring structures, called disks, containing 34 monomers. However, in the presence of TMV RNA the 34-mer appears to bind to a particular internal location on the RNA, and additional coat-protein molecules (probably in the form of small oligomers) add to the structure in an ordered fashion in both directions (see Chapter 3, this volume for discussion of the initiation process). The "elongation" phase of this assembly is an excellent example of ordered assembly, in which the coat protein requires both the RNA and previously assembled coat protein to act as a site upon which to bind.

The atomic resolution structure of the TMV coat protein in the 34-mer and in the virion, where it is bound to RNA, shows that several alpha helices are very flexible in the 34-mer but are rigidly positioned when bound to RNA (Stubbs et al., 1977; Champness et al., 1976; Bloomer et al., 1978). In addition, there appear to be very slight changes in the relative positions of some helices with respect to one another (Stubbs, 1984). It is not known whether the ordering of some helices or the shifting of overall position of some helices is

responsible for the formation of the binding site for additional coat protein; however, it certainly shows that some type of conformational change (as in model 3, Figure 5) in TMV coat protein might be responsible for "self-regulated" assembly.

Several nonviral homopolymerization systems clearly appear to utilize this type of control of assembly. *Salmonella typhimurium* flagellin appears to assemble in a similar "self-regulated" manner. In this case, circular dichroism measurements have documented a conformational change upon assembly (Uratani et al., 1972). The kinetics of flagellin assembly suggest that this conformational change after each subunit assembles is rate-limiting for the assembly process under laboratory conditions. Similar arguments have been made for actin polymerization (Wegner, 1976) and microtubule assembly (Erickson, 1974; Kirschner and Williams, 1974).

The molecular nature of the types of conformational changes that might occur within protein subunits upon their assembly have been discussed by Rossmann and colleagues (Chapter 2, this volume; Rossmann, 1984). In spite of the small changes in helix packing observed upon assembly of the TMV coat protein into virions, they argue rather persuasively that most protein structures cannot undergo major "elastic deformations", and they have catalogued the types of differences that have been seen between atomic resolution structures of free proteins and the same proteins bound to some ligand (small molecule, another protein, or nucleic acid). These are (1) order-disorder changes, (2) hinge movements, (3) conformational changes in loop or beta-bend regions and (4) alteration of binding surfaces between one subunit and another. In addition, very slight alterations in the overall folding would alter the positions of surface amino acid side chains, thereby changing the surface properties of the protein (Harrison, 1983). Presumably any of these, and perhaps other types of conformational changes not yet documented, could be responsible for the creation of the binding site for protein C in Figure 5, thus controlling the assembly process.

## CONCLUSION

For those who have studied or worked on the morphogenesis of bacteriophages, the concept of assembly pathways seems natural. It has always seemed that pathways of assembly must represent the most efficient mechanism of building a structure. We have all seen examples of the assembly-line concept in day-to-day life, from the way cars are built to the way buildings are constructed. As I have described in this review, the simplest of probability calculations bears out this feeling.

There is a wealth of descriptions of morphological pathways in the scientific literature. Most of the construction problems faced by organisms are dealt with by using pathways of assembly. However, a description on the molecular level of how pathways are controlled is blatantly missing from our knowledge. This then represents the challenge to investigators studying morphogenesis to determine the molecular basis for sorting protein molecules into a reaction sequence.

## ACKNOWLEDGEMENTS

I wish to thank my co-workers Mary F. Plishker, Monjula Chidambaram, Shaukat H. Rangwala, Karen Pollack, Roger Kylberg, Colleen Sari, John Schwarz, and Harryette Davis for their hard work, cheerfulness, and patience during the course of our studies and in the preparation of this article.

The preparation of this chapter was supported by Grant No. PCM 83-14241 and PCM 81-04523 from the National Science Foundation and GM28952 from the National Institutes of General Medical Sciences.

## REFERENCES

- Berget, P. B. and H. R. Warner. 1975. *J. Virol.*, 16, 1669.
- Berget, P. B. and J. King. 1978. *J. Mol. Biol.*, 124, 469.
- Berget, P. B. and J. King. 1983. In *Bacteriophage T4*. C. K. Mathews, E. M. Kutter, G. Mosig, and P. B. Berget, eds. p. 246. American Society of Microbiology, Washington, D. C.
- Bloomer, A., J. Champness, G. Bricogne, R. Staden, and A. Klug. 1978. *Nature*, 276, 362.
- Caspar, D. L. D. 1980. *Biophys. J.*, 32, 103.
- Champness, J. H., A. C. Bloomer, G. Bricogne, P.J. G. Butler, and A. Klug. 1976. *Nature*, 259, 20.
- Coombs, D. and F. A. Eiserling. 1977. *J. Mol. Biol.*, 116, 375.
- Crane, H. R. 1950. *The Scientific Mon.*, 70, 376.
- Drake, J. W. 1969. *Nature*, 221, 1132.
- Duda, R. and F. A. Eiserling. 1982. *J. Virol.*, 43, 714.
- Earnshaw, W. C., E. B. Goldberg, and R. A. Crowther. 1979. *J. Mol. Biol.*, 132, 101.
- Edgar, R. S. 1969. *Harvey Lect.*, 63, 263.
- Edgar, R. S. and I. Leilausis. 1968. *J. Mol. Biol.*, 32, 263.
- Edgar, R. S. and W. B. Wood. 1966. *Proc. Nat. Acad. Sci. (USA)*, 55, 498.
- Epstein, R. H., A. Bolle, C. M. Steinberg, E. Kellenberger, E. Boy de la Tour, R. Chevalley, R. S. Edgar, M. Sussman, G. H. Denhardt, and I. Leilausis. 1963. *Cold Spring Harbor Symp. Quant. Biol.*, 28, 375.

- Erickson, H. P. 1974. *J. Supramol. Struct.*, 2, 393.
- Harrison, S. C. 1983. *Adv. Virus Res.*, 28, 175.
- Kikuchi, Y. and J. King. 1975a. *J. Mol. Biol.*, 99, 645.
- Kikuchi, Y. and J. King. 1975b. *J. Mol. Biol.*, 99, 673.
- Kikuchi, Y. and J. King. 1975c. *J. Mol. Biol.*, 99, 695.
- King, J. 1971. *J. Mol. Biol.*, 58, 693.
- King, J. 1980. In *Biological Regulation and Development*. Volume 2. R. Goldberg, ed. p. 101. Plenum.
- King, J. and S. R. Casjens. 1974. *Nature*, 251, 112.
- King, J. and N. Mykolajewycz. 1973. *J. Mol. Biol.*, 75, 339.
- King, J. and W. B. Wood. 1969. *J. Mol. Biol.*, 39, 583.
- Kirschner, M. W. and R. C. Williams. 1974. *J. Supramol. Struct.*, 2, 412.
- Kozloff, L. M. 1981. In *Bacteriophage Assembly*. M. DuBow, ed. p. 327. Alan R. Liss.
- Laemmli, U. K. 1970. *Nature*, 227, 680.
- Meezan, E. and W. B. Wood. 1971. *J. Mol. Biol.*, 58, 685.
- Mosher, R. A. and C. K. Mathews. 1979. *J. Virol.*, 31, 94.
- Plishker, M. F., M. Chidambaram, and P. B. Berget. 1983. *J. Mol. Biol.*, 170, 119.
- Plishker, M. F. and P. B. Berget. 1984. *J. Mol. Biol.*, 178, 699.
- Rossmann, M.G. 1984. *Virol.*, 134, 1.
- Stubbs, G. 1984. In *Biological Macromolecules and Assemblies*, Vol. 1. F. A. Journak and A. McPherson eds. p. 149. Wiley.
- Stubbs, G., S. Warren, and K. Holmes. 1977. *Nature*, 267, 216.
- Terzaghi, B. E., E. Terzaghi, and D. Coombs. 1979. *J. Mol. Biol.*, 127, 1.
- Uratani, Y., S. Asakura, and K. Imahori. 1972. *J. Mol. Biol.*, 76, 85.
- Wagenknecht, T. and V. A. Bloomfield. 1977. *J. Mol. Biol.*, 116, 347.

Wegner, A. 1976. *J. Mol. Biol.*, 108, 139.

Wood, W. B. 1979. *Harvey Lect.*, 73, 203.

Wood, W. B. 1980. *Quart. Rev. Biol.*, 55, 353

Wood, W. B. and M. P. Conley. 1979. *J. Mol. Biol.*, 127, 15.

Wood, W. B., R. S. Edgar, M. Henninger, and I. Lielausis. 1968. *Fed. Proc.*, 27, 1160.