The DNA Enzymology of Protein Machines

B.M. ALBERTS

Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143

When I was asked to prepare this introduction, I very nearly declined. Being not nearly as wise as Alan Campbell, I knew that I could not attempt to match his scholarly introduction to the 1980 Cold Spring Harbor Symposium, the last Symposium to deal with these issues. Also, unlike him, my own contributions to the area of genetic recombination have been quite modest. I completely missed out on the two most striking enzymological discoveries in the genetic recombination field: in vitro, site-specific genetic recombination catalyzed by the λ integrase protein (Nash 1981), and the synapsis step of general recombination catalyzed by the *Escherichia coli* RecA protein (Radding 1982). Moreover, although I am a supposed expert in DNA replication, it took me a long time to realize the general significance of the seminal finding made by Ahmad Bukhari here at Cold Spring Harbor that the transposition of bacteriophage Mu involves a site-specific DNA replication process (Ljungquist and Bukhari 1977). As discussed in one of the most exciting talks given at this Symposium, Bukhari's pioneering work has now reached its logical conclusion with the reconstruction of the Mu transposition process in vitro (Mizuuchi et al., this volume). It is therefore especially fitting that this 49th Cold Spring Harbor Symposium is dedicated to his memory.

My own laboratory has spent a great deal of time characterizing the DNA replication apparatus of bacteriophage T4. As most scientists over 40 years old will remember, T4 is a large bacteriophage, and its genes seem to encode all of its own DNA replication and genetic recombination proteins. Over the past 18 years, I have received a general education in DNA enzymology from our continuing struggles with this system. I agreed to accept the challenge of preparing this introduction because I believe that some of the lessons we have learned will be helpful to others struggling with the enzymology of less well understood genetic systems. Thus, I deal here with general issues rather than with details.

Protein Machines

We have been surprised in recent years to realize that the T4 replication fork is moved by a multienzyme complex that can quite reasonably be viewed as a true "replication machine." As far as is now known, this replication machine is quite similar to the more complicated one used by *E. coli* (Kornberg 1980, 1984), as well as to the simpler apparatus used by bacteriophage T7 (Richardson 1983). In spite of these similarities,

the replication machines of the three organisms are largely composed of protein parts that are not interchangeable; e.g., in our system the T4 DNA polymerase cannot be replaced by either the T7 or the *E. coli* DNA polymerase. Yet, when present alone, each of these three DNA polymerases will synthesize a DNA chain in exactly the same way. The reason for this noninterchangeability of parts has become clear only recently: The T4 replication machine is a sophisticated DNA-handling apparatus, whose numerous protein parts move relative to each other as the replication fork moves, without disassembling. To expect the protein parts of two such machines to be interchangeable is like expecting to be able to interchange any two parts of two different automobiles.

One major lesson that can be derived from considering the elegance and efficiency of the T4 DNA replication machine is that nearly all of us have been greatly underestimating what can be accomplished by groups of proteins that work together on DNA. In retrospect, it seems obvious that complex living systems could not exist without the "high technology" of multicomponent protein machines. Underlying this technology are the ordered conformational changes in protein molecules that are caused by the hydrolysis of bound nucleoside triphosphate molecules. These ordered conformational changes enable complexes of proteins to move in coherent ways, freeing the biological process from the necessity of relying on random motions. One can confidently predict that essentially all of the major processes that occur in cells, including, of course, genetic recombination, will make use of such highly sophisticated machines. A major theme in this introduction is that acceptance of this view can profoundly change the way that one approaches a research problem, producing special opportunities, as well as special difficulties.

To make the above point, I will treat our investigation of the T4 bacteriophage DNA replication system as an experimental "case study" with implications for work in other systems. My emphasis will be on experiments, realizing that the prevalence of protein machines makes theoretical biology an especially precarious occupation. In fact, we know that proteins can do such fancy things on DNA that an enormous number of models are possible for almost any genetic process. Moreover, it is quite sobering to look back at how we have stumbled our way to an understanding of the DNA replication fork. The first discovery of discontinuous DNA synthesis was a real surprise, as was the later finding of RNA primers. Only in retrospect

did we come to realize that the asymmetric replication fork is elegantly designed to ensure that all of the DNA on both strands is synthesized with a very high fidelity (Alberts and Sternglanz 1977). It seems that one would need to learn how to think like a cell in order to be successful at predicting mechanisms. Among other things, this means that one would be required to keep in mind the many constraints that come from the finite error rates in all biological processes, while assuring that the many different events that occur simultaneously on a DNA molecule are compatible with each other. In the vast majority of cases, we are too ignorant to make useful predictions. As in the past, we will therefore have to stumble forward in the dark, conscious that we are exploring complex intracellular landscapes with a rather small lantern.

Because the T4 DNA replication apparatus will serve as a model for much of the discussion that follows, I need to begin by briefly outlining some of its central features.

The T4 DNA Replication Apparatus

By purifying the proteins corresponding to seven T4 replication genes (32, 41, 43, 44, 45, 61, and 62) and then mixing various combinations of these proteins with different types of DNA templates, we and others have been able to reconstruct the general mechanism of T4 replication fork movement (for review, see Nossal and Alberts 1983). A two-dimensional view of the T4 replication fork is shown in Figure 1. Only its most salient features will be outlined here:

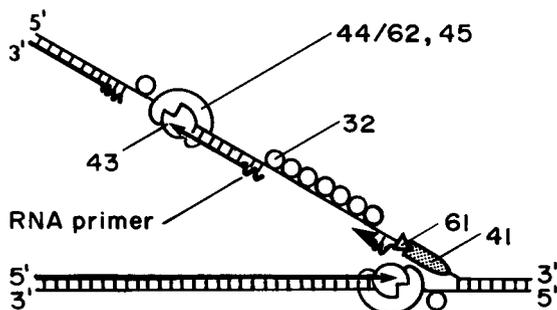1. There are two DNA polymerase molecules working

Figure 1. A two-dimensional view of the DNA replication fork of bacteriophage T4. The T4 DNA polymerase is the product of gene 43, and the gene-44/62 and -45 proteins that function with it are designated as "polymerase accessory proteins." The gene-32 protein binds to all of the single-stranded DNA at the fork; it is a helix-destabilizing protein, sometimes described as a "single-strand binding protein." The gene-41 protein is both a DNA helicase and a protein that is required, together with the gene-61 protein, for RNA primer synthesis on the lagging strand. The average length of the Okazaki fragments formed on the lagging strand is about 1500 nucleotides, both in vivo and in vitro.

at any given time, one on the leading strand and one on the lagging strand. In a process that requires periodic ATP hydrolysis, a complex of three polymerase accessory proteins — the products of T4 genes 44, 62, and 45 — acts to clamp down each DNA polymerase molecule, thereby converting it into a highly processive enzyme (Huang et al. 1981). These three proteins are thought to form a complex with the DNA polymerase (the product of gene 43) that creates a unit analogous to the more complicated E. coli DNA polymerase III holoenzyme with its seven different subunits (McHenry and Kornberg 1981).

2. In front of the T4 DNA polymerase unit on the leading strand, the DNA double helix is rapidly unwound into its two single strands by the combined actions of the DNA polymerase, the T4 gene-32 protein (a helix-destabilizing protein that binds tightly and cooperatively to all single-stranded DNA) and the T4 gene-41 protein (a DNA helicase on the lagging strand that uses GTP hydrolysis energy to force its way into the template helix). This unwinding normally proceeds at a rate of about 500 nucleotides per second, which is more than a million times faster than the estimated rate of spontaneous helix opening under physiological conditions (Alberts et al. 1980, 1983; Venkatesan et al. 1982).

Once positioned on the lagging strand, each functional unit of the 41 protein (an oligomer) remains with the same fork for a very long time; inside the cell, this unit of the 41 protein probably enters each fork only when the fork forms at an origin of replication, persisting until the fork finishes its DNA synthesis.

3. The lagging-strand template is a DNA single strand covered with the gene-32 protein. This strand provides sites at which a new, RNA-primed Okazaki DNA fragment is initiated about once every 4 sec. Each RNA primer is a pentaribonucleotide that is synthesized by a complex of the gene-61 and gene-41 proteins on the lagging strand and then elongated by the DNA polymerase unit on this strand (Liu and Alberts 1980; Nossal 1980). Thus, in addition to acting as a DNA helicase, the 41 protein plays a second key role at the fork: It forms a mobile site at which the 61 protein binds to help synthesize RNA primers. Since the 41 protein is only bound to the DNA at a replication fork, this mechanism presumably serves to prevent RNA primer synthesis elsewhere in the cell, thus allowing other regions of single-stranded DNA to remain single-stranded.

4. Like the DNA polymerase molecule on the leading strand, the molecule of DNA polymerase on the lagging strand can be shown to remain with its replication fork for a prolonged period of time, recycling every 4 sec or so to start a new Okazaki fragment on the lagging strand. Therefore, the lagging strand must be folded so as to bring the 3'-hydroxyl end of a completed Okazaki fragment adjacent to the start site for the next Okazaki fragment, as in
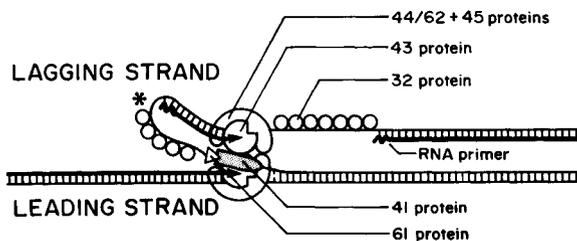
Figure 2. A schematic view of the seven proteins of the T4 DNA replication machine as they are thought to exist in an actual replication fork. The two-dimensional replication fork has been converted into the structure shown by folding the DNA on the lagging strand in such a way as to bring the DNA polymerase molecule on the lagging strand in close opposition to the DNA polymerase molecule on the leading strand. The lagging-strand DNA polymerase molecule is thereby held to the other replication proteins, allowing it to be retained for many successive cycles of Okazaki fragment synthesis, as shown in Fig. 3.

the example shown in Figures 2 and 3. The synthesis of a new RNA primer seems to be delayed until the preceding Okazaki fragment has been completed, because nearly every RNA primer that is synthesized in vitro at a fork starts a new DNA chain (Liu and Alberts 1980). The model for replication fork movement in Figure 3 therefore not only explains why the pattern of in vitro DNA synthesis on the lagging strand is unchanged by extreme polymerase dilutions (Alberts et al. 1983), but also why very few Okazaki fragments that are shorter than 500 nucleotides are made in this system, despite the fact that potential start sites for primers are quite frequent.

The lagging-strand DNA polymerase molecule moves rapidly ahead with the fork (Fig. 3), and in in vitro reactions the T4 replication apparatus lays down a series of unsealed Okazaki fragments on the lagging strand, which still retain intact RNA primers at their 5' ends. It therefore seems likely that, inside the cell, these RNA primers are removed and replaced by DNA in a separate DNA repair reaction, which is uncoupled from the replication fork itself and involves additional proteins.

In summary, the basic T4 replication apparatus consists of a moving complex of seven polypeptide chains, in which the entire DNA replication fork is embedded. The proteins have a total mass of at least 900,000 daltons, not including about 5×10^6 daltons of 32 protein that are bound to the regions of single-stranded DNA at a fork. As a group, the proteins at the fork proceed unidirectionally along the DNA at a rate of about 500 nucleotides/sec, faithfully replicating both strands of the template DNA helix as they go. In spirit, the T4 DNA replication apparatus is like a tiny sewing machine, composed of protein parts and powered by several different types of nucleoside triphosphate hydrolyses.

The T4 DNA Polymerase Can Be Incorporated into Different Types of Replication Complexes

Inside the cell infected with a T4 bacteriophage, both DNA replication and DNA repair utilize the same DNA polymerase molecule, the product of T4 gene 43 (Bernstein and Wallace 1983). Yet DNA replication and DNA repair are very different types of processes. Much of the DNA synthesis in DNA repair is thought to involve simple gap filling by the DNA polymerase, where it is important to avoid the type of DNA strand-displacement reaction that can generate a complete fork, including lagging-strand DNA synthesis. This fact becomes most obvious when considering the most frequent DNA repair event in the cell, the resealing of adjacent Okazaki fragments on the lagging strand. If even a small fraction of these repair reactions continue too far and create a replication fork, innumerable new forks would form on each lagging strand, making the DNA replication process hopelessly complex.

Figure 4 illustrates in a schematic way how different types of complexes containing the T4 DNA polymerase are likely to be formed at different sites on the same DNA molecule. In this view, only those special DNA sequences that direct the assembly of the 41 protein onto a DNA strand are replication origins (Fig. 4A). In principle, the different replication origins on the T4 chromosome (Kozinski 1983; Mosig 1983) could assemble somewhat different types of replication forks, depending on exactly which additional proteins are coassembled with the DNA polymerase and the 41 protein onto the DNA. Other special sites on the DNA might allow the DNA polymerase to become associated, not with the 41 protein, but with a different DNA helicase that is unable to participate in RNA primer synthesis (e.g., the T4 dda protein [Krell et al. 1979; Jongeneel et al. 1984]). DNA synthesis at these sites would generate "whiskers" of single-stranded DNA (Fig. 4B) that could either participate in the initiation of genetic recombination processes or be degraded in association with a "long-patch" type of DNA repair. Finally, other types of sites on the DNA, including each junction between a pair of unsealed Okazaki fragments, will bind a DNA polymerase complex that is prevented from using any type of DNA helicase (Fig. 4C). This type of DNA polymerase complex should be unable to carry out any strand-displacement DNA synthesis inside the cell and should instead fall off of the DNA after completing a "short-patch" type of DNA repair.

The above example provides one illustration of how a single protein (the T4 DNA polymerase) can participate in several different types of discrete protein machines. Perhaps the best studied case is that of the T4 gene-32 protein, which plays major roles in DNA replication, DNA repair, and genetic recombination (Alberts and Frey 1970). This helix-destabilizing protein has been shown to interact with a large number of different T4 proteins by both genetic (Mosig et al. 1979) and biochemical (Formosa et al. 1983) methods. These
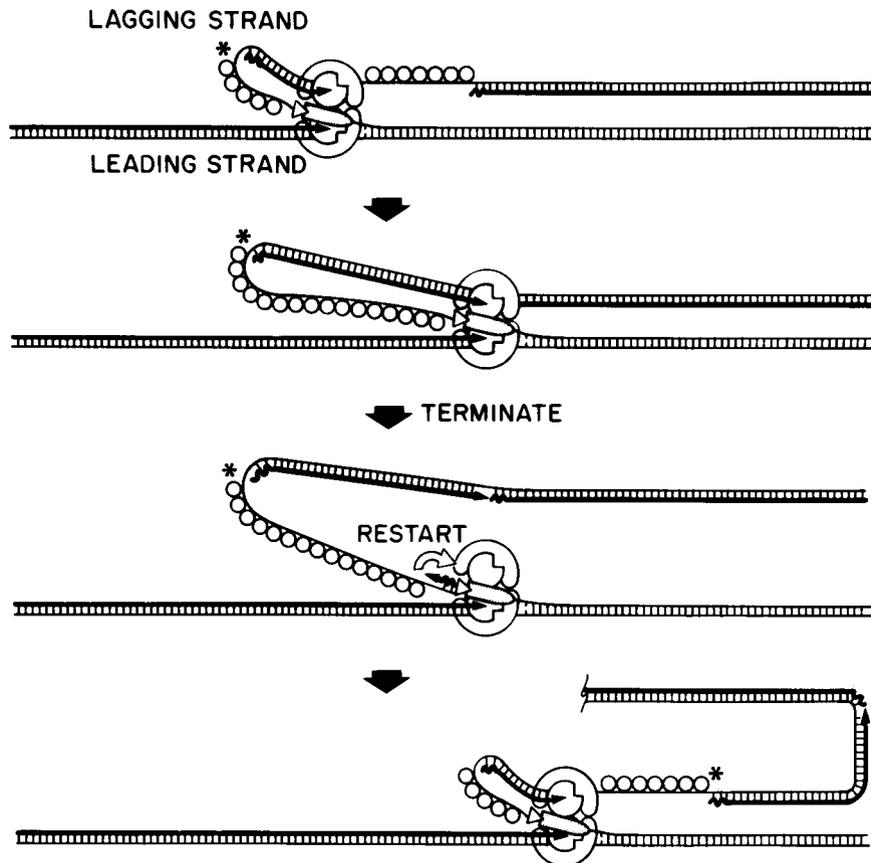
ALBERTS

LAGGING STRAND

LEADING STRAND

TERMINATE

RESTART

Figure 3. A model for DNA synthesis at a replication fork that allows the same DNA polymerase molecule to be repeatedly recycled to synthesize all of the Okazaki fragments on the lagging strand. The crucial step here occurs when the synthesis of each Okazaki fragment is completed. Rather than falling off of the entire fork assembly, the DNA polymerase molecule on the lagging strand remains bound to the assembly of replication proteins, letting go only of the DNA on the lagging strand. For the different symbols used to represent the T4 replication proteins, see Figs. 1 and 2.

interactions are thought to reflect alternate associations that the *32* protein undergoes in a variety of different multiprotein complexes.

Genetic Analyses Are Necessary but Not Sufficient

The seven T4 DNA replication proteins that move the fork in vitro were initially shown to be essential for T4 DNA replication in vivo by a genetic analysis that identified each one with a conditional-lethal mutant that is severely deficient in its DNA synthesis (Epstein et al. 1964; Warner and Hobbs 1967). Four of these seven proteins (products of genes *44, 62, 45,* and *41*) were initially purified by exploiting an in vitro complementation assay utilizing a cell lysate made from the respective type of mutant-infected cells (Barry and Alberts 1972; Morris et al. 1979a,b). Thus, the extensive genetic analysis of bacteriophage T4 begun by R. Edgar, R. Epstein, and their colleagues was absolutely essential for carrying out biochemical studies of the T4 replication process.

Genetics cannot be expected to do everything, however. For example, it has not been possible to learn anything about the precise role that each replication protein plays at the fork from detailed studies of the type of abortive replication observed when that protein is defective. Such information has only been gained from detailed in vitro studies of the reactions catalyzed by various combinations of the purified proteins on specifically modified DNA templates (see below). In addition, our recent in vitro studies have convinced us that a function carried out by a second DNA helicase (in addition to the gene-*41* protein), the T4 *dda* gene product, is essential for DNA replication inside the cell. This protein becomes absolutely necessary for in vitro DNA replication when RNA polymerase molecules are bound to the DNA template (Bedinger et al. 1983). Yet, the T4 *dda* protein can be removed by a mutation without seriously affecting T4 DNA replication in vivo (Behme and Ebisuzaki 1975). We predict therefore that there is a second T4 (or host) protein that can substitute for the *dda* protein when it is missing. In addition, possibly because T4 contains redundant sets of replication origins (Kozinski 1983; Mosig 1983), we still lack mutants in some of the functions that are thought to be needed for the initiation of replication forks at a T4 origin of replication, a process that has not yet been reconstructed in vitro.
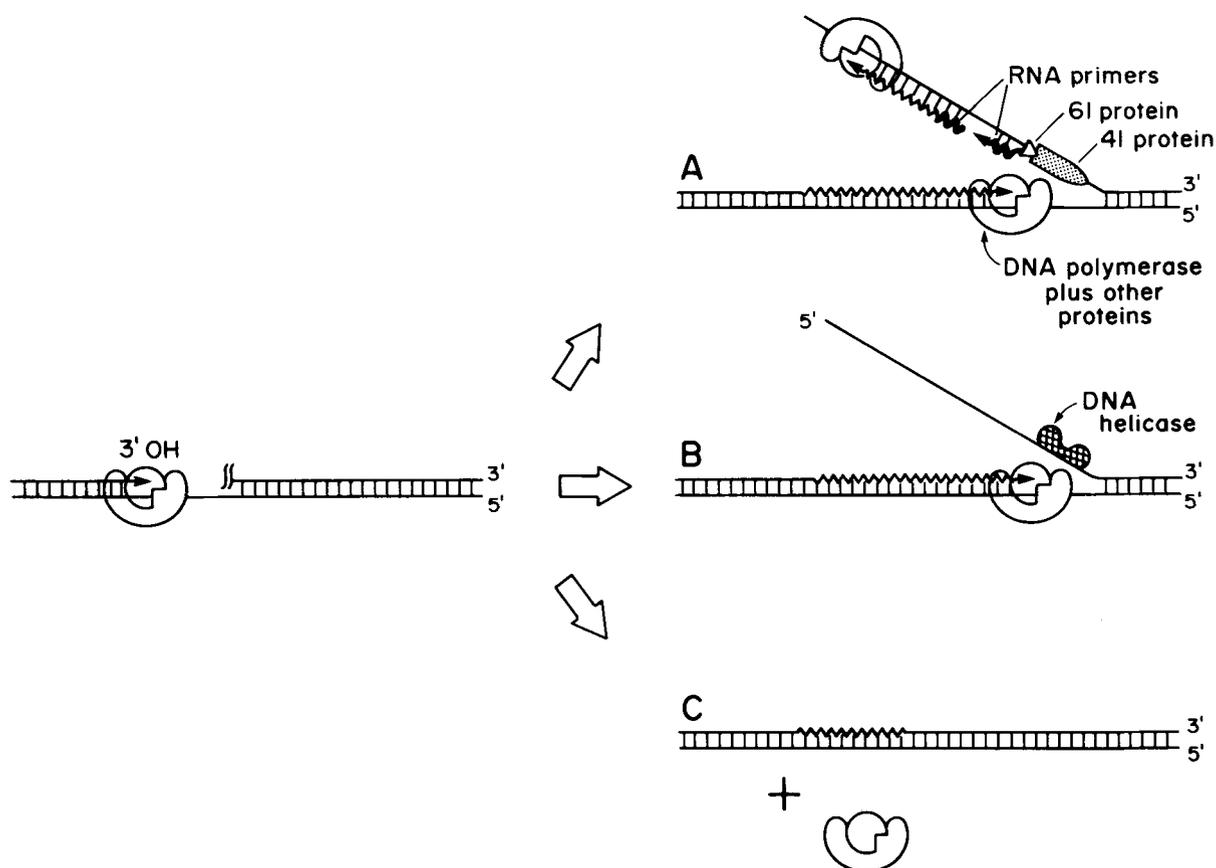
Figure 4. Diagram illustrating the results of assembling the T4 DNA polymerase into different replication protein complexes at different sites on the DNA. (*A*) A specific DNA sequence corresponding to a DNA replication origin directs the assembly of the T4 gene-*41* protein onto the lagging strand. As a result, a true replication fork is established with both leading- and lagging-strand DNA synthesis. (*B*) Another type of specific DNA sequence (or structure) is proposed to have allowed a DNA polymerase molecule to coassemble into a protein complex that contains a DNA helicase different from the gene-*41* protein. Because this DNA helicase is unable to participate in RNA primer synthesis, the DNA on the lagging strand remains single-stranded (see text). (*C*) A typical DNA repair reaction is shown, in which no DNA helicase is involved.

Proteins with redundant functions are not the only ones that are difficult to discover by a genetic analysis. All of us, geneticists and biochemists alike, tend to ignore small proteins. Biochemists usually electrophorese proteins of 10,000–15,000 daltons or less at the front of their SDS-polyacrylamide gels. Yet, by using two-dimensional, gradient-gel electrophoresis, Junko Hosoda has shown that many small proteins are synthesized from the T4 genome (for review, see Burke et al. 1983). Nearly all of these small proteins are still unidentified, no doubt because one rarely finds a mutant in a gene for a small protein, in view of its small target size.

Our DNA sequence analysis has revealed that there are unidentified reading frames for five small proteins clustered between T4 genes *41* and *61* (Fig. 5). These reading frames are also close to a primary origin of DNA replication that has been identified by Gisela Mosig and her collaborators on the other side of gene *61* (Mosig 1983). Because of the striking clustering of the known genes that synthesize interacting proteins on the T4 genome (Wood and Revel 1976), we suspect

that proteins encoded by one or more of the unidentified reading frames shown in Figure 5 play an important role in T4 DNA replication. In principle, recombinant DNA technology should make possible the identification, overproduction, and isolation of all of these proteins for in vitro studies. Likewise, adaptation to T4 of the in vitro mutagenesis techniques used in other organisms (Scherer and Davis 1979; Winston et al. 1983) should greatly facilitate obtaining T4 bacteriophages that are specifically mutated in these genes.

One Can Spend Too Much Clean Thought on Clean Proteins If One Does Not Have All the Parts

The T4 replication machine is constructed from proteins that are only rather loosely associated with each other. For example, most of the known protein-protein interactions are too weak to be detected by simple cosedimentation experiments on sucrose gradients, and six of the proteins shown in Figure 1 must be individually isolated from the crude lysate (the exception being the *44/62* protein, which is isolated as a tight complex
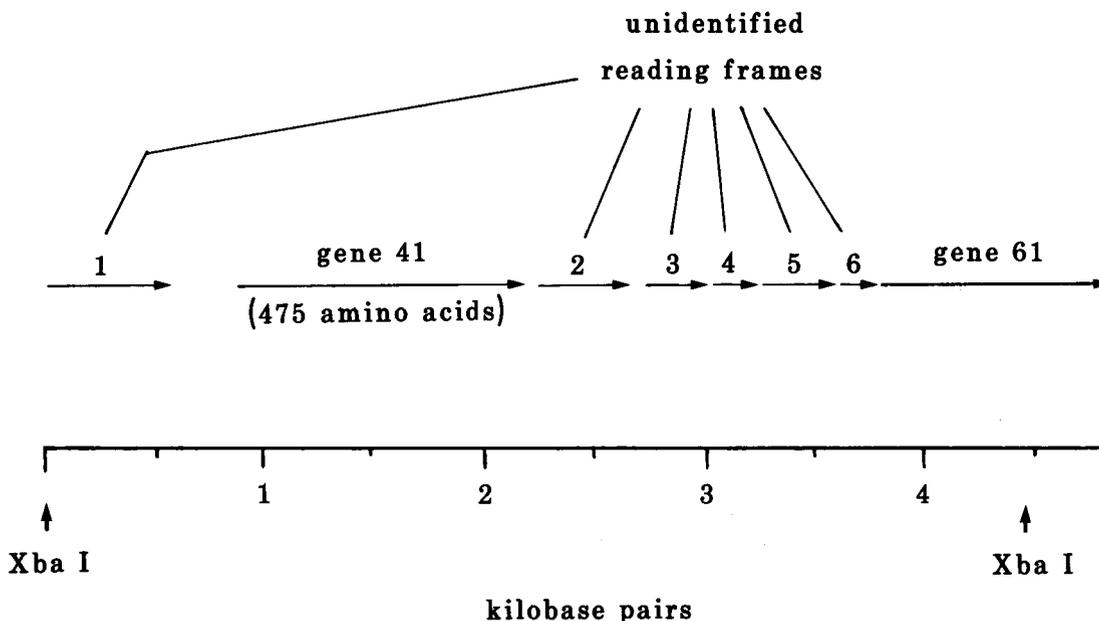
Figure 5. A sequenced early region of the T4 genome that contains five unidentified reading frames located between genes *41* and *61*. This region is suspected of coding for small proteins, some of which interact with either the gene-*41* or the gene-*61* proteins and are involved in T4 DNA synthesis (M. Nakanishi et al., unpubl.).

of two different protein chains). Not all protein machines fall apart in cell lysates in this way (the ribosome being a notable example of an especially stable complex). However, there is no reason to suspect that our case is unusual; intracellular protein concentrations are sufficiently high that even quite weak associations between proteins can be effective (Fulton 1982). Thus, one will generally need to isolate intracellular protein complexes piecewise, eventually reconstructing each of them by mixing together a number of different, highly purified protein subunits. Some of these purified proteins will have no detectable activity by themselves; e.g., this is true for the T4 gene-*45* protein. Other proteins will display only a partial activity that reveals almost nothing about their biological function. For example, the T4 *44/62* protein is a very weak, DNA-dependent ATPase (Mace and Alberts 1984a); only when it is stimulated by the *45* protein does this ATPase activity "come alive," serving to facilitate DNA synthesis by the T4 DNA polymerase (*43* protein).

A more subtle example relates to our initial discovery that the *41* protein plays an important role in leading-strand DNA synthesis, as well as being required for the synthesis of RNA primers on the lagging strand (Liu et al. 1979). This realization was made possible by the development of an assay for measuring the rate of movement of an individual DNA polymerase molecule on a double-helical template, which demonstrated that in vitro fork rates are much slower than the in vivo fork rate unless the *41* protein is added. If one is permitted to extrapolate to other situations, our experience suggests that it is important to attempt to reproduce intracellular reaction rates in in vitro systems, as well as the correct reaction pathway.

It is often quite difficult to know when a protein machine has been well enough reconstructed that its in vitro properties warrant an extensive study. The situation is complicated by the fact that the number of different experiments that can be done in any system that contains three or more components is enormously greater than one's intuitive expectations. For example, varying both the relative and absolute concentrations of three protein components in a mixture and measuring their activity as a function of the concentrations of nucleotide cofactors, Mg^{++}, specific anions and cations, pH, temperature, and so forth could provide enough experiments for a lifetime. The number of different experiments already published with the three-component *E. coli* system — RecA protein, single-strand binding protein (SSB), and DNA — should conclusively prove this point. But, to continue with this arbitrary example, does the genetic recombination machine of *E. coli* also contain other, as yet unisolated protein components that would drastically change many of the results obtained in such experiments? In part, one's experimental approach to this problem depends on how complicated one expects the recombination apparatus to be; my own present bias is to suspect a greater complexity.

Affinity Chromatography Allows the Different Parts of a Protein Machine to Select Each Other Out

To identify some of the protein-protein interactions in the T4 replication apparatus, we have linked the T4 gene-*32, -43, -44/62, -45,* and *-41* proteins to separate agarose matrices and used these matrices (each containing ~1 mg/ml of their respective bound protein) as affinity columns for the chromatography of T4-in-

fected cell extracts. To distinguish the bacteriophage-encoded proteins from the *E. coli* host proteins that are also present in the lysate, the T4 proteins were selectively labeled by the addition of radioactive amino acids to the infected cells. We found that each T4 replication protein column selected a different subset of proteins from the extract that were not bound by a control column of albumin (Alberts et al. 1983). The most striking results were obtained with the column that contained the immobilized gene-*32* protein. As shown in Table 1, the *32* protein bound more than 10 different T4-induced proteins specifically; nearly all of these could be identified as specific T4 gene products that have an important role in DNA replication or genetic recombination (Formosa et al. 1983). In addition, one major host protein also bound to the column. This host protein is unusual in that it also binds to the T4 *uvsX* protein (the T4 RecA protein analog) and to DNA; its function in *E. coli* is unknown (see Formosa and Alberts, this volume).

Experiments like the one described in Table 1 reveal that protein affinity columns are an enormously useful tool for finding the various components of a large protein machine. In principle, any protein can be used to produce such a column, although to make a 0.1-ml column at the high protein concentrations required to detect weak interactions requires at least 100 μg of the protein. (Fortunately, after an experiment, the columns can be rinsed with a buffer containing 50% glycerol and stored unfrozen at −20°C for reuse.) We were encouraged to begin experiments of this type by the results of Dr. Jack Greenblatt, who had made use of similar protein affinity columns in his elegant study of the interaction of λ N protein with RNA polymerase (Greenblatt 1981; see also Ratner 1974).

In principle, a possible alternative to the use of protein affinity columns is the development of special solvents that greatly enhance the strength of protein-protein interactions so as to allow most of the pieces of a protein machine to remain bound to each other during all of the successive chromatographic steps required for their purification. It is not clear that a general solvent of this type is possible, although it would seem reasonable to attempt to exploit the large exclusion volumes of aqueous polymer solutions for this purpose (Al-

bertsson 1960). These polymers have proven quite useful for increasing the effective concentrations of interacting biological macromolecules during enzymatic reactions that occur on DNA (Fuller et al. 1981; Zimmerman and Pheiffer 1983), but I am not aware of experiments that have used them successfully in column buffers.

The Subassemblies of a DNA Replication Machine Have Activities That Are Best Measured on Special DNA Templates

A major breakthrough in our own work came in 1975, when we first mixed all of the T4 replication proteins together with an appropriate DNA template and were able to copy a double-stranded DNA template efficiently (Morris et al. 1975). However, in actuality, that achievement taught us very little about the DNA replication process. With all seven replication proteins present, the reaction was too complex to figure out what each of the proteins was doing. Determination of the role of each T4 replication protein required that "partial reactions" be set up, in which DNA templates that were simpler to replicate than an intact DNA double helix were provided with less than the full set of proteins.

The structures of a few of the various types of DNA templates that we have used are illustrated in Figure 6. Primed, single-stranded DNA molecules (Fig. 6A) were useful for showing that the T4 DNA polymerase prefers DNA single strands coated with *32* protein over the bare DNA as a template (Huberman et al. 1971), as well as for working out details of the ATP-dependent stimulation of the DNA polymerase catalyzed by the polymerase accessory proteins (Piperno and Alberts 1978; Huang et al. 1981; Roth et al. 1982; Mace and Alberts 1984b). Circular DNA single strands (Fig. 6B) were useful for studying the RNA-primed, de novo initiation of DNA chains (Liu and Alberts 1980, 1981; Nossal 1980) since, unlike a linear DNA molecule, they have no free 3'-hydroxyl end that can fold back to prime DNA synthesis (the use of these templates was pioneered by Arthur Kornberg's laboratory, as detailed in an earlier Symposium volume [Kornberg 1979]). Finally, circular, double-stranded bacteriophage fd DNA nicked by the fd gene-*2* protein (Fig. 6C) provides a unique start site for a "rolling circle" DNA replication that has been useful for studying the collision of replication forks with RNA polymerase molecules (Bedinger et al. 1983).

Recently, we have begun to take advantage of the availability of technology for the direct chemical synthesis of up to 80-nucleotide-long DNA chains of any desired sequence (e.g., see Urdea et al. 1983), which enables small DNA molecules of essentially any structure to be designed for use as replication templates. Two types of DNA molecules that are presently being used in an attempt to position the various T4 replication proteins by means of DNA footprinting studies are sketched in Figure 6, D and E (M. Munn and M.

Table 1. Major Protein Species Bound to *32*-Protein Agarose Columns

Protein	Function
T4 *32* protein	helix-destabilizing protein
T4 *43* protein	T4 DNA polymerase
T4 *45* protein	DNA polymerase accessory protein
T4 *uvsX* protein	T4 recombination: RecA protein analog[a]
T4 *uvsY* protein	T4 recombination
T4 *dda* protein	DNA-dependent ATPase: helicase
T4 *46/47* protein	T4 recombination: exonuclease
T4 RNase H	removes RNA from RNA:DNA hybrids
BP-1	unknown; 30,000 daltons; T4-encoded
BP-2	unknown; 32,000 daltons; host-encoded

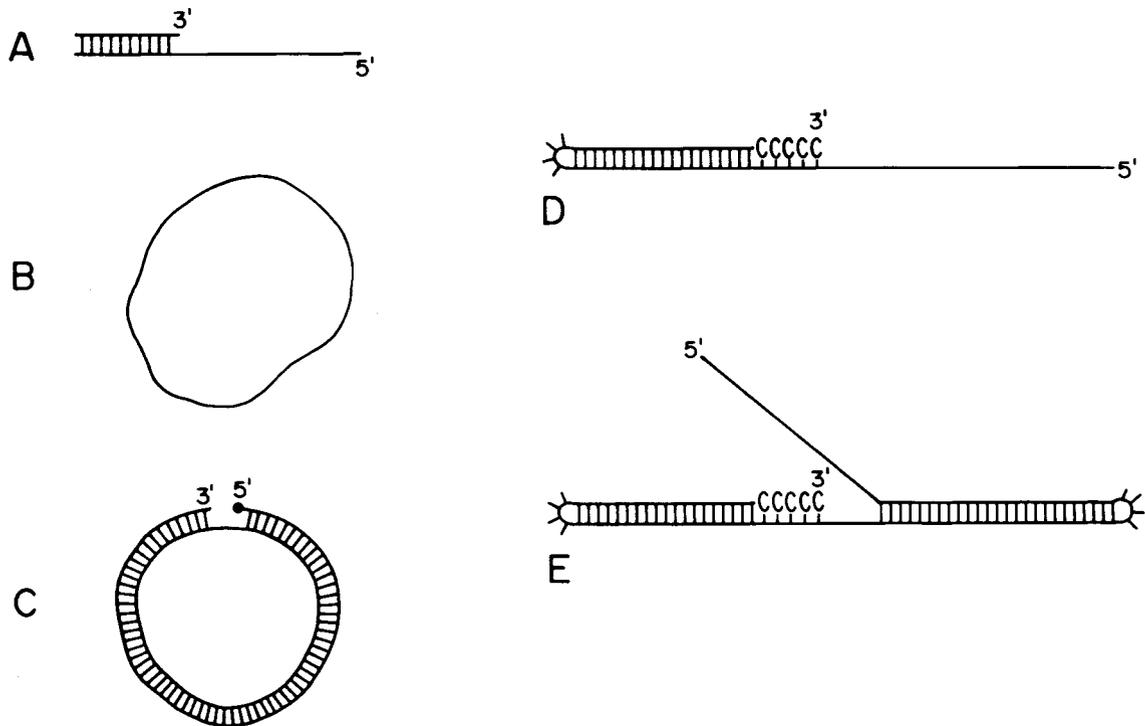[a]T. Minagawa (pers. comm.); T. Formosa (unpubl.)

Figure 6. Some special types of DNA molecules that have been useful for studying DNA replication mechanisms. The molecules shown in *D* and *E* are defined-sequence molecules that have been chemically synthesized. In addition to having structures that mimic aspects of a DNA replication fork, they contain a contiguous run of C residues at their 3' ends; these are necessary to keep the T4 DNA polymerase in place in the presence of dCTP (M. Munn and M. Urdea, unpubl.).

Urdea, unpubl.). As the technology for making such custom-designed DNA molecules continues to improve, synthetic DNAs will undoubtedly become widely used as substrates for studying many of the different reactions that occur on DNA. For example, they should help allow the detailed mechanism of many of the "partial reactions" that represent specific steps in genetic recombination pathways to be dissected in vitro.

New Technologies Can Help Revolutionize DNA Enzymology

There is a strong tendency for laboratories to become comfortable with a limited group of methods and approaches and to continue using them even when much better ones have become available. This is especially true when the new technologies are expensive, involving capital costs for new machines. As with the steel industry, however, one can make the argument that such major capital investments are overdue and cost-effective at the present time.

Everyone is familiar with the use of new gene cloning, sequencing, and expression technologies for overproducing gene products. These new techniques now allow the production of gram quantities of interesting proteins that were once difficult to obtain pure even in microgram amounts. But there are also other uses of new technologies that are perhaps not so obvious. Some of these have begun to change our approach to

the DNA enzymology of the T4 bacteriophage-infected cell, as I describe briefly below.

Throughout our work on the T4 system, we have made extensive use of DNA cellulose chromatography. The rationale for this approach was explained in 1968, in my first talk at a Cold Spring Harbor Symposium. The T4 bacteriophage makes a large commitment to DNA-related processes, as evidenced by the fact that 8% of the total early T4 protein made in a normal infection binds to DNA, compared with only 1% of the total *E. coli* proteins (Fig. 4 in Alberts et al. 1969). A subsequent experiment performed with DNA-free "minicells" of *E. coli* showed that the vast majority of the DNA cellulose–binding proteins in *E. coli* are bound to the *E. coli* chromosome in vivo (Alberts 1970), a finding in agreement with our belief that at least most of the T4 proteins that bind to DNA cellulose columns under our conditions have a DNA-related function.

We, like everyone else, now routinely use electrophoresis on SDS-containing polyacrylamide slab gels to analyze each chromatography step during protein purifications (Studier 1973). Figure 7 shows such an analysis of the total DNA binding proteins obtained when 200 g of T4-infected cells were extracted and fractionated on a large DNA cellulose column. This column was rinsed in a buffer containing 0.1 M NaCl and then eluted in salt steps of 0.3 M, 0.5 M, and 2.0 M NaCl. The particular T4 mutant used for the infection greatly overproduces the gene-*32* protein (Gold et al.
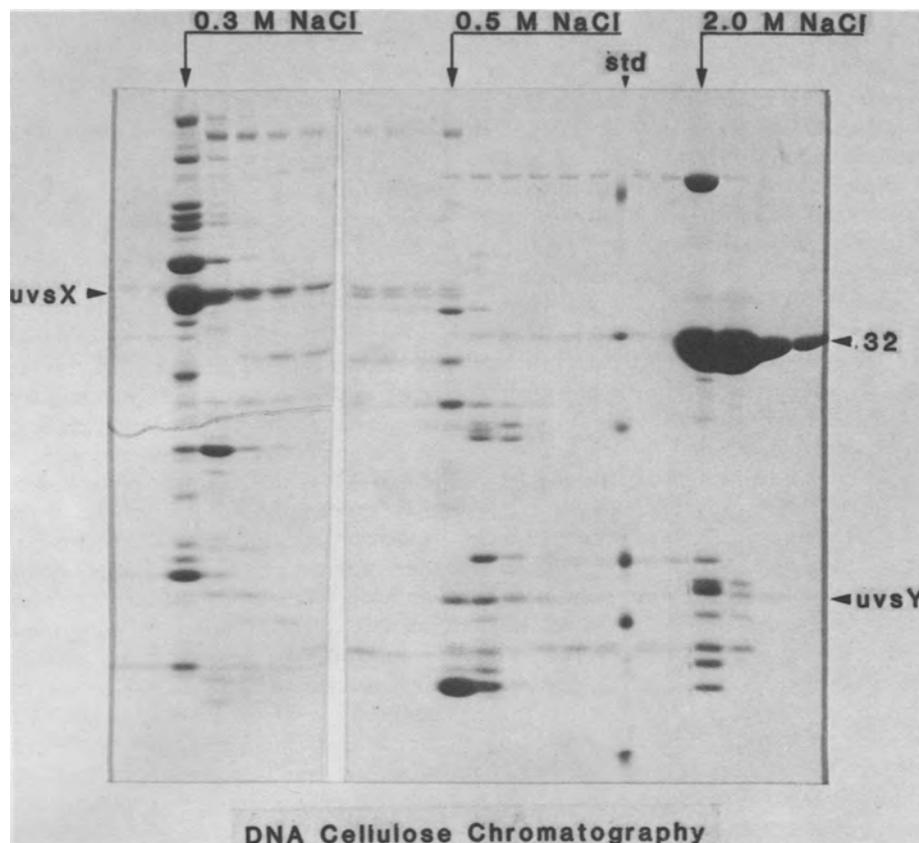
Figure 7. Analysis by SDS-polyacrylamide gel electrophoresis of the total proteins that bind to a single-stranded, calf thymus DNA cellulose column when a crude extract prepared from 200 g of T4-infected cells is loaded onto a 300-ml column in the presence of 0.1 M NaCl. Successive fractions have been analyzed in adjacent gel lanes, beginning with an elution with buffer containing 0.3 M NaCl, as indicated. The *uvsX* protein is the T4 RecA protein analog, and it has a molecular mass of about 40,000 daltons; the *uvsY* protein is likewise involved in genetic recombination in T4-infected cells, and it has a molecular mass of about 16,000 daltons (see Formosa and Alberts, this volume). Because the cells used for this fractionation were infected with T4 bacteriophage that were mutated in genes *33* and *55*, they greatly overproduce the gene-*32* protein (34,500 daltons). This gel was stained with Coomassie blue and therefore includes some *E. coli* host proteins. It represents the first step in the purification of the gene-*61* protein, which is the doublet of two equally stained bands that migrate at about the same rate as the *uvsX* protein in the fractions just preceding the 0.5 M elution (M. Munn and J. Barry, unpubl.).

1976), which accounts for the huge amount of this protein obtained in the 2 M NaCl elution. About 50 other distinct protein bands are also detected by this analysis. Of these bands, 10 correspond to known T4 proteins that we have purified to essential homogeneity, starting from such DNA cellulose eluates.

Most of the remaining 40 protein bands in Figure 7 represent unknown proteins that are normally discarded during the subsequent chromatographic steps that are used to purify the proteins of interest. This is not only wasteful, but frustrating, since one can be certain that many of the proteins being thrown out are even more interesting than the ones being saved. Since we have nearly run out of interesting T4 mutants that can be used to screen for altered DNA-binding proteins, some alternative approach is required.

One approach that we have been too slow to use involves the sequencing of the 20 to 30 aminoterminal amino acids in each protein band. This can now be done with standard technologies that require only about 0.3 μg of a 30,000-dalton protein (10 pmoles) and about 24 hr of machine time per protein (Hewick et al. 1981). If all of us routinely obtained this information for each protein we find, a quick reference to a computer bank of such sequences would suffice to enable every "newly discovered" protein to be correctly identified and indexed with respect to work done in other laboratories. In the long run, this should save the funding agencies more than the cost of the new facilities required to make such sequencing routine. In our case (and others), the sequence information obtained could also be used to construct a synthetic DNA probe to find and characterize the gene that encodes the protein (Noyes et al. 1979). The cloned DNA sequence could then be employed to create mutations in that gene, as required to test the protein's biological role.

Another approach that is possible now, but which was unforeseen in 1968, involves a powerful application of DNA cellulose chromatography that was ini-

tially developed by Gaudray et al. (1981). The technique takes advantage of the fact that all proteins that bind to specific DNA sequences also seem to have a substantial, but greatly reduced affinity for any DNA sequence (von Hippel 1979). A DNA cellulose column that contains a large amount of any readily obtained DNA is first loaded with a crude extract and rinsed to remove all non-DNA-binding proteins. Instead of performing salt elutions, as illustrated in Figure 7, the column is eluted with buffer that contains a small amount of a particular cloned DNA molecule, chosen to contain a specific DNA sequence that one of the cellular proteins should recognize. This experiment must be carried out at a salt concentration that is high enough to cause the protein of interest to equilibrate freely between the DNA on the column and the DNA in the mobile phase. The necessary conditions are determined empirically, by repeating the DNA elution after sequential washes of a single column at progressively higher concentrations of salt. Many of the DNA cellulose–bound proteins will be constantly partitioning back and forth between the stationary and mobile phase. However, because the stationary phase contains much more DNA than the mobile phase, only the few proteins that strongly prefer a DNA sequence that is highly enriched in the mobile phase will move rapidly enough through the column to be eluted. Enormous purifications of those proteins that specifically recognize the DNA in the mobile phase can therefore be obtained.

This method can be used effectively with any cloned DNA sequence, with preelution with a similar DNA of different sequence used as a control. It can be used even more effectively with chemically synthesized DNA molecules of any desired sequence. This makes possible the immediate search for a protein that might recognize a short DNA sequence, even one that is only suspected of being interesting because of its nonrandom distribution in a sequenced genome. A number of such DNA sequences that should be tested in this way have already been identified in the T4 genome. We also plan to elute DNA cellulose columns with the type of synthetic DNA molecules illustrated previously in Figure 6, D and E. In this way, one can probe for proteins among the set displayed in Figure 7 that bind preferentially to a specific type of DNA *structure*, rather than to any particular DNA sequence. The elution with synthetic DNA molecules that contain special structures generated during genetic recombination, such as a cross-strand exchange, should likewise provide a powerful tool for studies of some of the proteins that catalyze the genetic recombination process.

CONCLUSION

In ending his introduction to the 1978 Symposium, Arthur Kornberg made a series of eight predictions about the future of DNA replication (Kornberg 1979). In an attempt to follow in this tradition, I would like to make one prediction, which is that we will see a marked resurgence of interest and activity in the enzymology of prokaryotic organisms in the coming years, including both *E. coli* and the long-forgotten organism, bacteriophage T4.

On what do I base this statement? We have just been through an era in which enzymology has been generally ignored, with students seemingly eager to clone any eukaryotic gene in preference to a day in the cold room with a DEAE-cellulose column. As some of the initial excitement wears off, it is becoming clear that to understand such fascinating problems as the control of eukaryotic gene expression will require the same kind of serious protein biochemistry that was required to work out DNA replication enzymology. There is no reason to think that any central genetic mechanism will be less complex than DNA replication and every reason to believe that the same type of sophisticated multiprotein complex will be involved. If we think of molecular genetics in terms of "protein machines" rather than in terms of sequential reactions that are carried out by individual proteins, then it becomes abundantly clear why it is important to study the detailed enzymology of all of these processes in organisms that are simpler than the higher eukaryotes — even if one is self-centered enough to care only about human biology. In a number of simpler organisms, the combined power of genetics and biochemistry makes it possible to find and study every piece of a multiprotein complex, as is required to come to any real understanding of the mechanism involved. In contrast, this goal presently seems beyond reach for most higher eukaryotic systems.

Through combinations of the types of experiments described in this paper, I am convinced that it will be possible to assign both genes and functions to all of the T4-induced proteins that are involved in the various aspects of its DNA metabolism. A task that seemed unimaginable to me at the time of the 1968 Cold Spring Harbor Symposium now seems like a realistic goal. The drastic change in what is possible in biological research has been largely due to the impressive development of many new technologies over the course of the past 16 years. One cannot help but wonder how much further this can go. Will the experimental tools available to the biological scientist increase in their power and speed as much in the next 16 years as they have in the past 16? We will only know the answer to this question when the 65th Cold Spring Harbor Symposium is held in June of the year 2000. I do know, however, that there will be many exciting advances in our understanding of how cells work in the interim and that we are all very lucky to be biologists at such a productive time.

ACKNOWLEDGMENTS

sults that made this article possible. Their names appear on the articles cited on which I am a coauthor.

REFERENCES

ALBERTS, B.M. 1970. Function of gene 32-protein, a new protein essential for the genetic recombination and replication of T4 bacteriophage DNA. *Fed. Proc.* **29:** 1154.

ALBERTS, B.M. and L. FREY. 1970. T4 bacteriophage gene 32: A structural protein in the replication and recombination of DNA. *Nature* **227:** 1313.

ALBERTS, B. and R. STERNGLANZ. 1977. Recent excitement in the DNA replication problem. *Nature* **269:** 655.

ALBERTS, B.M., F.J. AMODIO, M. JENKINS, E.D. GUTMANN, and F.L. FERRIS. 1969. Studies with DNA-cellulose chromatography. I. DNA-binding proteins from *Escherichia coli. Cold Spring Harbor Symp. Quant. Biol.* **33:** 289.

ALBERTS, B.M., J. BARRY, P. BEDINGER, T. FORMOSA, C.V. JONGENEEL, and K.N. KREUZER. 1983. Studies on DNA replication in the T4 bacteriophage in vitro system. *Cold Spring Harbor Symp. Quant. Biol.* **47:** 655.

ALBERTS, B.M., J. BARRY, P. BEDINGER, R.L. BURKE, U. HIBNER, C.-C. LIU, and R. SHERIDAN. 1980. Studies of replication mechanisms with the T4 bacteriophage *in vitro* system. *ICN-UCLA Symp. Mol. Cell. Biol.* **19:** 449.

ALBERTSSON, P.-A. 1960. *Partition of cell particles and macromolecules.* Wiley, New York.

BARRY, J. and B. ALBERTS. 1972. *In vitro* complementation as an assay for new proteins required for bacteriophage T4 DNA replication: Purification of the complex specified by T4 genes 44 and 62. *Proc. Natl. Acad. Sci.* **69:** 2717.

BEDINGER, P., M. HOCHSTRASSER, C.V. JONGENEEL, and B.M. ALBERTS. 1983. Properties of the T4 bacteriophage DNA replication apparatus: The T4 dda DNA helicase is required to pass a bound RNA polymerase molecule. *Cell* **34:** 115.

BEHME, M.T. and K. EBISUZAKI. 1975. Characterization of a bacteriophage T4 mutant lacking DNA-dependent ATPase. *J. Virol.* **15:** 50.

BERNSTEIN, C. and S.S. WALLACE. 1983. DNA repair. In *Bacteriophage T4* (ed. C. Mathews et al.), p. 138. American Society for Microbiology, Washington, D.C.

BURKE, R.L., T. FORMOSA, K.S. COOK, A.F. SEASHOLTZ, J. HOSODA, and H. MOISE. 1983. Use of two-dimensional polyacrylamide gels to identify T4 prereplicative proteins. In *Bacteriophage T4* (ed. C. Mathews et al.), p. 321. American Society for Microbiology, Washington, D.C.

EPSTEIN, R.H., A. BOLLE, C. STEINBERG, E. KELLENBERGER, E. BOY DE LA TOUR, R. CHEVALLEY, R.S. EDGAR, M. SUSMAN, G.H. DENHARDT, and A. LIELAUSIS. 1964. Physiological studies of conditional lethal mutants of bacteriophage T4D. *Cold Spring Harbor Symp. Quant. Biol.* **28:** 375.

FORMOSA, T., R.L. BURKE, and B.M. ALBERTS. 1983. Affinity purification of T4 bacteriophage proteins essential for DNA replication and genetic recombination. *Proc. Natl. Acad. Sci.* **80:** 2442.

FULLER, R.S., J.M. KAGUNI, and A. KORNBERG. 1981. Enzymatic replication of the origin of the *E. coli* chromosome. *Proc. Natl. Acad. Sci.* **78:** 7370.

FULTON, A.B. 1982. How crowded is the cytoplasm? *Cell* **30:** 345.

GAUDRAY, P., C. TYNDALL, R. KAMEN, and F. CUZIN. 1981. The high affinity binding site on polyoma virus DNA for the viral large-T protein. *Nucleic Acids Res.* **9:** 5697.

GOLD, L., P. O'FARRELL, and M. RUSSELL. 1976. Regulation of gene 32 expression during bacteriophage T4 infection of *Escherichia coli. J. Biol. Chem.* **251:** 7251.

GREENBLATT, J. 1981. Regulation of transcription termination by the *N* gene protein of bacteriophage lambda. *Cell* **24:** 8.

HEWICK, R.M., M.W. HUNKAPILLAR, L.E. HOOD, and W.J. DREYER. 1981. A gas-liquid solid phase peptide and protein sequenator. *J. Biol. Chem.* **256:** 7990.

HUANG, C.-C., J.E. HEARST, and B.M. ALBERTS. 1981. Two types of replication proteins increase the rate at which T4 DNA polymerase traverses the helical regions in a single-stranded DNA template. *J. Biol. Chem.* **256:** 4087.

HUBERMAN, J.A., A. KORNBERG, and B.M. ALBERTS. 1971. Stimulation of T4 bacteriophage DNA polymerase by the protein product of T4 gene 32. *J. Mol. Biol.* **62:** 39.

JONGENEEL, C.V., P. BEDINGER, and B.M. ALBERTS. 1984. Effects of the bacteriophage T4 dda protein on DNA synthesis catalyzed by purified T4 replication proteins. *J. Biol. Chem.* (in press).

KORNBERG, A. 1979. Aspects of DNA replication. *Cold Spring Harbor Symp. Quant. Biol.* **43:** 1.

———. 1980. *DNA replication.* W.H. Freeman, San Francisco.

———. 1984. DNA replication. *Trends Biochem. Sci.* **9:** 122.

KOZINSKI, A.W. 1983. Origins of T4 DNA replication. In *Bacteriophage T4* (ed. C. Mathews et al.), p. 111. American Society for Microbiology, Washington, D.C.

KRELL, H., H. DURWALD, and H. HOFFMANN-BERLING. 1979. A DNA-unwinding enzyme induced in bacteriophage T4-infected *E. coli* cells. *Eur. J. Biochem.* **93:** 387.

LIU, C.-C. and B.M. ALBERTS. 1980. Pentaribonucleotides of mixed sequence are synthesized and efficiently prime de novo DNA chain starts in the T4 bacteriophage DNA replication system. *Proc. Natl. Acad. Sci.* **77:** 5698.

———. 1981. Characterization of RNA primer synthesis in the T4 bacteriophage in vitro DNA replication system. *J. Biol. Chem.* **256:** 2821.

LIU, C.-C., R.L. BURKE, U. HIBNER, J. BARRY, and B.M. ALBERTS. 1979. Probing DNA replication mechanisms with the T4 bacteriophage in vitro system. *Cold Spring Harbor Symp. Quant. Biol.* **43:** 469.

LJUNGQUIST, E. and A.I. BUKHARI. 1977. State of prophage Mu DNA upon induction. *Proc. Natl. Acad. Sci.* **74:** 3143.

MACE, D.C. and B.M. ALBERTS. 1984a. The complex of T4 bacteriophage gene 44 and 62 replication proteins form an ATPase that is stimulated by DNA and by T4 gene 45 protein. *J. Mol. Biol.* **177:** 279.

———. 1984b. Characterization of the stimulatory effect of T4 gene 45 protein and the gene 44/62 protein complex on DNA synthesis by T4 DNA polymerase. *J. Mol. Biol.* **177:** 313.

McHENRY, C. and A. KORNBERG. 1981. DNA polymerase III holoenzyme. In *The enzymes* (ed. P.D. Boyer), vol. 14, p. 39. Academic Press, New York.

MORRIS, C.F., L.A. MORAN, and B.M. ALBERTS. 1979a. Purification of gene 41 protein of bacteriophage T4. *J. Biol. Chem.* **254:** 6797.

MORRIS, C.F., N.K. SINHA, and B.M. ALBERTS. 1975. Reconstruction of the bacteriophage T4 DNA replication apparatus from purified components: Rolling circle replication following de novo chain initiation on a single-stranded circular DNA template. *Proc. Natl. Acad. Sci.* **72:** 4800.

MORRIS, C.F., H. HAMA-INABA, D. MACE, N.K. SINHA, and B. ALBERTS. 1979b. Purification of the gene 43, 44, 45, and 62 proteins of the bacteriophage T4 DNA replication apparatus. *J. Biol. Chem.* **254:** 6787.

MOSIG, G. 1983. Relationship of T4 DNA replication and recombination. In *Bacteriophage T4* (ed. C. Mathews et al.), p. 120. American Society for Microbiology, Washington, D.C.

MOSIG, G., A. LUDER, G. GARCIA, R. DANNENBERG, and S. BOCK. 1979. In vivo interactions of genes and proteins in DNA replication and recombination of phage T4. *Cold Spring Harbor Symp. Quant. Biol.* **43:** 501.

NASH, H.A. 1981. Integration and excision of bacteriophage lambda: The mechanism of conservative site-specific recombination. *Annu. Rev. Genet.* **15:** 143.

NOSSAL, N.G. 1980. RNA priming of DNA replication by bacteriophage T4 proteins. *J. Biol. Chem.* **255:** 2176.

Nossal, N.G. and B.M. Alberts. 1983. The mechanism of DNA replication catalyzed by purified bacteriophage T4 DNA replication proteins. In *Bacteriophage T4* (ed. C. Mathews et al.), p. 71. American Society for Microbiology, Washington, D.C.

Noyes, B.E., M. Mevarechi, R. Stein, and K.L. Agarwal. 1979. Detection and partial sequence analysis of gastrin mRNA by using an oligodeoxynucleotide probe. *Proc. Natl. Acad. Sci.* **76:** 1770.

Piperno, J.R. and B.M. Alberts. 1978. An ATP stimulation of T4 DNA polymerase mediated via T4 gene 44/62 and 45 proteins: The requirement for ATP hydrolysis. *J. Biol. Chem.* **253:** 5174.

Radding, C.M. 1982. Strand transfer in homologous genetic recombination. *Annu. Rev. Genet.* **16:** 405.

Ratner, D. 1974. The interaction of bacterial and phage proteins with immobilized *E. coli* RNA polymerase. *J. Mol. Biol.* **88:** 373.

Richardson, C.C. 1983. Bacteriophage T7: Minimal requirement for the replication of a duplex DNA molecule. *Cell* **33:** 315.

Roth, A.C., N.G. Nossal, and P.T. Englund. 1982. Rapid hydrolysis of deoxynucleoside triphosphates accompanies DNA synthesis by T4 DNA polymerase and T4 accessory proteins 44/62 and 45. *J. Biol. Chem.* **257:** 1267.

Scherer, S. and R.W. Davis. 1979. Replacement of chromosome segments with altered DNA sequences constructed in vitro. *Proc. Natl. Acad. Sci.* **76:** 4951.

Studier, F.W. 1973. Analysis of bacteriophage T7 early RNAs and proteins on slab gels. *J. Mol. Biol.* **79:** 237.

Urdea, M.S., J.P. Merryweather, G.T. Mullenbach, D. Coit, U. Heberlein, P. Valenzuela, and P.J. Barr. 1983. Chemical synthesis of a gene for human epidermal growth factor urogastrone and its expression in yeast. *Proc. Natl. Acad. Sci.* **80:** 7461.

Venkatesan, M., L.L. Silver, and N.C. Nossal. 1982. Bacteriophage T4 gene 41 protein, required for the synthesis of RNA primers, is also a DNA helicase. *J. Biol. Chem.* **257:** 12426.

von Hippel, P.H. 1979. On the molecular basis of the specificity of interaction of transcriptional proteins with genome DNA. In *Biological regulation and development* (ed. R. Goldberger), vol. 1, p. 279. Plenum Press, New York.

Warner, H.R. and M.D. Hobbs. 1967. Incorporation of uracil-C^{14} into nucleic acids in *E. coli* infected with bacteriophage T4 and T4 amber mutants. *Virology* **33:** 376.

Winston, F., F. Chumley, and G.R. Fink. 1983. Eviction and transplacement of mutant genes in yeast. *Methods Enzymol.* **101:** 211.

Wood, W.B. and H.R. Revel. 1976. The genome of bacteriophage T4. *Bacteriol. Rev.* **40:** 847.

Zimmerman, S.B. and B.H. Pheiffer. 1983. Macromolecular crowding allows blunt end ligation by DNA ligases from rat liver or *E. coli*. *Proc. Natl. Acad. Sci.* **80:** 5852.

The DNA Enzymology of Protein Machines

B.M. Alberts

References	This article cites 50 articles, 27 of which can be accessed free at: http://symposium.cshlp.org/content/49/1.refs.html Article cited in: http://symposium.cshlp.org/content/49/1#related-urls
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here**

To subscribe to *Cold Spring Harbor Symposia on Quantitative Biology* go to:
http://symposium.cshlp.org/subscriptions