NUCLEIC ACID-PROTEIN RECOGNITION

ACADEMIC PRESS, INC.
NEW YORK SAN FRANCISCO LONDON

In Vitro DNA Replication Catalyzed by Six Purified T4 Bacteriophage Proteins

BRUCE ALBERTS,[1] JACK BARRY,[1] MICHAEL BITTNER,[2]
MONIQUE DAVIES, HIROKO HAMA-INABA,[3]
CHUNG-CHENG LIU,[1] DAVID MACE,[4]
LARRY MORAN,[5] CHARLES F. MORRIS,[2]
JEANETTE PIPERNO,[6] AND NAVIN K. SINHA[7]
Department of Biochemical Sciences
Frick Chemical Laboratory
Princeton University
Princeton, New Jersey

INTRODUCTION

The goal of our work has been to understand the basic mechanism of DNA replication, with special emphasis on the protein and nucleic acid interactions involved. It appears that in all systems thus far exam-

[1] Present address: Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, California.
[2] Present address: Department of Biological Chemistry, Washington University School of Medicine, St. Louis, Missouri.
[3] Present address: National Institute of Radiological Science Anagawa, Chiba-shi, Japan.
[4] Present address: Roche Institute of Molecular Biology, Nutley, New Jersey.
[5] Present address: Université de Genève, Institut de Biologie Moléculaire, Geneva, Switzerland.
[6] Present address: Department of Biochemistry, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania.
[7] Present address: Waksman Institute of Microbiology, Rutgers University, New Brunswick, New Jersey.

ined closely, replication is carried out by a multienzyme complex of several proteins (e.g., see refs. 1–4) quite possibly interwoven with the DNA to create a particle with a unique three-dimensional conformation (5). However, attempts to purify intact DNA replication forks with these proteins attached have thus far failed, apparently reflecting a basic lability of the entire structure (6–8). This lability of the "replication apparatus" greatly complicates *in vitro* studies of replication mechanisms. In this context, it is relevant to note the extent to which the comparative stability of ribosome and RNA polymerase protein complexes has facilitated the advances in our knowledge of the biochemistry of translation and transcription.

Inspired by the reconstitution of functional ribosomal particles from their individual components (9), we set out to approach the replication problem by individually isolating each of the proteins of one DNA-replication apparatus, trusting that we could eventually get this structure to self-assemble *in vitro* by incubating intracellular concentrations of these proteins with DNA. With this goal in mind, we focused our attention on the T4 bacteriophage replication system. Its major advantage was that the genetics of T4 replication had been well worked out. In addition, on the order of 60 replication forks are established in each T4-infected cell (10); this makes the T4 system a relatively rich source for biochemical isolation of replication proteins, especially with the discovery of T4 mutants which overproduce some of these gene products (11–14).

GENETIC CHARACTERIZATION OF THE T4 BACTERIOPHAGE REPLICATION SYSTEM

In 1963, R. H. Epstein, R. S. Edgar, and their collaborators presented their now classical analysis of the conditional-lethal mutants of T4 bacteriophage (15). These mutants, which mapped at that time into 47 different complementation groups, or genes, included alterations in five of the presently identified six protein components of the T4 DNA replication apparatus: the products of T4 genes 32, 41, 43, 44, 45, and 62. These mutations are characterized by the fact that infection under nonpermissive conditions gives rise to little or no DNA synthesis (15–17) even though the normal deoxyribonucleoside triphosphate precursors are available (18). In addition, temperature-sensitive mutants have been isolated for the products of genes 32, 41, 43, and 45,

which cause T4 DNA replication to cease within 1 min after a shift from low (25°C) to high temperature (42°C) (19). Since replication forks appear to stop completely before moving even one-fifth the length of one mature T4 genome following this shift to 42°C (20), each of these four proteins appears to be involved directly in the DNA polymerization process. The other two gene products (genes 44 and 62) appear to function as a single tight complex of two polypeptide chains (21). The few phage mutants available carrying temperature-sensitive defects in gene 44 replicate their DNA normally at 42°C if allowed 10 min of prior infection at 30°C (19; J. Karam, personal communication). It is thus possible that these two gene products function only in the initial stages of T4 DNA synthesis. However, no firm conclusion can be drawn in this respect, especially since the altered polypeptide may be stabilized to heat inactivation as soon as the 44–62 complex forms.

The first of the T4 replication gene products to be identified and extensively characterized was the product of gene 43, T4 DNA polymerase (22–29). The gene for DNA polymerase maps in a cluster of replication genes (genes 41, 43, 62, 44, and 45) which surround the preferred T4 replication origin as defined by Mosig and co-workers (30). This tight clustering suggests that these gene products interact with each other (31); likewise, their close linkage to a replication origin raises the possibility that at least one of the proteins acts with DNA sequence specificity in the initiation of replication forks.

Besides the above six proteins, additional gene products were identified in 1963 whose alteration leads to either "DNA synthesis-delay" or "DNA synthesis-arrest" phenotypes (15). The number of genetically identified T4 gene products with such partial effects on DNA synthesis has increased since then (e.g., see refs. 32–35), although very little is known concerning their biochemistry. These proteins would appear to play a less central role in T4 DNA synthesis, and it is still not certain that they are *directly* involved in the *in vivo* replication process. Finally, a series of elegant, although indirect, experiments suggest that some of the T4-induced enzymes required for synthesis of deoxyribonucleoside triphosphates could act at the replication fork as well (36,37).

Before reviewing the current status of our work analyzing the biochemistry and enzymology of the T4 replication system, a general view of DNA replication mechanisms will be presented. This will help us in considering possible functions for those T4 replication proteins whose raison d'être is not yet understood.

THE GENERAL STRUCTURE OF A REPLICATION FORK

Despite the genetic complexity indicated for the T4 replication apparatus, most of us working on the biochemistry of DNA replication in the early 1960's tended to view replication as a relatively simple process. J. Cairns had shown by radioautography that the two parental polynucleotide chains become separated from each other, and at about the same time become paired with newly made complementary chains, at a structure termed a "replication fork" (38). The simplest mechanism to draw (and therefore the favorite) was one in which DNA polymerases (3,39,40) continuously elongate the DNA chains on the two sides of this fork, as schematically shown in Fig. 1A. Because of the antiparallel orientation of strands in the DNA double helix, only one of these DNA polymerases could proceed by chain elongation in the 5′ to 3′ chain direction (by attack of a 3′-OH of a polymer chain end on the activated 5′ end of an incoming deoxyribonucleoside 5′-triphosphate monomer, as observed). A second type of DNA polymerase had to be postulated for which the growing chain end carried the 5′ triphosphate activation, and the incoming 5′-triphosphate monomer was "backed in" to present its 3′-OH for the polymerization reaction. Efforts to find such a second DNA polymerase were unsuccessful. Moreover, pulse-labeled incorporation of [^{3}H]thymidine revealed a perponderance of 1000–2000 nucleotide-long "Okazaki pieces" of DNA at the growing fork (41,42), synthesized in the 5′ to 3′ chain direction only (43) and subsequently joined together by DNA ligase (44–46). These observations suggested that daughter DNA chains grow discontinuously at the replication fork, and in conjunction with detailed electron microscopic analysis of replicating molecules (47), led most workers in the field to switch to favor the type of model for replication schematically illustrated in Fig. 1B. Here, owing to the asymmetry of syntheses on the leading and lagging sides of the fork, only one type of DNA polymerase is required, of the type observed. However, fresh complications arise due to the (at first sight perverse) inability of known DNA polymerases to start DNA chains *de novo* (39,40). Such new chain starts are required at frequent intervals on the lagging side of the fork (Fig. 1B). Partly for this reason, a new primer-generating polymerase was sought, and eventually found, in the *Escherichia coli* replication system (48,49). Its properties, combined with data from other systems (50), suggest that such an enzyme functions *in vivo* to create short RNA primers (approximately 10 nucleotides long), and that these primers are first elongated by DNA polymerase and then erased (and replaced by DNA) prior to ligase

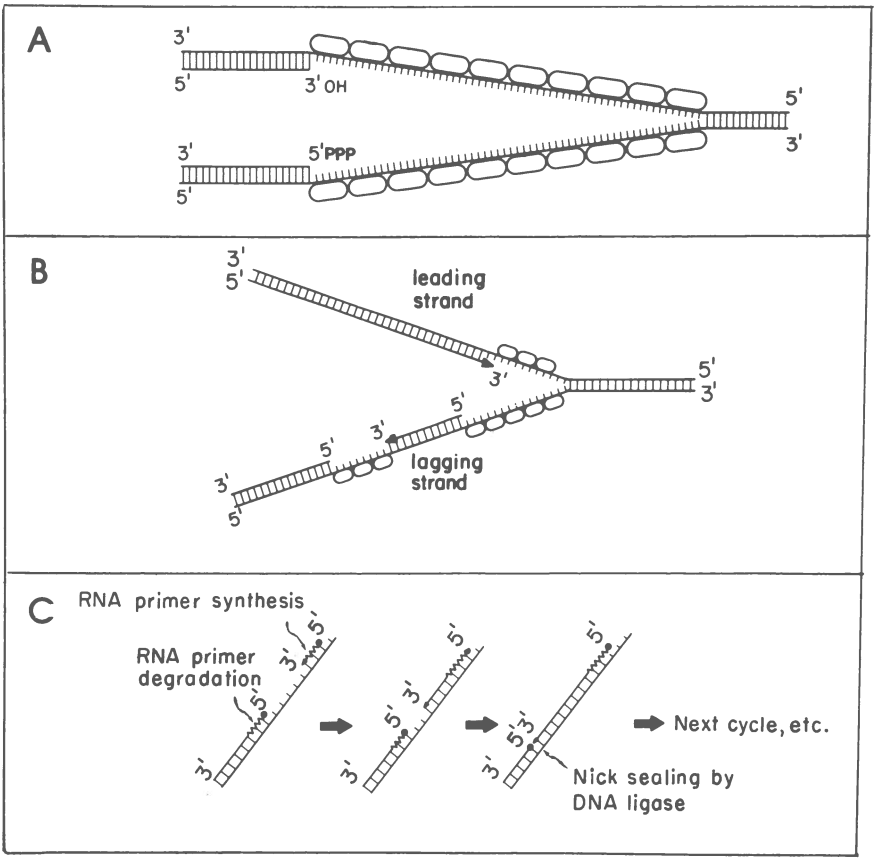

Fig. 1. Schematic representation of possible replication fork structures. (A) "Everybody's favorite" replication fork of the early 1960's; both strands grow continuously in opposite chemical directions. (B) The current view of a replication fork, in which "lagging"- and "leading"-strand syntheses differ, but are carried out by the same type of DNA polymerase. (C) Lagging-strand synthesis in more detail: a cycle of synthesis and degradation of the RNA primer believed to start Okazaki-pieces [for details, see Kornberg (3)]. Note that a role for DNA-unwinding protein has been suggested in both (A) and (B).

sealing of the lagging-strand DNA pieces. One such cycle of RNA primer synthesis and degradation is schematically illustrated in Fig. 1C.

Remnants of the postulated RNA primer disappear so rapidly in complete replication systems that one suspects that primer degradation may be tightly coupled to primer utilization. In the limit, both

processes might be catalyzed by the same enzyme complex, thereby maximizing the efficiency of primer removal.

FIDELITY CONSTRAINTS ON THE MECHANISM

The type of mechanism for DNA replication suggested in Fig. 1B and 1C seems at first sight much more complex and unwieldy than the mechanism in Fig. 1A. Is it possible that nature has been foolish and wasteful in engineering these crucial processes? We have previously (51) suggested a rationale for the mechanism observed based on the fact that DNA replication must be an extremely accurate process. Indeed, the observed templating fidelity is such that only one mistake is made in 10^9 to 10^{10} base-pair replications in *E. coli* (52).

The key observation was that of Brutlag and Kornberg (53), who showed that for polymerization, both T4 and *E. coli* DNA polymerases have a very strong requirement for a Watson–Crick base-paired residue at the 3′-OH primer terminus to which nucleotides are being added. When confronted with a template-primer with a terminal mismatch, these polymerases make use of their built-in $3' \rightarrow 5'$ exonuclease activities to clip off unpaired primer residues by hydrolysis; this continues until enough nucleotides are removed to regenerate a base-paired terminus and create an active template-primer. As a result, these polymerases will efficiently remove their own polymerization errors (see also refs. 25–27,29). This self-correcting feature allows DNA polymerase to select for the proper template base-pairing of each added nucleoside triphosphate in a separate backward reaction, in addition to its strong selection for base-pairing of nucleoside triphosphates during the initial polymerization. Theoretically, the intrinsic mistake frequency at this first step due to base mispairings can be reduced by the same factor in the second proofreading step. It has been argued from chemical considerations that the intrinsic mistake frequency is unlikely to be less than 10^{-5} (54). In this case, at least one proofreading step (perhaps two) would have been absolutely required to approach a tolerable mutational load ($10^{-5} \times 10^{-5} = 10^{-10}$).

Accepting this view, it becomes clear why no DNA polymerase is able to start a new polynucleotide chain *de novo,* whereas DNA-dependent RNA polymerases can do so: the necessity to be self-correcting gives DNA polymerase a stringent requirement for a perfectly base-paired primer terminus; to ask this enzyme to start synthesis in the complete absence of primer, without losing any of its discrimination between base-paired and unpaired template 3′-OH

termini, seems contradictory. In contrast, the RNA polymerases need not be self-correcting, inasmuch as relatively high error rates can be tolerated during transcription.

This line of reasoning likewise suggests an explanation for the failure to find a second DNA polymerase which adds deoxyribonucleoside 5′-triphosphates in such a way as to cause chains to grow in the 3′ → 5′ direction, as required in Fig. 1A. Despite the relatively simple type of replication fork mechanism that this would allow, for such a 3′ → 5′ polymerase the growing 5′ chain end (rather than the incoming mononucleotide) carries the triphosphate activation. Thus, mistakes in polymerization cannot be hydrolyzed away without a special enzymatic system for reactivating the bare 5′ chain end thus created.

Finally, this view suggests why an erasable ribonucleotide-containing primer might be preferred for priming, rather than a non-erased DNA primer which might otherwise seem more economical. The argument that a self-correcting polymerase cannot start chains *de novo* also implies its converse: that an enzyme which starts chains *de novo* cannot do a good job of self-correcting. Thus, any enzyme which primes discontinuous synthesis will of necessity make a relatively inaccurate copy (e.g., one error in 10^5). Even if the amount of this copy which is retained in the final product constitutes 1% of the total genome (10 nucleotides per 1000 nucleotide DNA fragment), the resulting increase in overall mutation rate could be enormous. Thus, it seems reasonable to suggest that the use of ribonucleotides in synthesis of primer was of great evolutionary advantage, since it automatically marked these sequences as "bad copy" to be removed.

In conclusion, it seems likely that the type of mechanism for DNA replication sketched in Fig. 1B and 1C, sometimes viewed as the whimsical result of historical accident, was specially selected to ensure that the entire DNA sequence is very accurately copied (i.e., copied by a self-correcting type of DNA polymerase). Additional complexities in the mechanism, including poorly understood requirements for protein cofactors which hydrolyze nucleoside triphosphates to nucleoside diphosphates (4,55,56), may be similarly designed to help generate fidelity in as yet unexplained ways (see, for example, refs. 57,58).

KINETIC CONSTRAINTS ON THE MECHANISM

In prokaryotic systems, the rate of polymerization at the replication fork is on the order of 10^3 nucleotides per second (1–3). While this is

not an unusual turnover number for a simple enzyme, it does seem remarkably fast for a mechanism as complex as DNA replication, where helix unwinding, faithful templating, and exonuclease proofreading must accompany each polymerization step. Clearly the mechanism had to be designed for speed as well as for fidelity, since even at 10^3 nucleotides per second it takes 40 min to replicate the *E. coli* chromosome, and bacterial generation times less than this can be attained only by resorting to dichotomous chromosome growth (59,60).

At least some of the proteins which are not polymerases in the replication apparatus are likely to have been designed to increase the rates of reactions, including the rates of helix opening and template-nucleotide base pairings. The so-called "DNA-unwinding" or "DNA melting" proteins are of ubiquitous occurrence, destabilize the DNA helix, and appear to hold single-stranded DNA template strands in a conformation which is particularly favorable for templating for their homologous DNA polymerase (61–68). They [and possible ATP-driven proteins (69)] may be required at replication forks in part to lower activation energies for helix destabilization and base-pairing reactions.

A second way in which protein components of the replication apparatus are likely to serve in rate accelerations is by bringing other proteins with sequential functions in the replication process into close proximity. Thus, for example, tying primer-generating and primer-erasing proteins in a complex with the traveling DNA polymerase would ensure that they are available immediately when needed. For example, with 100 molecules per cell of a freely diffusible protein ($\sim 10^{-7}$ *M*), the fastest possible association rates would still give an average delay of about 0.1 sec between the time that a site becomes available and its actual recognition.* Clearly, any protein which must function as often as once per every 1000 nucleotides polymerized (once per second) might work more efficiently if kept located at the replication fork in a complex with polymerase (for further details, and some possible implications for DNA-unwinding protein function, see ref. 71).

* For this calculation, we chose a diffusion-controlled rate constant of $10^8\ M^{-1}\ \text{sec}^{-1}$, which is about the theoretical maximum and equal to the association-rate constant observed for the lactose repressor protein binding to its DNA operator in 0.1 *M* KCl (70).

IN VITRO RESULTS WITH THE T4 REPLICATION SYSTEM

ISOLATION OF THE GENE 32 PROTEIN

In our attempts to decipher the biochemistry of the T4 replication system, we have relied on the preexisting genetic analysis to help generate new methods for assay and identification of the relevant protein components. DNA-cellulose chromatography was developed as one such tool (72,73). This method assumes that many of the proteins which function in association with intracellular DNA will recognize purified DNA as a substrate and bind tightly to it *in vitro*. Indeed, when infected-cell extracts were chromatographed on DNA columns, about 20 different T4-induced DNA binding proteins were resolved by polyacrylamide gel electrophoresis (74). By comparing these DNA-binding proteins from mutant and wild-type infections, it was possible to use the genetics to identify protein species as the product of particular bacteriophage genes. In this way, the gene 32 protein was identified, isolated, and subsequently characterized as a "DNA-unwinding" or "DNA melting" protein (61,62,72,74,75).

The gene 32 protein has a monomer molecular weight of 35,000 daltons, and at high concentrations it can self-aggregate to form large multimeric complexes (62,76). As schematically illustrated in Fig. 2B, it binds tightly and cooperatively to completely cover single-stranded DNA, melting all weak intrastrand hairpin helicies and holding the

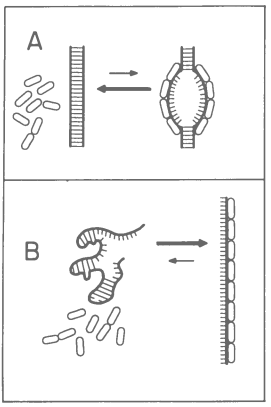

Fig. 2. Helix-melting equilibria in the presence of T4 gene 32 protein under physiological conditions. (A) Perfect double helices; equilibrium toward left. (B) The imperfectly base-paired helices of single-stranded DNA; equilibrium toward right.

DNA in an extended conformation with a 4.6 Å translation distance per nucleotide (75). Since the nucleotide bases in this complex appear to be left uncovered and exposed for templating processes, there is little base specificity in the binding (61,67).

Because it binds much more tightly to single-stranded than to double-stranded DNA, 32 protein catalyzes DNA denaturation, lowering the midpoint (T_m) of the helix–coil transition at physiological ionic strengths by an estimated 40°C. Since the T_m of T4 DNA under physiological salt conditions *in vitro* is about 85°C (77), even with unwinding protein present, long helical regions of T4 DNA should not melt inside the cell except above 45°C. Regions of perfect helix might be transiently invaded by 32 protein *in vivo* at 37°C. However, the reverse reaction would be faster than the forward reaction, so that any opened helical region should rapidly reform, as indicated in Fig. 2A. In contrast, special proteins which nucleate the cooperative binding of 32 protein at specific regions of double helix might suffice to keep the neighboring DNA denatured [see article by von Hippel *et al.* (67)]. In particular, the observed binding of 32 protein to the gene 43 protein (T4 DNA polymerase, ref. 62), and possibly to the gene 44–62 protein (see below), should help destabilize the DNA helix immediately ahead of the replication fork. In addition, cooperative binding of 32 protein in this prefork region should be nucleated on the adjacent section of lagging strand, where 32 protein is already complexed with single-stranded DNA (see Fig. 1B).

An additional possibility is suggested by the fact that proteolytic removal of a terminal 8000 dalton peptide from 32 protein greatly increases the rate at which it denatures T4 DNA *in vitro* (78,79). Moise and Hosoda have proposed that this region of 32 protein normally forms a protective "flap" which reduces 32 protein DNA affinity, except when this flap is altered by specific protein–protein interactions, such as those in the replication complex (79).

At any rate, Nossal has observed that 32 protein will allow *some* strand displacement synthesis by the T4 DNA polymerase on nicked double-helical templates, whereas the polymerase otherwise cannot proceed except on denatured DNA strands (80). This suggests that, whatever the mechanism, this DNA unwinding protein can open helical regions for the T4 DNA polymerase and expose their otherwise unavailable strands for templating. Note that unwinding protein has been schematically diagrammed as having a helix opening role in the replication forks shown earlier (Figs. 1A and 1B). Other possibilities have been outlined in detail previously (71).

ISOLATION OF THE GENE 41 PROTEIN, THE GENE 45 PROTEIN AND THE GENE 44–62 PROTEIN COMPLEX

In early 1971, we were faced with the problem of isolating the four remaining genetically identified T4 replication proteins besides the "DNA-unwinding" or "DNA melting" protein (gene 32) and the DNA polymerase (gene 43) known at that time. Following the approach which had been successful with gene 32 protein, we at first tried to use standard double-label radioisotope techniques to identify and thus purify the wild-type analog of mutationally altered proteins. However, although we now know that the gene 41 protein and the gene 44–62 protein complex can interact with DNA (see below), neither these proteins nor the gene 45 protein bound in substantial amounts to our DNA-cellulose columns (unpublished results of L. Moran). This meant that cascades of chromatographic techniques with less resolution had to be used for their purification. Experience with such techniques convinced us that we would be unlikely to obtain a biologically relevant product using double-label techniques alone, and that an additional assay was needed which would allow biological activity to be monitored throughout the purification.

In developing such an assay, the classical enzymological approach seemed unworkable: not only did we not know what type of activity the gene 41, 44, 45, and 62 proteins might have, but the genetic results raised the possibility that each protein might act as part of a large multienzyme complex, with individual components inactive on their own. Faced with a similar situation, Edgar and Wood had developed an "*in vitro* complementation assay" for T4 tail-fiber assembly; here appropriate mutationally deficient cell lysates were used to test for exogenously added proteins which could overcome the deficiency (81). Following on this example, we developed an *in vitro* DNA-synthesizing system which shows a requirement for the products of T4 genes 41, 44, 45, and 62, and which could therefore be used to provide an assay for their purification despite the fact that their function in replication was unknown (21,82). This system consists of a concentrated T4-infected cell lysate which, when supplied with deoxyribonucleoside triphosphates, supports a brief period of rapid DNA synthesis using the endogenous DNA as template. When such lysates are prepared from replication-defective cells, the DNA synthesis they support is specifically stimulated by addition of extracts which contain the missing gene product (82).

Using this complementation test as an assay, it has been possible to

TABLE I
Properties of the Bacteriophage T4 DNA Replication Proteins

Gene number	Molecular weight	Designation	Separate *in vitro* activities detected
32	35,000 (oligomeric aggregates)	T4 "DNA-unwinding" or "DNA-melting" protein	Catalyzes DNA denaturation and DNA renaturation; stimulates T4 DNA polymerase
41	58,000	?	DNA-dependent GTPase (and ATPase); long DNA single-strands optimal
43	110,000	T4 DNA polymerase	Template-dependent dNTP polymerization with associated 3′ to 5′ proofreading exonuclease
44 | 62	4 × 34,000 2 × 20,000	?	45 protein stimulated, DNA-dependent ATPase; short DNA single strands optimal Stimulates T4 DNA polymerase in reaction requiring 45 protein and ATP hydrolysis
45	2 × 27,000	?	Participates in the above two reactions with 44–62 protein

purify separately the gene 41 protein, the gene 45 protein, and a tight complex of gene 44 and 62 proteins to greater than 95% homogeneity in an active form (4,83). In retrospect, the development of successful purification procedures required an activity assay: nearly half of the fractionation techniques attempted purified the relevant proteins (detected by double-label counting) in a functionally denatured form and were therefore avoided.

Each of the six T4 replication proteins can now be obtained in greater than 10-mg quantities, free of detectable endonuclease activity and nearly homogeneous as judged by polyacrylamide gel electrophoresis. Table I summarizes some of their relevant properties, as we presently understand them.

As previously reported (21), the gene 44 and 62 proteins are isolated as a tight complex (180,000 daltons) which appears to contain four molecules of 44 protein (34,000 daltons each) and two molecules of 62 protein (20,000 daltons each). Thus far, these two different polypeptide chains have been separable only by agents [such as sodium dodecyl sulfate (SDS) and guanidine] which denature proteins. The other four replication proteins contain only a single type of subunit, although the gene 32 protein appears to exist mainly as a 6–8 S ag-

gregate, and the gene 45 protein as a dimer. The gene 41 protein appears to be a monomer as judged by its sedimentation rate; however, there are other indications that in its functional state it may be at least a dimer (see below).

Note that the existence of these replication proteins as oligomers raises the possibility of multisite activity and/or cooperativity, and that this may be important in the replication process.*

PARTIAL REACTIONS

The most striking *in vitro* activities of the T4 replication proteins require that all six of them be present simultaneously, and we shall discuss physical evidence that they form a multienzyme complex. Since analysis of individual functions in such a multicomponent system is difficult, it is instructive to study in detail as many "partial reactions" as possible. Here we include the DNA polymerase and proofreading exonuclease activities of 43 protein (which utilize single-stranded DNA templates), and the denaturation (and renaturation) of DNA catalyzed by 32 protein. The properties of these two proteins have been previously discussed in the literature to such an extent that they need not be reviewed again here (see refs. 1–3,39,51,67,71).

The 44–62 Protein. Included in Table I is the fact that the 44–62 protein complex is an ATPase which can hydrolyze either ribo- or deoxyribo-ATP to the corresponding diphosphate and inorganic phosphate. As shown in Fig. 3, this ATPase reaction is strongly stimulated by DNA (Fig. 3A) and by addition of 45 protein (Fig. 3B). Shown in Fig. 4 is the fact that the same gene 45 and 44–62 proteins facilitate T4 DNA polymerase (gene 43 protein) utilization of long single-stranded DNA templates. This stimulatory reaction appears to be completely dependent on the hydrolysis of the β-γ bond of ATP, since ATP removal, or its substitution by the nonhydrolyzable analog

* One intriguing possibility is that the DNA strands near the fork are convoluted in such a way as to bring leading and lagging strand DNA polymerase molecules into close juxtaposition, and that bifunctional replication proteins join these two molecules together (4). Another possibility is that the enzyme involved in making the primer for *de novo* chain starts acts on a palindromic DNA recognition sequence (84,85), and is a dimer with a twofold axis of symmetry so as to be able to prime both lagging and leading strand pieces simultaneously. Such coupling between the two sides of a replication fork might be designed to prevent imbalanced synthesis, and is one way to account for the observation that short DNA pieces can be recovered from both strands after pulse labeling (42,86).

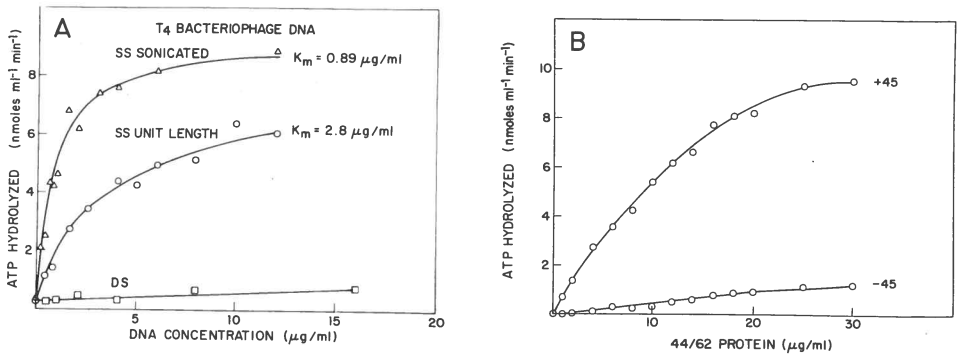

Fig. 3. Characteristics of the 44–62 protein ATPase reaction. (A) Dependence on DNA. Varying amounts of double-stranded (DS) and single-stranded (SS) sonicated and unit-length T4 DNA's were incubated with 8 μg of 44–62 protein and 10 μg of 45 protein per milliliter in a mixture containing 10 m*M* Tris-HCl pH 7.4, 25 m*M* KCl, 5 m*M* $MgCl_2$, 1 m*M* dithiothreitol (DTT), 75 μg of BSA per milliliter, and 0.25 m*M* γ-labeled [^{32}P]ATP. After 5 min at 37°C, aliquots were mixed with charcoal and analyzed for nonadsorbable [^{32}P]phosphate. (B) Dependence on 45 protein. Varying amounts of 44–62 protein with and without 10 μg of 45 protein per milliliter were incubated for 5 min at 37°C in a mixture containing 10 m*M* Tris-HCl, pH 7.4, 25 m*M* KCl, 5 m*M* $MgCl_2$, 1 m*M* DTT, 400 μg of BSA and 35 μg of sonicated single-stranded calf thymus DNA per milliliter, and 0.25 m*M* γ-labeled [^{32}P]ATP. Reactions were analyzed for liberated [^{32}P]phosphate as in (A).

AMPP(NH)P, eliminates the effect of the 44–62 and 45 proteins on polymerase activity (Fig. 4).

Analysis by classical methods reveals that all the DNA product made in the experiment in Fig. 4 is template-linked (both stimulated and control reactions), as expected for 3′-OH template end priming; thus, no new primer sites are being generated. Instead, the stimulation of DNA synthesis turns out to be due to the fact that the T4 DNA polymerase normally continues along the DNA template for a short distance before falling off, and the accessory proteins (44–62 and 45) increase this distance. In theory, either increasing K_F (the average number of nucleotides polymerized per second by a bound DNA polymerase molecule) and/or enhancing the average sticking time of an active polymerase will increase the observed rate of DNA synthesis, providing that the template length is longer than the typical sticking-distance for polymerase alone. Direct measurements of the rate of polymerization in the stimulated and unstimulated reactions obtained with 2- to 20-sec incubations, to be described in detail elsewhere (87), allows us to estimate for T4 DNA polymerase alone a microscopic polymerization rate (K_F) of ~300 nucleotides per second

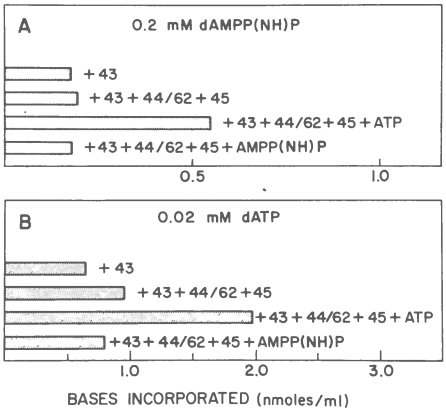

Fig. 4. ATP hydrolysis is required for the stimulation of DNA polymerase by gene 44–62 and 45 proteins. DNA synthesis reactions contained 10 m*M* Tris-HCl, pH 7.4, 25 m*M* KCl, 5 m*M* $MgCl_2$, 1 m*M* DTT, 75 μg of BSA per milliliter, 0.1 m*M* each dCTP and dGTP, 0.11 m*M* [^{3}H]dTTP, and 5 μg/ml unit-length T7 single-stranded DNA. (A) DNA polymerized using 0.2 m*M* dAMPP(NH)P as the fourth deoxyribonucleoside triphosphate substrate for DNA polymerase (this dATP analog is not hydrolyzable by 44–62 protein). (B) DNA polymerized using 0.02 m*M* dATP. Where added, the concentrations of ATP and the nonhydrolyzable ATP analog AMPP(NH)P were 0.25 m*M*. The proteins were added where designated at the following concentrations (per milliliter): 1 μg of 43, 1 μg of 44–62, and 5 μg of 45 protein. Assays were incubated 5 min at 37°C, and acid-insoluble radioactivity was determined by standard techniques. At 0.2 m*M* dATP, the dATP hydrolysis by the 44–62 protein will completely mask the ATP hydrolysis requirement seen in (A) and (B), and in (B) a slight stimulation by the 44–62 and 45 proteins is seen even at 0.02 m*M* dATP for the same reason.

and a sticking-time of 2–3 sec under our conditions (see Fig. 2B). Addition of 44–62 and 45 proteins increases the K_F to about 800 nucleotides per second, along with a slight increase in sticking-time (to about 5 sec).

Note that these values mean that T4 DNA polymerase runs at close to the *in vivo* rate of polymerization (estimated at 1000 nucleotides per second) in our simple *in vitro* system, even without the further stimulation expected from addition of 32 protein (62). We conclude, therefore, that *the remaining proteins of the replication apparatus need only increase the processiveness (by tying down polymerase) to account for the observed speed of in vivo fork movement.*

The fact that hydrolysis of ATP is absolutely required to observe an effect of 44–62 and 45 proteins on polymerase rate and processiveness might suggest that these accessory proteins act as a "DNA-walking machine," generating a motive force which helps drive the polym-

erase down its template (88–90). However, measurements of the stoichiometry of hydrolysis reveal that less than one ATP is hydrolyzed per every ten deoxyribonucleotides polymerized in these reactions (unpublished results of J. Piperno). Thus, if a walking machine exists, it must either operate with an unusually large step distance (greater than 10 template nucleotides traversed per hydrolysis event), or generate its motive force only at those occasional sites on the template at which the polymerase balks [e.g., at tight intrastrand hairpin helicies (91,92)]. Another mechanism needing very little hydrolysis is one in which ATP is utilized to drive an ordered allosteric change in the 44–62 protein, which then "irreversibly" assembles or modifies a 45 protein-polymerase complex. This mechanism provides an alternative explanation (besides the 44–62 assembly hypothesis mentioned previously) for why the available temperature-sensitive mutants in gene 44 continue their replication if shifted to the nonpermissive temperature once DNA synthesis has started (19), while at least some gene 45 temperature-sensitive mutants stop replicating instanteously. However, such a role for the large 44–62 complex (180,000 daltons) seems unpalatable at the present state of our knowledge, inasmuch as we lack a clear theoretical justification for requiring ATP in this type of function. In contrast, the theoretical justification for using ATP-driven allosteric changes to generate concerted protein movements (88) and/or to increase templating fidelity (57,58) are well established.

Possible additional insight into the mechanism of action of 44–62 protein can be gained by examining the relative abilities with which different DNA's activate its DNA-dependent ATP hydrolysis. Such studies reveal no DNA sequence specificity, but strongly suggest that DNA chain ends bind and activate the 44–62 protein complex much more effectively than do DNA middles (about 10^4-fold better on a per nucleotide basis; Fig. 3A and unpublished results of J. Piperno). This is expected if, as seems reasonable, the 44–62 protein recognizes the same DNA chain end that the DNA polymerase uses as a primer.

The Gene 45 Protein. The 45 protein must play an important role in the activities of 44–62 protein, since both ATPase and polymerase stimulation require its presence at about 5 μg/ml. In addition to its role in DNA replication, genetic studies reveal that the 45 protein is absolutely required for late mRNA transcription (93), and it has been found to bind tightly to the T4-modified form (but not the unmodified form) of the host RNA polymerase (94). An understanding of the role of 45 protein in transcription would be extremely useful for deter-

mining its function in replication; unfortunately late T4 mRNA transcription is sufficiently complex as to discourage such studies as an easy approach to replication function [for review, see Wu *et al.* (95)]. Tests for an activity of 45 protein alone (as a nucleoside triphosphatase, polymerase, deoxyribonuclease, and/or ribonuclease), have thus far all been negative. Moreover, 45 protein did not have any obvious effect on the *in vitro* RNA synthesis catalyzed by RNA polymerase in preliminary experiments (unpublished results of M. Davies).

The Gene 41 Protein. The gene 41 protein, like the gene 44–62 protein, is an enzyme capable of splitting nucleoside triphosphates to nucleoside diphosphates and inorganic phosphate. In both cases, single-stranded DNA is required to activate this process. However, whereas the 44–62 protein works best with short DNA strands (suggesting a preference for DNA chain ends, as described), the gene 41 protein hydrolyzes nucleoside triphosphates more and more poorly as intact T4 DNA chains are reduced in size by shearing (unpublished results of C. C. Liu and C. F. Morris). Moreover, there is an exponential increase in hydrolysis rate as the 41 protein concentration is raised. Both these observations suggest the possibility that a dimer of 41 protein with two DNA sites simultaneously bound is the active species for nucleoside triphosphate hydrolysis, although other explanations are clearly possible.

The preferred nucleotide for hydrolysis by 41 protein is rGTP, with dGTP, rATP and dATP also being quite active substrates (rates being at least one-fifth that with rGTP). However, what role (if any) this hydrolysis plays in DNA replication is unclear. First of all, the K_m observed for rGTP in the reaction is about 3 m*M*, with a V_{max} representing only about 30 molecules of nucleotide hydrolyzed per minute per 41 protein molecule present. Second, if enough 32 protein is present to cover the single-stranded DNA added, the DNA stimulation of hydrolysis disappears. All single-stranded DNA within the cell is probably covered with 32 protein, and intracellular concentrations of rGTP are only about 1 m*M* (18). Therefore, we tend to view this partial reaction more as a probe of 41 protein conformation and affinities than as a reaction of certain biological significance. For example, none of the T4 replication proteins effect the low level of nucleoside triphosphate hydrolysis observed in the absence of DNA, and only 32 protein effects the DNA-stimulated reaction. We therefore believe that the 32 protein inhibition is due to its effect on DNA conformation, and find no evidence that either it or the other T4 replication proteins form a tight complex with the 41 protein.

COMPLEX FORMATION BETWEEN REPLICATION PROTEINS

For near-maximal stimulation of DNA synthesis, the *in vitro* complementation assay requires that the following concentrations of the purified replication proteins be added to each appropriate mutant-infected crude extract (per milliliter): 5 μg of 41 protein, 0.3 μg of 43 protein, 0.5 μg of 44–62 protein, and 2 μg of 45 protein. For comparison, typical concentrations used in the *in vitro* replication system containing only purified proteins are 50 μg of 41 protein, 1 μg of 43 protein, 10 μg of 44–62 protein, 10 μg of 45 protein, and 200 μg of 32 protein per milliliter. Even these latter concentrations are less than or equal to the estimated *in vivo* concentrations, so that there is no reason to doubt that each protein is functioning much as it does within the cell. The fact that relatively high concentrations of some of the proteins are required (in large molar excess over polymerase) implies either that these proteins are needed in many copies per replication fork, or that their continuous rapid association and dissociation from the replicating structure is required for efficient DNA synthesis.

Evidence that the high protein concentrations required may be in part due to the weakness of complex formation comes from our attempts to demonstrate directly that a multienzyme complex of the replication proteins exists. Conventional experiments in which the replication proteins are mixed together and cosedimented through sucrose gradients fail to reveal a rapidly sedimenting complex. Thus, a method capable of detecting weaker complexes had to be devised. For this purpose, we utilize small amounts of single-stranded DNA cellulose as template for the replication proteins, and then rapidly collect the immobilized DNA molecules with their associated proteins as a powder on Millipore filters. After a brief wash on the filter (< 1 min), the DNA-cellulose is scraped into an SDS-containing buffer (to solubilize bound proteins) and analyzed for Coomassie Blue-stainable protein bands following polyacrylamide gel electrophoresis.

Such experiments have revealed the protein–protein and protein–DNA interactions drawn as solid lines in Fig. 5 (unpublished results of M. Davies). Only the gene 43 and gene 32 proteins bind tightly enough to DNA under our DNA synthesis conditions to be isolated directly as a DNA–protein complex. However, once the 32 protein is bound to the single-stranded DNA, the 44–62 protein complex will bind. Moreover, the gene 45 protein is obtained in the complex if, and only if, the 44–62 protein is present.

Thus far we have not been able to detect any binding in these experiments of the gene 41 protein, even though the DNA stimulation

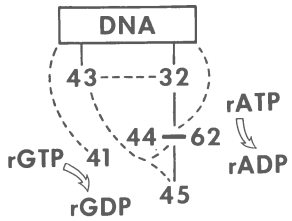

Fig. 5. Schematic representation of known protein–protein and protein–DNA interactions in the T4 replication apparatus. Solid lines represent interactions inferred from experiments in which protein complexes on DNA-cellulose were isolated by rapid filtering of reaction mixes, as described in the text. The dashed line connecting 32 and 43 represents a direct protein–protein interaction detected by Huberman *et al.* (62). Other dashed lines are inferred from the DNA-dependence of the indicated ATPase and GTPase reactions, and from the stimulation of the 43 protein (DNA polymerase) by the 44–62 and 45 proteins (see Fig. 4 and refs. 4 and 87).

of its GTPase activity reveals that at least a weak DNA interaction must exist (dashed lines). Also shown in Fig. 5 as dashed lines are interactions which we can infer from several other types of experiments (see Fig. 5 legend). With additional interactions probably remaining to be discovered, it seems quite reasonable to view the T4 replication apparatus as a large multienzyme complex. This conclusion is supported by functional studies which reveal synergistic effects requiring all six replication proteins (see below).

CONCERTED PARTIAL REACTIONS CARRIED OUT BY THE SIX T4 REPLICATION PROTEINS

The model illustrated earlier in Fig. 1B predicts that two DNA polymerase molecules work at any one time in the replication fork. The 44–62 protein ATPase might be traveling along with one or both of these polymerase molecules, and, in the simplest view, another replication protein (e.g., gene 41 protein?) might function solely to synthesize an RNA primer. To test these expectations, we set out to study the predicted reactions on the two sides of the replication fork in Fig. 1B separately.

As a model for lagging-strand synthesis, we tested for ribonucleoside triphosphate-dependent *de novo* chain initiations on a T4 single-stranded DNA template (Fig. 6A), analogous to the *E. coli dna*G protein-catalyzed priming reaction observed on ϕX174 and G4 phage DNA templates (48,49). As a model for leading-strand synthesis, a rapid and efficient strand-displacement reaction on a T4 DNA

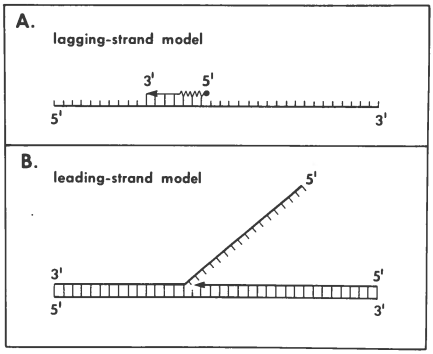

Fig. 6. Predicted "half-replication fork" partial reactions based on the mechanism in Fig. 1B. (A) Lagging-strand model: a ribonucleoside triphosphate-dependent *de novo* chain start on a single-stranded template. (B) Leading-strand model: efficient strand-displacement synthesis on a double-stranded template.

duplex was sought, as schematically illustrated in Fig. 6B. It turns out that the T4 replication proteins catalyze both of these reactions in Fig. 6 in an efficient manner. However, we were surprised to find that both the leading-strand and lagging-strand model reactions require all six of the T4 replication proteins for maximal DNA synthesis.

A "Lagging-Strand" Model Reaction. Incorporation data strongly suggesting RNA-primed DNA chain starts *in vitro* are presented in Fig. 7. Using T4 single-strands as a template, one obtains extensive DNA synthesis which requires the presence of all six of the replication proteins. As shown, this synthesis is completely dependent upon the addition of a mixture of rCTP, rGTP, and rUTP, in addition to the rATP present. Alkaline sucrose gradient sedimentation of denatured DNA strands reveals that the small amount of product made in the absence of either rCTP, rGTP, and rUTP (not shown) or 32 protein (Fig. 8B) cosediments with the template used,* as expected for template-priming by foldback at the 3′-OH end (25). In contrast, essentially all of the product made at early times in the complete reaction sediments with a mean chain length much smaller than the template (Fig. 8A). These results demonstrate that ribonucleoside triphosphate-dependent *de novo* initiation of new DNA chains is occurring. This reaction proceeds under conditions which approximate intracellular

* Note that the template used in the experiment in Fig. 8 was T4 single-stranded DNA which had been sheared approximately in fifths so as to cosediment with the T7 DNA marker.

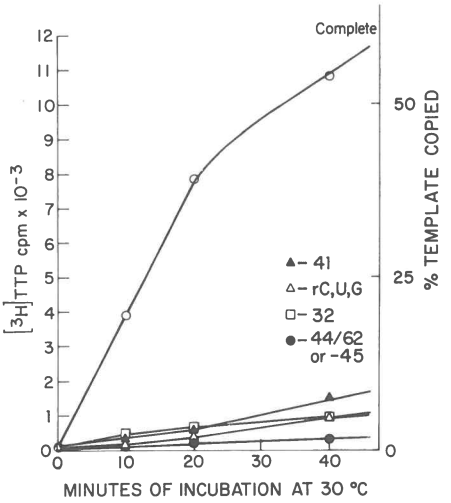

Fig. 7. Time course of ribonucleotide-dependent DNA synthesis on a single-stranded T4 DNA template. The "complete" reaction components consist of 43 protein, 1 μg/ml; 41 protein, 80 μg/ml; 44–62 protein, 8 μg/ml; 45 protein, 25 μg/ml; 32 protein, 125 μg/ml; dNTP's 0.1 m*M* each, rATP, 0.5 m*M*; nuclease-free BSA, 50 μg/ml; $MgSO_4$, 5 m*M*; KCl, 25 m*M*; Tris-HCl (pH 7.4), 20 m*M*; dithriothreitol, 0.5 m*M*; alkali-denatured T4 single-stranded DNA 10 μg/ml; and rCTP, rUTP, rGTP, 0.1 m*M* each. [^{3}H]TTP was used to label DNA products made. For the incomplete reactions, only the specified component was deleted. At the specified times, aliquots were removed, and incorporation into acid-insoluble material was measured.

concentrations of salts, and with protein concentrations not greater than those estimated to be present *in vivo*. Therefore, it appears to be biologically relevant.

The sites at which new DNA chains can be generated are not restricted to T4 DNA. It appears that a ribo-dependent reaction similar to that shown in Fig. 7 proceeds on many DNA templates (e.g., single-stranded T4, fd, and G4 bacteriophage DNA's), but that not all DNA's are active (e.g., single-stranded T7 and synthetic DNA's). Moreover, studies show that omitting a single-ribonucleoside triphosphate (C, U, or G) diminishes the ribo-dependent DNA synthesis observed by an amount which depends on the particular DNA template used; thus, for example, omission of rCTP eliminates only two-thirds of the ribo-dependent reaction on an fd DNA template, while completely eliminating (>95%) all ribo-dependent synthesis on a T4 DNA template (unpublished results of C. F. Morris). In both cases, physiologically reasonable concentrations of rCTP are needed for half-maximal stimulation (20–60 μM). Thus, it seems likely that the T4

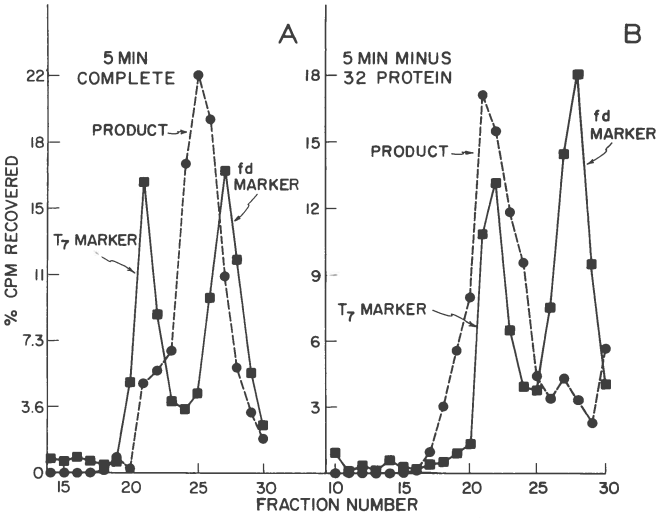

Fig. 8. "Complete" reactions on T4 single-stranded templates produce newly started DNA strands. Reaction conditions were essentially as in Fig. 7, except that the T4 DNA strands used as template had been sheared so as to cosediment with the T7 DNA marker. After incubation for 5 min at 30°C, 20 m*M* EDTA was added to stop the reaction. The samples were then mixed with ^{14}C-labeled DNA standards (solid lines) and centrifuged through 5–30% alkaline sucrose gradients (0.8 *M* NaCl, 0.2 *M* NaOH, 5 m*M* EDTA). (A) Complete reaction: product much smaller than template strands. (B) Reaction with only 32 protein omitted. An indistinguishable profile is obtained with rCTP, rUTP, rGTP omitted instead, or if T4 DNA polymerase is used alone. In these reactions, the product cosediments with template strands from the earliest times, as expected for template-primer chain starts.

replication proteins are recognizing some special sequence or structure on the DNA single-strand and generating a short RNA primer there (3,50). To explain why all six T4 proteins are required, we suggest that a special multienzyme complex is formed between them, and that this is needed to generate the correct environment for the postulated synthesis and/or utilization of the primer RNA.

Note that the regular DNA-dependent RNA polymerase utilized for transcription is not involved in our reactions: we are unable to detect the activity of this enzyme in any of our purified proteins; moreover, the ribonucleoside triphosphate requirement for DNA synthesis is retained even after addition of high levels of rifampicin (10 μg/ml) or antibody to T4-modified RNA polymerase.

A "Leading-Strand" Model Reaction. The T4 DNA polymerase by itself will not utilize nicked double-helical DNA's as a template, as it

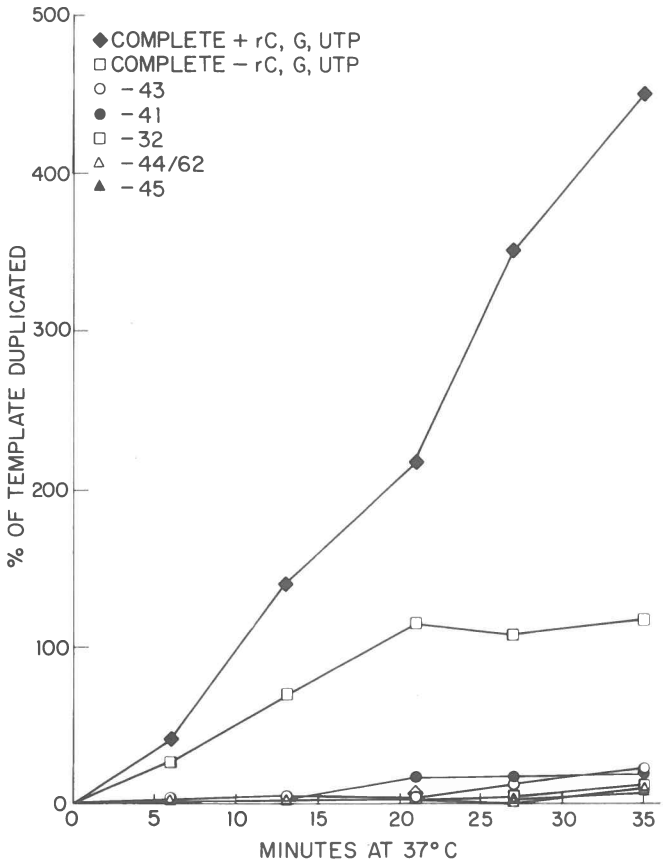

Fig. 9. Time course of synthesis on a double-stranded T4 DNA template. The complete reaction components were essentially the same as those specified in Fig. 7, the major changes being addition of K^+ to 67 m*M*, substitution of 4 μg of intact T4 double-stranded DNA per milliliter as template, and removal of rCTP, rGTP, and rUTP. For the incomplete reactions, only the specified component was deleted. Note that only the one reaction indicated contained the three ribonucleoside triphosphates.

is unable to copy base-paired regions (22,62). However, in 1974 Nancy Nossal made the important observation that addition of high concentrations of 32 protein allows a limited amount of DNA synthesis to begin on such templates (80). Much of the product DNA made is reversibly denaturable, re-forming hairpin-type double-helical structures after heating and quick cooling (80). This is the expected result if branch migration (96) repeatedly interrupts the polymerization process by knocking the DNA polymerase off the growing chain end (4,97,98).

When additional T4 replication proteins are added to this system, such interuptions are apparently prevented. One observes, instead, the rapid and efficient strand-displacement reaction expected for leading strand synthesis. All double-stranded DNA's thus far tested function as templates for this synthesis, including the DNA from bacteriophages T4, T7, lambda, PM2, and G4; *E. coli;* SV40 virus; and monkey cells. The protein requirements for extensive DNA synthesis on a double-stranded T4 DNA template are illustrated in Fig. 9. Once again, all six T4 replication proteins are required; note that an amount of product equal to the input template is made within 20 min at 37°C. Alkaline sucrose gradient analysis demonstrates that long continuous DNA strands are being synthesized, and a test for S1 nuclease susceptibility of product and template reveals that more than 99% of each is fully denaturable at all extents of reaction (unpublished results of N. K. Sinha).

Further addition to this system of a mixture of rCTP, rGTP, and rUTP can be seen to stimulate the observed rate of DNA synthesis about twofold (Fig. 9), probably reflecting a ribo-dependent lagging-strand synthesis which ensues (see below).

COMPLETE REACTIONS

When leading-strand and lagging-strand reactions proceed simultaneously, they should generate a replication fork resembling that described earlier in Fig. 1B. This is readily testable when circular templates, such as PM2 or fd bacteriophage DNA, are used. For example, following *de novo* chain initiation on a fd template, polymerization should proceed around the circular, single-stranded DNA molecule to yield a circular double helix with one strand-specific interruption. Further synthesis on these double-stranded circles should induce strand displacement and generate a "rolling circle" mode of DNA synthesis (99), as schematically illustrated in Fig. 10. Because the initial template is exclusively (+)-strand, the 32-protein coated, single-stranded regions (formed as intermediates on these rolling-circle tails) are expected to be located exclusively on the (−)-strand. These regions therefore cannot renature with each other to confuse the analysis. [In contrast, renaturation between complementary rolling-circle tails can readily occur when the initial template is a circular double-stranded DNA molecule (such as PM2 or SV40), and different DNA circles begin synthesis so as to spin out single-strands of opposite polarity.]

The only components used in the "complete" reaction on fd DNA

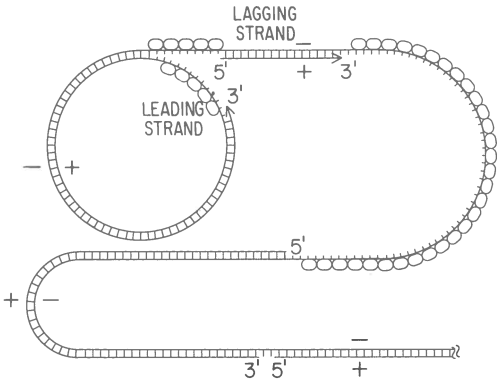

Fig. 10. Schematic representation of a "rolling-circle" model of DNA synthesis, initiated by a *de novo* start on a single-stranded circular viral DNA [(+)-strand]. Gene 32 protein is shown bound to the single-stranded DNA exposed by fork movement [(−) strand].

templates are the six replication proteins, rNTP's, dNTP's, bovine serum albumin (BSA) as a protective agent, Mg^{2+}, and salts. Beginning with 3 μg of fd DNA template per milliliter, we have been able to make as much as 80 μg of product DNA per milliliter in a 30-min incubation at 37°C. If any of the T4 proteins (44–62, 43, 41, or 45) is left out of the reaction mixture, one finds less than 1% the amount of DNA synthesis seen in the complete reaction. Therefore, each is clearly essential.

There is a ten-to twentyfold decrease in synthesis when rUTP, rGTP, and rCTP are omitted and a threefold decrease when rCTP alone is omitted. This strong requirement for ribonucleotides suggests that an RNA primer is being made. However, we have not yet been able to detect radioactive ribonucleotides incorporated into polymer in experiments in which the fate of high specific-activity rCTP has been monitored by thin-layer chromatography on PEI-cellulose. At the same time, we have not found a significant ribonuclease or ribonuclease H activity in the system which will act on exogenously added synthetic RNA substrates. It is nevertheless possible that an RNA primer is being both made and degraded in our system, perhaps by a cis-acting RNase activity intrinsic to the replication complex.

If the "rolling-circle" scheme in Fig. 10 is actually being carried out by our system, at late times in the course of the reaction long single-stranded product DNA molecules (multiple fd genome lengths) should appear and these should be predominantly (−)-strand. In contrast, the shorter product molecules (less than one fd genome length)

should be predominantly (+)-strand. Moreover, template DNA strands should remain circular and intact throughout the course of the reaction.

In order to test these predictions, the reaction products were sedimented through a neutral sucrose gradient to isolate the largest (most rapidly sedimenting) rolling circles. These molecules were then denatured and sedimented through an alkaline sucrose gradient as separated single strands. The original ^{3}H-labeled template strands all remained at the position expected for intact circles, as anticipated. Fractions across this alkaline gradient were pooled into size classes and aliquots were either self-hybridized (control) or hybridized with an excess of fd viral DNA [(+)-strand]. S1 nuclease digestion was used to determine how much double-stranded DNA had formed. The results showed that more than 90% of the single-stranded DNA larger than 12×10^6 daltons is (−)-strand, while 75% of the single-stranded DNA smaller than 2×10^6 daltons is (+)-strand, which is again consistent with Fig. 10 (see ref. 83).

In order to examine the DNA products by electron microscopy, aliquots were removed from reactions on an fd DNA template and examined after cytochrome speading on a formamide hypophase (100). This type of analysis is not strictly quantitative, since a substantial fraction of DNA molecules are not well enough spread to be scored, and thus a special minor class of product can easily be overlooked. Moreover, the shear forces generated by spreading appear to fragment some of the longer DNA molecules (especially those containing substantial single-stranded interruptions). Therefore, although simple linear double-stranded products are seen, it is not clear that they were present as such before spreading.

Despite these qualifications, it is clear that a major class of product DNA has precisely the geometry predicted from the replication mechanism suggested in Fig. 10. Double-stranded DNA circles of fd length are the predominant structures seen at early times in the reaction. On further incubation, long double-stranded DNA tails are seen attached to some of these circles (Fig. 11). As expected from the rolling-circle model, the DNA at the point of attachment of circle and tail is nearly always single-stranded. Moreover, double- and single-stranded regions frequently alternate in this region, as in the schematic diagram in Fig. 10 (4,83). We therefore conclude that the (+)-strand complement made on the single-stranded tail is synthesized by a discontinuous mechanism, as predicted (see lagging strand in Fig. 1B). Similar structures to those shown in Fig. 11 are seen whether fd or G4 bacteriophage DNA's (single-stranded), or PM2 bacteriophage or

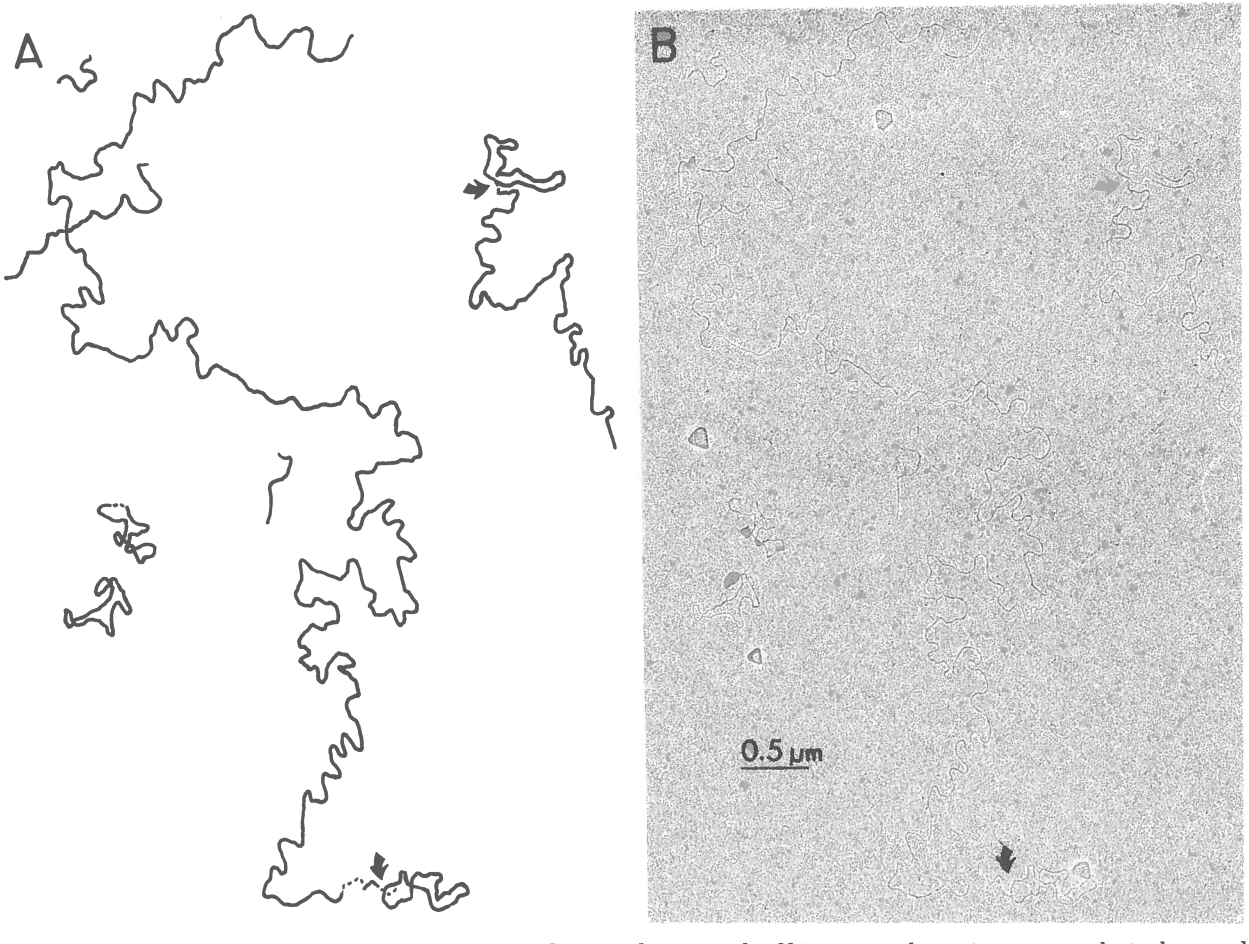

Fig. 11. Electron microscopic visualization of reaction products made on circular fd DNA templates. A, tracing with single-strands dotted; B, actual micrograph. Arrows denote the origin of rolling-circle tails. Note that two typical rolling-circle molecules are present in this field,

SV40 virus DNA's (double-stranded) provide the initial circular template. In all cases, although an amount of DNA equivalent to multiple copies of the template is made in the first 30 min of incubation, only a small percentage of double-stranded circular molecules appear as circles with attached tails. In a typical electron microscopic field, many of these tails are greater than 50 μm in length (10^8 daltons). Thus the mechanism of polymerization at these replication forks appears to be highly processive, and we can calculate a rate of polymerization per active rolling circle of at least 800 nucleotides per second at 37°C. Note that this is close to the rate at which the T4 replication fork moves *in vivo* (10).

THE FIDELITY OF SYNTHESIS

In the complete reactions just described, all (>99%) of the product DNA made is denatured by 3 min of heating at 100°C followed by rapid cooling, as judged by its becoming sensitive to S1 endonuclease degradation. As a control, when nitrous acid cross-linked DNA molecules are added to the same reaction mixtures, they remain S1 nuclease-resistant following such heat treatment. Thus, our DNA products do not contain the hairpin inversions found when either *E. coli* DNA polymerase I (97,98) or T4 DNA polymerase plus 32 protein (80) utilize double-stranded DNA as template. When the DNA product made on a circular G4 DNA template was treated with several different restriction endonucleases, a pattern of DNA fragments indistinguishable from that expected for natural G4 DNA was obtained, as judged by agarose gel electrophoresis. Moreover, transfection assays revealed that this product DNA was highly infectious (unpublished results of N. K. Sinha). We conclude therefore that DNA templates are copied with a high degree of fidelity in our *in vitro* system, despite the extensive amount and rapid rate of DNA synthesis observed.

FUTURE DIRECTIONS

It seems likely that special DNA sequences or structures unique to the T4 genome are required for initiation of new replication forks *in vivo*. In contrast, our *in vitro* T4 system appears to be able to establish replication forks almost equally well on most double-helical DNA

molecules. Although not yet proved, it seems likely that these forks can arise by strand displacement from a nick, followed by *de novo* initiation of complementary strand synthesis on this displaced single-strand. Other modes of replication fork initiation may also occur.

Once begun, the replication fork propagation reaction seen *in vitro* closely mimics that found *in vivo*. Since we have available a completely defined system in which functioning replication forks are formed, by raising and lowering the concentration of individual proteins, nucleoside triphosphates, and salts, we can distort the normal mechanism in ways which should reveal its crucial parameters. The problems which need solving include discovering which proteins control the Okazaki-piece length on the lagging strand, and what the signals are that they recognize. Moreover, is synthesis continuous or discontinuous on the leading strand? We would also like to know which proteins affect the fidelity of synthesis and whether ATP hydrolysis is involved in their action.

For this last project, biological assays for DNA infectivity should allow the fidelity of the DNA product synthesized *in vitro* to be quantitated with great sensitivity. Circular DNA templates such as ϕX174 seem optimal for such tests, since unit-length product DNA can be recovered from the long rolling-circle tails by appropriate restriction enzyme cleavage, and then assayed directly for infectivity. With a defined mutant DNA as template, and appropriate base analogs as substrates, the reversion frequency from mutant to wild-type transfecting DNA can be measured. This should allow us to determine how well the normal discrimination mechanisms are working, as conditions are varied *in vitro*.

The field of comparative replication enzymology also promises to be a fruitful one. Bacteriophage T4 appears to replicate its DNA with only six crucial proteins, while *E. coli* may require more than ten proteins to do the same job [see ref. 101 and article by R. McMacken *et al.*, this volume (49)]. Since the two systems are likely to be closely homologous, it seems reasonable to suppose that a particular part of the *E. coli* mechanism has been short-circuited in the bacteriophage system. It has been reported that, in addition to the *E. coli* DNA unwinding protein, the four *E. coli* proteins, *dna*B, *dna*C, "n" and "i" are required to set up a special protein–DNA complex, which then acts to create a signal for *dna*G priming of a *de novo* chain start on bacteriophage ϕX174 single-stranded templates (49,101,102). In contrast, the *dna*G priming mechanism operates independently of the above four proteins on a bacteriophage G4 DNA template (49). If the T4 system

resembles the *E. coli* system acting on G4 rather than ϕX174 DNA, both systems would contain about the same number of proteins. If this analogy is correct, besides matching the gene 32 protein and the *E. coli* DNA unwinding protein, one might choose the T4 gene 41 protein as a *dna*G homolog. This would leave the 44–62, 45, and 43 protein set as an analog of the large *E. coli* DNA polymerase holoenzyme, with its four different protein subunits (49,103). If such homologies can be pinned down, a comparison of results obtained with different replication systems should greatly accelerate progress toward an understanding of the basic mechanisms involved. Also important in this regard is the bacteriophage T7 DNA replication apparatus, which may contain even fewer crucial protein components than the T4 system described here (104).

Knowledge of the properties of one DNA-unwinding protein allowed proteins of homologous function to be identified in more complex biological systems, where the mutants and genetic analyses otherwise crucial for the identification were unavailable (63,65,66). Already, it is clear that enzymes which hydrolyze ribonucleoside triphosphates in a DNA-dependent manner are likely to be intimately involved in all replication systems. Thus, for example, it might be technically feasible to find proteins in eukaryotes which are homologous to the T4 bacteriophage-induced ATPase and GTPase described above.

A major advantage of working with the T4 system is that the profligate replication of T4-infected cells, combined with the identification of overproducing phage variants (13,14,68), makes possible the isolation of each of the six replication proteins in greater than 5-mg quantities starting from only 100 g of infected cells. Thus, there is no reason why these proteins cannot be obtained in large enough quantities for sequencing, chemical modification studies, and crystallization. Only when the detailed three-dimensional structures are available for all of these proteins can one hope to attain the type of chemical understanding which should provide the ultimate answers to the mechanistic problems raised in this review.

ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health and the American Cancer Society and benefited greatly over the years from the skillful technical assistance of Frank Amodio, Linda Frey, Judy Goldberg, Linda McAfee, and Mei Lie Wong (in chronological order).

REFERENCES

1. Klein, A., and Bonhoeffer, F. (1972) *Annu. Rev. Biochem.* **41,** 301.
2. Gefter, M. (1975) *Annu. Rev. Biochem.* **44,** 45.
3. Kornberg, A. (1974) "DNA Synthesis" Freeman, San Francisco, California.
4. Alberts, B., Morris, C. F., Mace, D., Sinha, N., Bittner, M., and Moran, L. (1975) *In* "DNA Synthesis and its Regulation" (M. Goulian, P. Hanawalt, and C. F. Fox, eds.), Vol. 3, pp. 241–269. Benjamin, New York.
5. Alberts, B. (1971) *In* "Nucleic Acid-Protein Interactions-Nucleic Acid Synthesis in Viral Infection," *Proc. Miami Winter Symp., 1971* (D. W. Ribbons, J. F. Woessner, and J. Schultz, eds.), Vol. 2, pp. 128–143, North Holland Publishing Co., Amsterdam.
6. Fuchs, E., and Hanawalt, P. (1970) *J. Mol. Biol.* **52,** 301.
7. Miller, R. C., Jr., and Buckley, P. (1970) *J. Virol.* **5,** 502.
8. Manoil, C., Sinha, N. K., and Alberts, B. (1977) *J. Biol. Chem.* (in press).
9. Traub, P., and Nomura, M. (1968) *Proc. Natl. Acad. Sci. U.S.A.* **59,** 777.
10. Werner, R. (1968) *Cold Spring Harbor Symp. Quant. Biol.* **33,** 501.
11. Gold, L. M., O'Farrell, P. Z., Singer, B., and Stormo, G. (1973) *In* "Virus Research," *Second ICN-UCLA Symp. Mol. Biol.* (C. F. Fox, and W. S. Robinson, eds.), pp. 205–225, Academic Press, New York.
12. Krisch, H. M., Bolle, A., and Epstein, R. H. (1974) *J. Mol. Biol.* **88,** 89.
13. Wiberg, J. S., Mendelsohn, S., Warner, V., Hercules, K., Aldrich, C., and Munro, J. L. (1973) *J. Virol.* **12,** 775.
14. Karam, J. D., and Bowles, M. G. (1974) *J. Virol.* **13,** 428.
15. Epstein, R. H., Bolle, A., Steinberg, C. M., Kellenberger, E., Boy de la Tour, E., Chevalley, R., Edgar, R. S., Susman, M., Denhardt, G. H., and Lielausis, A. (1963) *Cold Spring Harbor Symp. Quant. Biol.* **28,** 375.
16. Kozinski, A. W., and Felgenhauer, Z. Z. (1967) *J. Virol.* **1,** 1193.
17. Warner, H. R., and Hobbs, M. D. (1967) *Virology* **33,** 376.
18. Mathews, C. K. (1972) *J. Biol. Chem.* **247,** 7430.
19. Riva, S., Cascino, A., and Geiduschek, E. P. (1970) *J. Mol. Biol.* **54,** 85.
20. Curtis, M. J., and Alberts, B. (1976) *J. Mol. Biol.* **102,** 793.
21. Barry, J., and Alberts, B. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69,** 2717.
22. Goulain, M., Lucas, Z. J., and Kornberg, A. (1968) *J. Biol. Chem.* **243,** 627.
23. deWaard, A., Paul, A. V., and Lehman, I. R. (1965) *Proc. Natl. Acad. Sci. U.S.A.* **54,** 1241.
24. Warner, H. R., and Barnes, J. E. (1966) *Virology* **28,** 100.
25. Englund. P. T. (1971) *J. Biol. Chem.* **246,** 5684.
26. Nossal, N. G., and Hershfield, M. S. (1971) *J. Biol. Chem.* **246,** 5414.
27. Muzyczka, N., Poland, R. L., and Bessman, M. J. (1972) *J. Biol. Chem.* **247,** 7116.
28. Lehman, I. R. (1974) *In* "Methods in Enzymology" (L. Grossman and K. Moldave, eds.), Vol. 29, p. 46. Academic Press, New York.
29. Lo, K.-Y., and Bessman, M. J. (1976) *J. Biol. Chem.* **251,** 2475.
30. Marsh, R. C., Breschkin, A. M., and Mosig, G. (1971) *J. Mol. Biol.* **60,** 213.
31. Stahl, F. W. (1967) *J. Cell. Physiol.* **70,** Suppl. 1, 1.

32. Yegian, C. D., Mueller, M., Selzer, G., Russo, V., and Stahl, F. W. (1971) *Virology* **46,** 900.
33. Mufti, S., and Bernstein, H. (1974) *J. Virol.* **14,** 860.
34. Kutter, E. M., and Wiberg, J. S. (1968) *J. Mol. Biol.* **38,** 395.
35. Wu, J., and Yeh, Y. (1975) *J. Virol.* **15,** 1096.
36. Tomich, P. K., Chiu, C.-S., Wovcha, M. G., and Greenberg, G. R. (1974) *J. Biol. Chem.* **249,** 7613.
37. Chiu, C.-S., Tomich, P. K., and Greenberg, G. R. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **73,** 757.
38. Cairns, J. (1963) *J. Mol. Biol.* **6,** 208.
39. Kornberg, T., and Kornberg, A. (1974) *In* "The Enzymes" (P. D. Boyer, ed.), 3rd ed., Vol. 10, pp. 119–144. Academic Press, New York.
40. Bollum, J. F. (1975) *Prog. Nucleic Acid Red. Mol. Biol.* **15,** 109.
41. Sakabe, K., and Okazaki, R. (1966) *Biochim. Biophys. Acta* **129,** 651.
42. Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K., Kainuma, R., Sugino, A., and Iwatsuki, N. (1968) *Cold Spring Harbor Symp. Quant. Biol.* **33,** 129.
43. Sugino, A., and Okazaki, R. (1972) *J. Mol. Biol.* **64,** 61.
44. Gellert, M. (1967) *Proc. Natl. Acad. Sci. U.S.A.* **57,** 148.
45. Richardson, C. C., Masamune, Y., Live, T., Jacquemin-Sablon, A., Weiss, B., and Fareed, G. (1968) *Cold Spring Harbor Symp. Quant. Biol.* **33,** 151.
46. Lehman, I. R. (1974) *In* "The Enzymes" (P. D. Boyer, ed.), 3rd ed., Vol. 10, pp. 237–260. Academic Press, New York.
47. Inman, R., and Schnös, M. (1971) *J. Mol. Biol.* **56,** 319.
48. Bouché, J.-P., Zechel, K., and Kornberg, A. (1975) *J. Biol. Chem.* **250,** 5995.
49. McMacken, R., Bouché, J.-P., Rowen, L., Weiner, J., Ueda, K., Thelander, L., McHenry, C., and Kornberg, A., this volume.
50. Reichard, P., Eliasson, R., and Söderman, G. (1974) *Proc. Natl. Acad. Sci. U.S.A.* **71,** 4901.
51. Alberts, B. (1973) *In* "Molecular Cytogenetics" (B. Hamkalo and J. Papaconstantinou, eds.), pp. 233–251. Plenum, New York.
52. Drake, J. W. (1969) *Nature (London)* **221,** 1132.
53. Brutlag, D., and Kornberg, A. (1972) *J. Biol. Chem.* **247,** 241.
54. Topal, M. D., and Fresco, J. R. (1976) *Nature (London)* **263,** 285.
55. Wickner, S., and Hurwitz, J. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72,** 3342.
56. Wickner, W., and Kornberg, A. (1973) *Proc. Natl. Acad. Sci. U.S.A.* **70,** 3679.
57. Hopfield, J. J. (1974) *Proc. Natl. Acad. Sci. U.S.A.* **71,** 4135.
58. Ninio, J. (1975) *Biochimie* **57,** 587.
59. Oishi, M., Yoshikawa, H., and Sueoka, N. (1964) *Nature (London)* **204,** 1069.
60. Helmstetter, C. E. (1969) *Annu. Rev. Microbiol.* **23,** 223.
61. Alberts, B. M., and Frey, L. (1970) *Nature (London)* **227,** 1313.
62. Huberman, J. A., Kornberg, A., and Alberts, B. M. (1971) *J. Mol. Biol.* **62,** 39.
63. Sigal, N., Delius, H., Kornberg, T., Gefter, M. L., and Alberts, B. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69,** 3537.
64. Reuben, R. C., and Gefter, M. L. (1973) *Proc. Natl. Acad. Sci. U.S.A.* **70,** 1846.
65. Banks, G. R., and Spanos, A. (1975) *J. Mol. Biol.* **93,** 63.
66. Herrick, G., Delius, H., and Alberts, B. (1976) *J. Biol. Chem.* **251,** 2142.
67. von Hippel, P. H., Jensen, D. E., Kelly, R. C., and McGhee, J. D., this volume.
68. Gold, L., Lemaire, G., Martin, C., Morrissett, H., O'Conner, P., O'Farrell, P., Russel, M., and Shapiro, R., this volume.
69. Abdel-Monem, M., and Hoffmann-Berling, H. (1976) *Eur. J. Biochem.* **65,** 431.

70. Riggs, A., Bourgeois, S., and Cohn, M. (1970) *J. Mol. Biol.* **53,** 401.
71. Alberts, B. (1974) *In* "Mechanism and Regulation of DNA Replication" (A. R. Kolber and M. Kohiyama, eds.), pp. 133–148. Plenum, New York.
72. Alberts, B. M., Amodio, F. J., Jenkins, M., Gutmann, E. D., and Ferris, F. L. (1968) *Cold Spring Harbor Symp. Quant. Biol.* **33,** 289.
73. Alberts, B., and Herrick, G. (1971) *In* "Methods in Enzymology" (K. Moldave and L. Grossman, eds.), Vol. 21, p. 198. Academic Press, New York.
74. Alberts, B. M. (1970) *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **29,** 1154.
75. Delius, H., Mantell, N. J., and Alberts, B. (1972) *J. Mol. Biol.* **67,** 341.
76. Carroll, R. B., Neet, K. E., and Goldthwait, D. A. (1975) *J. Mol. Biol.* **91,** 275.
77. Marmur, J., and Doty, P. (1962) *J. Mol. Biol.* **5,** 109.
78. Hosoda, J., Takacs, B., and Brack, C. (1974) *FEBS Lett.* **47,** 338.
79. Moise, H., and Hosoda, J. (1976) *Nature (London)* **259,** 455.
80. Nossal, N. G. (1974) *J. Biol. Chem.* **249,** 5668.
81. Edgar, R. S., and Wood, W. B. (1966) *Proc. Natl. Acad. Sci. U.S.A.* **55,** 498.
82. Barry, J., Hama-Inaba, H., Moran, L., Alberts, B., and Wiberg, J. (1973) *In* "DNA Synthesis in Vitro" (R. D. Wells and R. B. Inman, eds.), pp. 195–214. Univ. Park Press, Baltimore, Maryland.
83. Morris, C. F., Sinha, N. K., and Alberts, B. M. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72,** 4800.
84. Cairns, J. (1973) *Br. Med. Bull.* **29,** 188.
85. Schaller, H., Uhlmann, A., and Geider, K. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **73,** 49.
86. Sternglanz, R., Wang, H. F., and Donegan, J. J. (1976) *Biochemistry* **15,** 1838.
87. Mace, D., and Alberts, B. (1977) In preparation.
88. Hill, T. L. (1969) *Proc. Natl. Acad. Sci. U.S.A.* **64,** 267.
89. MacKay, V., and Linn, S. (1974) *J. Biol. Chem.* **249,** 4286.
90. Wilcox, K., and Smith, H. (1976) *J. Biol. Chem.* **251,** 6127.
91. Doty, P., Boedtker, H., Fresco, J. R., Haselkorn, R., and Litt, M. (1959) *Proc. Natl. Acad. Sci. U.S.A.* **45,** 482.
92. Sherman, L. A., and Gefter, M. (1976) *J. Mol. Biol.* **103,** 61.
93. Wu, R., and Geiduschek, E. P. (1975) *J. Mol. Biol.* **96,** 513.
94. Ratner, D. (1974) *J. Mol. Biol.* **88,** 373.
95. Wu, R., Geiduschek, E. P., Rabussay, D., and Cascino, A. (1973) *In* "Virus Research," *Second ICN-UCLA Symp. Mol. Biol.,* (C. F. Fox, and W. S. Robinson, eds.), pp. 181–199, Academic Press, New York.
96. Lee, C. S., Davis, R. W., and Davidson, N. (1970) *J. Mol. Biol.* **48,** 1.
97. Inman, R. B., Schildkraut, C. L., and Kornberg, A. (1965) *J. Mol. Biol.* **11,** 285.
98. Masamune, Y., and Richardson, C. C. (1971) *J. Biol. Chem.* **246,** 2692.
99. Gilbert, W., and Dressler, D. (1968) *Cold Spring Harbor Symp. Quant. Biol.* **33,** 473.
100. Inman, R. B. (1974) *In* "Methods in Enzymology" (L. Grossman and K. Moldave, eds.), Vol. 29, p. 451. Academic Press, New York.
101. Wickner, S., and Hurwitz, J. (1975) *In* "DNA Synthesis and its Regulation," *ICN-UCLA Symp. Mol. Cell. Biol.* (M. Goulian, P. Hanawalt, and C. F. Fox, eds.), Vol. 3, pp. 227–240. Benjamin, New York.
102. Weiner, J. H., McMacken, R., and Kornberg, A. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **73,** 752.
103. Wickner, S., and Hurwitz, J. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **73,** 1053.
104. Hinkle, D. C., and Richardson, C. C. (1975) *J. Biol. Chem.* **250,** 5523.