STUDIES OF REPLICATION MECHANISMS WITH THE T4 BACTERIOPHAGE *IN VITRO* SYSTEM[1]

Bruce M. Alberts, Jack Barry, Pat Bedinger, Rae Lyn Burke, Urszula Hibner[2], Chung-Cheng Liu and Richard Sheridan

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

ABSTRACT. Seven T4 bacteriophage-coded proteins have been purified which are essential for replication fork movement both *in vivo* and *in vitro*. Experiments are reported which enable us to assign functions to the individual proteins and to position them at least tentatively at the fork. The seven proteins are assigned to one of the following four functional classes: helix-destabilizing protein, DNA polymerase, DNA polymerase accessory proteins, and RNA priming proteins. Comparison is made with the simpler replication system encoded by bacteriophage T7, and with the more complex replication system of the *E. coli* host. We suggest that more replication proteins are required for organisms with larger genomes, due to the fact that a greater accuracy is needed in their DNA replication process.

INTRODUCTION

We have previously reported the development of a multienzyme *in vitro* DNA replication system which replicates double-stranded DNA templates efficiently (1-5). As summarized in Table 1, the system now consists of seven highly-purified bacteriophage T4-encoded proteins, the products of T4 genes 32, 41, 43, 44, 45, 61, and 62. By mutational analyses, each of these proteins has been shown to play a major role in DNA replication *in vivo* (6-8), and each is essential for efficient replication fork movement *in vitro*

[1] This work was supported by Grant No. GM 24020 from the National Institutes of Health.

[2] Present address: Institut Pasteur, Parasitologie Experimentale, 28 Rue du Dr Roux, Paris 15, France

ISBN 0-12-048850-7

(4,5,9). In Figure 1, we present a genetic map of T4 bacteriophage, emphasizing the genes corresponding to these seven replication proteins. Comparison to Table I confirms the general tendency for functionally interacting T4 gene products to map in adjacent genes (10).

It has been our goal to dissect and reconstruct the T4-encoded replication apparatus in order to understand the protein-protein and protein-DNA interactions generally involved in DNA replication. Therefore, after establishment of the basic *in vitro* replication system (1,2), we have attempted to decipher the details of replication fork movement by breaking down the process into simpler components. Progress in this regard was greatly aided by the discovery of the gene 61 protein as the seventh T4 protein involved in our system (4).

The gene 61 protein is one of the two RNA priming proteins (4,16), and it is required for DNA synthesis in such small amounts that it was previously overlooked as a trace contaminant in our 32 protein preparations. This protein has now been highly purified and identified as the T4 gene 61 product by several criteria, including two-dimensional gel electrophoresis (14). In wild type T4 infected *E. coli* lysates, the 61 protein may be identified as an extremely basic protein of 40,000 daltons (Figure 2A). The spot corresponding to 61 protein is absent when a 61 amber mutant is used to infect *E. coli* (Figure 2B) and reappears when this amber mutation is suppressed by growth in an appropriate host (Figure 2C). Mutants in T4 gene 61 show a delayed start in T4 DNA synthesis *in vivo*, although they are not completely defective (18). This fact suggests that some *E. coli* host protein can partially compensate *in vivo* for the absence of this protein.

Replication forks can be started *in vitro* by covalent addition onto a random nick in a double-helical DNA template molecule (4,5,9). In this way, the fork propagation aspect of replication can be analyzed in a relatively simple well-defined system, divorced from the biochemical complexities inherent to the mechanism used for fork initiation within the cell (whereby replication bubbles are generated at specific sites (19)).

After a brief overview, our current understanding of the process of replication fork movement will be outlined by discussing some model reactions organized with respect to what these reactions have taught us about three separate aspects of the DNA replication process: RNA priming, helix destabilization, and base-pairing fidelity. Summarized elsewhere (20,21) are the results of our study of the T4 DNA

Table I

Properties of the seven T4 bacteriophage-coded proteins required for efficient DNA replication fork movement

Functional class	DNA polymerase	Helix-destabilizing protein	Polymerase-accessory proteins		RNA-priming proteins	
Genetic locus	gene 43	gene 32	genes 44 & 62;	gene 45	gene 41	gene 61
Activities	5'→3' polymerase 3'→5' exonuclease	Cooperative binding to SS DNA	SS-DNA termini-dependent rATPase, dATPase		Long SS-DNA-dependent rGTP, dGTPase rATP, dATPase	Binds DNA
Current Purity (%)	99	99	99	95	90	70
Polypeptide Molecular Weight (kD)	110	35	34 & 20;	27	58	40

Table I. The proteins listed are those used in the studies presented in this report. They were purified as previously described (11,12,13), except that the gene 61 protein was prepared by a modification (14) of earlier procedures (4).

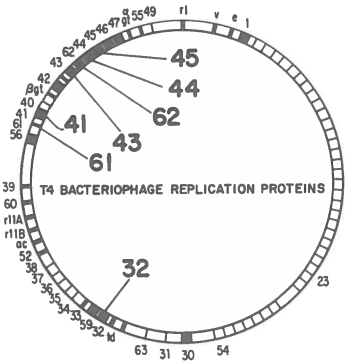

FIGURE 1. Genetic map of T4 bacteriophage with replication protein genes emphasized (after ref. 15).

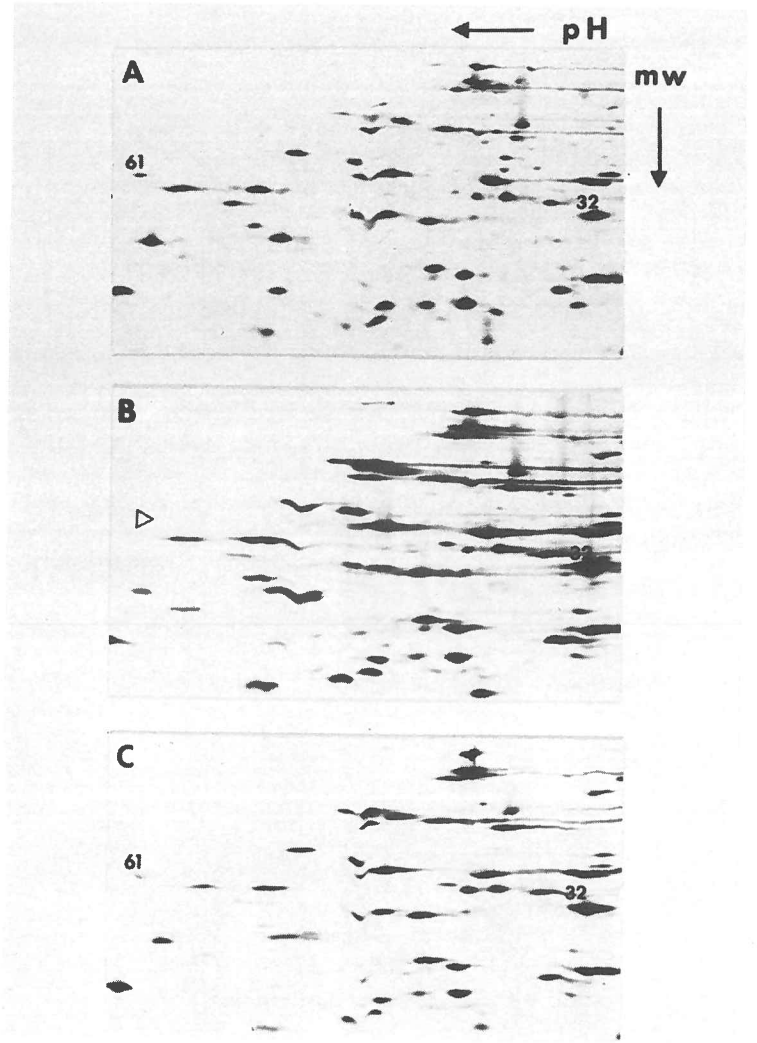

FIGURE 2. Two-dimensional gel electrophoresis of T4 bacteriophage early proteins identifies the gene 61 protein. (A) T4 wild type infected *E. coli* B (su); (B) T4 am E219 (gene 61) infected *E. coli* B (su); (C) T4 am E219 (gene 61) infected *E. coli* B40 (suI). *E. coli* cells infected with T4 at a multiplicity of 10 were labeled by the addition of [^{14}C] mixed amino acids from 3 to 7.5 min post infection at 37°C. NEPHGE gels were run according to O'Farrell et al. (17), with the protein spots detected by autoradiography.

topoisomerase, an enzyme thought to be a central component in the important (but separable) biological process of replication fork initiation.

DNA SYNTHESIS IN THE *IN VITRO* SYSTEM

When mixed together, the seven purified T4 replication proteins listed in Table I can catalyze extensive *in vitro* DNA synthesis. As previously demonstrated, any natural double-stranded DNA template can be replicated efficiently, with at least several copies of the input DNA template synthesized during a 30 minute incubation (5). DNA synthesis starts by covalent addition onto the 3'OH end at a nick in a double-helical DNA template molecule, and it continues by a strand displacement mode to generate a replication fork with the expected leading and lagging strand geometries (see Figure 11, below).

In our present system, DNA appears to be synthesized continuously on the leading side of the fork, while being made discontinuously on the lagging side of the fork where Okazaki pieces are generated (5). In addition to having the structure expected from *in vivo* studies as judged by electron microscopy (1,2,5), the replication forks produced *in vitro* move in a highly processive fashion and travel at approximately the *in vivo* rate of 500 nucleotides per second (4,5). Moreover, as will be discussed below, the template DNA sequences are copied with an impressive base-pairing fidelity (22).

Circular DNA templates are used with great efficiency, due to the fact that, once established, replication forks can continue indefinitely around the template in a "rolling circle" mode of synthesis (1,5,9). Moreover, it is not important for such "rolling circles" that the Okazaki pieces generated on the lagging strand of the replication fork be sealed to each other, since the newly-synthesized DNA on the rolling circle tails need not be reutilized as a template. In this connection, we note that the additional proteins essential for the joining of adjacent Okazaki pieces (including T4 DNA ligase (23) and an enzyme which degrades RNA primers) are not included in the mixture of the seven T4 replication proteins listed in Table I; correspondingly, the Okazaki pieces made on the lagging strand in this *in vitro* system are not sealed (5).

RNA PRIMING

Using either bacteriophage fd DNA or denatured T4 DNA as a single-stranded DNA template for modeling lagging strand

synthesis, both we and the Nossal laboratory have been able to show that only DNA, the 41 protein, and the 61 protein are required for the DNA-templated synthesis of short RNA oligonucleotides from ribonucleoside triphosphates (4,24,25,26). When T4 DNA polymerase and the four dNTPs are added, these "RNA primers" can be used to initiate *de novo* DNA chain starts. In the seven protein *in vitro* replication system, it is possible to demonstrate that more than 90% of these newly-synthesized RNA oligonucleotides become covalently attached onto the 5' end of newly-initiated DNA chains, from which they can be recovered by deoxyribonuclease digestions (25).

We have carried out an extensive analysis of the sequence of these RNA oligonucleotides (25,26; see also ref. 24). The results obtained for the *in vitro* products agree completely with the sequence of the RNA primers isolated from T4-infected cells, as reported by the Okazaki laboratory (27). Strikingly, all of the primers made *in vitro* in the presence of 4 rNTPs are five nucleotides long, even though they are very heterogeneous in sequence. Sequence analysis of 5' terminally-labeled primers, using nucleotide-specific enzymatic cleavages followed by electrophoretic sizing (28), reveals that the major products have the sequence pppApCpNpNpN, where (at least in positions 3 and 4) N can be any one of the four ribonucleotides.

If GTP and UTP are omitted, we obtain a substantial amount of pentaribonucleotide product of sequence pppApCpNpNpN, where N can now be either A or C. Correspondingly, we observe *de novo* DNA chain starts in the seven protein system with only ATP and CTP present (4). However, now tetra-, tri-, and dinucleotide products also accumulate. It therefore appears that the synthesis of all of the RNA primers is template directed, and that prematurely-aborted products smaller than pentanucleotides are made only if an A or C in the DNA sequence calls for the insertion of a UTP or a GTP when these nucleoside triphosphates are missing.

The nature of the RNA primers made *in vitro* can be most readily seen by a two-dimensional fractionation procedure (29), consisting of electrophoresis on cellulose acetate at pH 3.5 (which separates oligonucleotides primarily according to their nucleotide composition), followed by homochromatography on a DEAE cellulose thin-layer plate (which separates oligonucleotides primarily according to their chain length). The result of one such analysis is shown in Figure 3A. Here the products are those which become labeled with [γ^{32}P]ATP when the primer synthesis reaction is carried out in the presence of ATP and CTP only. The distinct spots detected by autoradiography have a mobility in the homochromatography dimension corresponding to 5'-triphosphate terminated tri-,

tetra- and pentanucleotides, as indicated by the numbered brackets. Note that in the electrophoresis dimension, all of these oligonucleotides move relatively slowly, as expected for products that contain only A and C (30).

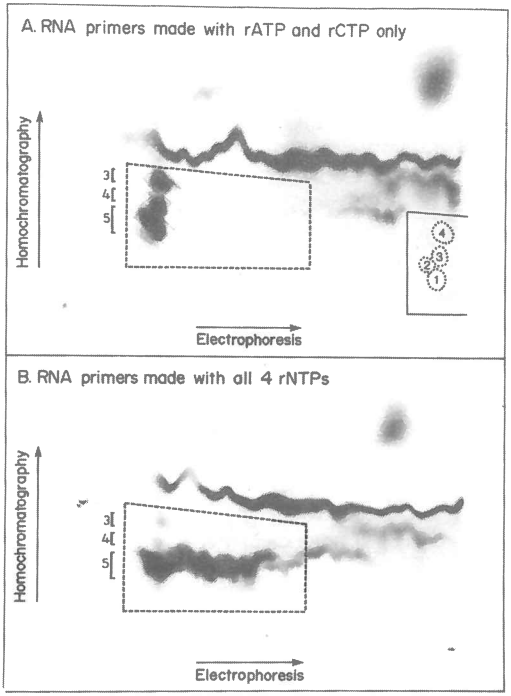

FIGURE 3. The T4 gene 41 and 61 proteins synthesize a mixture of different short oligoribonucleotides of defined length. The reaction mixture described previously (4) contained [γ^{32}P]rATP to label the 5' end of the RNA products beginning with pppA. (A) Only rATP and rCTP present. (B) All four ribonucleoside triphosphates present. Incubation was for 15 min at 30°C, after which time the reaction mixture was spotted onto DE-81 filter paper and washed (4). The material remaining on the paper was eluted with triethylamine carbonate, desalted, and analyzed in the two-dimensional fractionation system described by Jay et al. (29). All of the significant reaction products migrate within the area demarked by the dotted box; the numbered brackets denote the chain length of the oligonucleotide spots, as determined by partial hydrolysis followed by polyacrylamide gel electrophoresis (25).

Figure 3B presents an identical analysis of the [γ^{32}P]-ATP-labeled RNA oligonucleotides synthesized in the presence of all four ribonucleoside triphosphates. Under these conditions, almost all of the RNA products move as pentanucleotides in the homochromatography dimension. At the same time, there is a great heterogeneity in the electrophoretic mobility (and thus the nucleotide composition) of these short RNAs. Those pentanucleotides which move the most rapidly in the electrophoresis dimension are expected to contain the most U and G. Very similar patterns of oligonucleotides are seen whether single-stranded T4 or fd DNA serves as the template.

We conclude that the sequence of the RNA made is heterogeneous, that the synthesis is template-directed, and that the synthetic apparatus automatically terminates RNA chain growth at a length of five. This absolute self-termination of the reaction implies that the RNA priming proteins automatically release the 3'OH end of the completed primer, thus avoiding a competition of primase and polymerase for chain extension at the same 3'OH end (Figure 4).

There are 40 to 50 distinguishable sites on the fd DNA molecule at which primers could be made with the sequence pppApCpNpNpN, where N is either C or A; yet only five major and five minor DNA chain starts are seen with ATP and CTP present (4,25). There must therefore be some restriction on

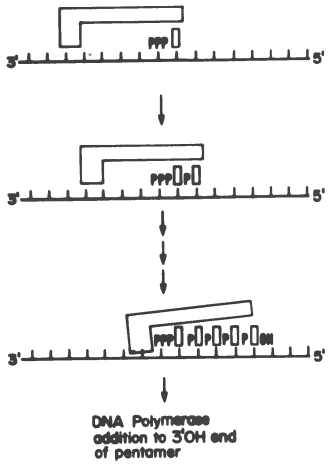

FIGURE 4. Schematic view of a primer synthesizing system which stops automatically at a chain length of 5, and then exposes the 3'OH end of the primer to the DNA polymerase.

the selection of *de novo* chain start sites besides the mere ability to template for the sequence ppppApC. However, this restriction cannot be severe, since any single-stranded DNA tested which can template for A and C will work (26), including the synthetic random copolymer poly d(I,T).

An interesting sidelight is that, whereas only pppA starts are seen on a T4 DNA template, on all other natural DNAs tested many of the RNA primers made start with pppG (26). This limitation to pppA starts appears to be due to the substitution of glycosylated 5-hydroxymethyl cytosine for cytosine in T4 DNA; when a cytosine-containing T4 DNA is used as a template, the primers made can start with either pppA or pppG (26).

As noted in Table I (see above), there are two sets of proteins in the T4 replication system which hydrolyze nucleoside triphosphates to nucleoside diphosphates and inorganic phosphate. As we shall see below, the contributions of these DNA-dependent hydrolyses to the replication process can be separately assessed and shown to be essential by making use of different synthetic nucleotide inhibitors (ATPγS being used to inhibit the 44/62 protein ATPase selectively, and GTPγS being used to inhibit the 41 protein GTPase selectively). One can thereby ask how the splitting of GTP to GDP by the gene 41 protein is related to the function of this protein in RNA priming. Two observations are of particular relevance here: first of all, very long DNA chains and DNA circles are the best cofactors for the stimulation of the 41 GTPase, which suggests that this protein runs rapidly along its bound DNA strand, driven by conformational changes induced by GTP hydrolysis (26,31). Secondly, while a limited RNA primer synthesis will occur *in vitro* without GTP hydrolysis by the 41 protein, the prolonged synthesis of RNA primers requires such hydrolysis. In combination, these two results suggest that the 41 protein acts as a "mobile promotor" (32,33), which marks the spots where successive RNA primers are to be synthesized on the lagging strand of a replication fork. Why this arrangement might be useful to the replication process is best discussed after a description of the separate role which the 41 protein appears to play in DNA helix destabilization at the fork.

HELIX DESTABILIZATION AT THE FORK

We have previously observed that the T4 DNA polymerase alone can make DNA at a rate of about 250 nucleotides per sec on a single-stranded DNA template (2,34), whereas in most reactions on a double-stranded DNA template the polymerase moves much more slowly (or not at all; ref. 4). These

results suggest that the rate of opening of the DNA double-helix ahead of a replication fork normally limits the rate at which the leading strand DNA polymerase molecule can synthesize DNA. As reviewed elsewhere (35), it appears that a variety of different helix-destabilizing forces must act in concert to open up the parental DNA helix at a replication fork at the rate of DNA chain polymerization *in vivo* (about 500 nucleotides per sec). We have studied the source of helix-destabilization in the T4 system in detail, using two different model systems.

Penetration of the DNA polymerase through the hairpin helices in a single-stranded DNA molecule. The individual contribution of each replication protein to the helix opening at a fork can be detected sensitively by observing the rate of DNA polymerase movement through the relatively weak, imperfect helices which form on a single-stranded DNA template, where even modest helix-destabilizing forces seem to have an observable effect (36). Results obtained by Chao Huang and John Hearst at the University of California at Berkeley reveal that the polymerase accessory proteins (44/62 and 45 proteins) accelerate the rate of helix penetration by the polymerase by a factor of about 5. A very similar stimulation of T4 DNA polymerase is detected when only the T4 helix-destabilizing protein (the gene 32 protein) is added. Interestingly, the 44/62 and 45 proteins interact synergistically with the 32 protein; consequently, the rate of polymerase travel over the strongest hairpin helices in the template is increased more than 50-fold in the presence of all four additional proteins (37).

By using the synthetic ATP analogue, ATPγS, Huang and Hearst have shown that, while ATP hydrolysis is required for formation of an apparent accessory protein-DNA polymerase complex, it is not required for the subsequent efficient traversal of hairpin helices by the complex. Thus, in this model system, helix destabilization seems to come from two sources: the expected shift in the helix-coil equilibrium caused by the cooperative binding of the gene 32 protein to a template strand (38), plus the presence of an accessory protein "clamp" which allows the DNA polymerase to continue its DNA synthesis in a highly processive fashion, even when pushing against a helical region in the template (37). The data of Newport and von Hippel also supports this view of the accessory protein complex (39).

Penetration of the DNA polymerase through a perfect double-helix. In contrast to the above results using a single-stranded DNA template, all five T4 proteins (the 43,

44/62, 45 and 32 proteins) must be present in order to obtain any extensive DNA synthesis on double-stranded DNA templates (4,9). We therefore conclude that both contributions to helix destabilization are required before the polymerase can make any substantial headway through the much more stable structure represented by a long perfect DNA helix.

In order to examine further the mechanism of helix destabilization at an actual DNA replication fork, we have developed a second model system, in which we can measure the rate at which a leading strand DNA polymerase molecule moves through a perfect double-helix (4). The principle of this method is schematically outlined in Figure 5. A very brief polymerization reaction is carried out in which leading-strand DNA synthesis is primed on a pre-existing nick on a double-helical template. High specific activity dTTP is used to radioactively label the DNA product, and 5-hydroxymethyl dCTP (dHMCTP) replaces dCTP as the fourth DNA precursor nucleotide (dHMCTP, the natural DNA precursor for the T4 system, is utilized indistinguishably from dCTP in the T4

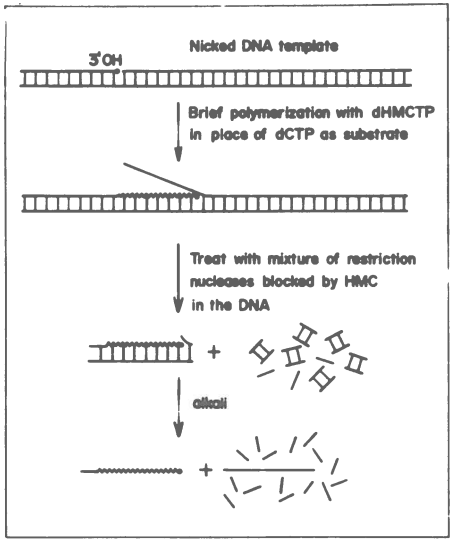

FIGURE 5. Schematic outline of a method for measuring DNA polymerase rates on a double-helical DNA template. The newly-synthesized, radioactive DNA product chain is indicated by the zigzag line. The length of this product chain is measured as a function of polymerization time by exploiting the fact that some restriction endonucleases will not cut DNA unless it has C residues on both DNA strands (4).

in vitro reactions). The substitution of HMC for C on one of the two strands of a double-helix renders the DNA product completely resistant to cutting by certain DNA restriction endonucleases. Therefore, after a very brief DNA synthesis reaction has been stopped and the DNA product deproteinized, exhaustive treatment of the total DNA with a mixture of these restriction nucleases removes all but the newly-synthesized length of DNA helix. The number of bases polymerized per second at the chain growing point is then determined from a plot of the length of this nuclease-resistant DNA product as a function of DNA synthesis time.

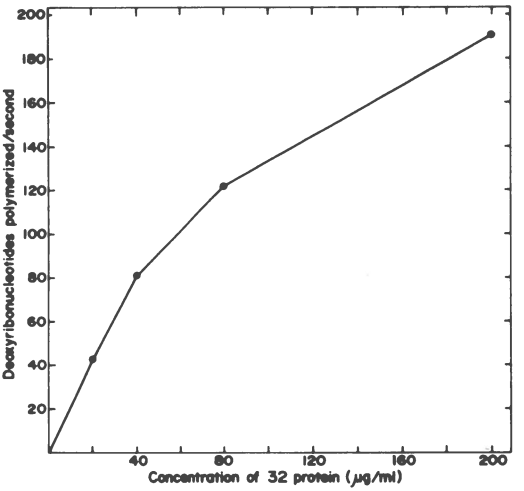

FIGURE 6. The rate at which the T4 DNA polymerase-accessory protein complex moves through a double-helical DNA template depends on the concentration of free 32 protein. A standard reaction mixture was used, consisting of 10 μg/ml cytosine-containing T4 DNA, 160 μM each of dGTP, dTTP, dATP, and dHMCTP (in place of dCTP), 0.5 mM DTT, 60 mM potassium acetate, 33 mM Tris acetate, pH 7.8, 10 mM magnesium acetate, 250 μM ATP, 250 μM GTP, 200 μg/ml nuclease-free bovine serum albumin, 28 μg/ml 44/62 protein, 34 μg/ml 45 protein, 3.4 μg/ml 43 protein (DNA polymerase), 1 mCi/ml [α^{32}P]dTTP and 1 μg/ml 32 protein. After a 2 min preincubation at 37°C, additional 32 protein was added to the reaction mixture to raise the 32 protein concentration to the indicated level and start the reaction. At various times within the next few minutes, the reaction was stopped and processed according to Figure 5, as previously described (4). An analysis similar to that illustrated in Figures 7 and 8 below was used to determine the fork rate at each 32 protein concentration.

When we used the method outlined in Figure 5 for measuring polymerization rates in a five-protein reaction on a double-helical template (43, 44/62, 45 and 32 proteins only), we found that this rate was less than 20% of the estimated *in vivo* rate of T4 replication fork movement (<100 nucleotides/second; ref. 4). A striking effect of the free helix-destabilizing protein concentration has been observed more recently in these experiments. A compilation of the results obtained is presented in Figure 6, where it can be seen that fork rates increase dramatically as the 32 protein concentration is raised (early time points were analyzed, so that the 32 protein added was always present in vast excess over that which became bound to DNA single-strands). Such a result suggests that the 32 protein exerts a "pressure" at the fork which is roughly proportional to the concentration of free 32 protein present. The data thereby argues that this helix-destabilizing protein plays an active role in helix unwinding ahead of the leading strand DNA polymerase molecule, in contrast to a "passive" role in which the 32 protein would merely be needed to bind up the single-stranded template DNA displaced by this polymerase.

A special role of the gene 41 protein in helix destabilization. Considerably greater *in vitro* fork rates than those seen in Figure 6 had been estimated earlier for a reaction with all seven of the T4 replication proteins listed in Table I present, based on the average amount of DNA synthesized per rolling circle intermediate on a circular DNA template (1,5). This discrepancy was explained when we found that the addition of one of the RNA priming proteins to the five protein mixture dramatically increases the rate of fork movement measured by the Figure 5 assay (4). In Figure 7, we present an autoradiogram illustrating the progressive increase in the length of the nuclease-resistant DNA product in reactions with and without the gene 41 protein present. It is clear that a faster rate of DNA synthesis is induced in a fraction of replication forks in the presence of the gene 41 protein.

A quantitative analysis of this data, obtained by using a densitometer to measure DNA concentrations, is shown in Figure 8. In the reaction without 41 protein present (lanes A, B and C), there is a rather sharp "front" representing the largest molecules present at each time point. As expected, the labeled product DNA size, measured as the midpoint of the front, increases continuously during the incubation period (see also ref. 4). The presence of 41 protein induces a clear second component composed of much larger DNA molecules (lanes D, E and F). The position of this faster "front"

likewise increases continuously with time, the quantitation in this case indicating that this class of replication forks is moving at least 3.2-fold more rapidly than the forks in the control reaction without 41 protein. More recent experiments have confirmed these findings, and shown that, with addition of fresh preparations of the gene 41 protein, nearly all of the forks are induced to move at the more rapid rate. These recent experiments have also revealed that fork rates are not increased by the further addition of the 61 protein and the remaining rNTPs to the reactions containing 41 protein, indicating that the rate stimulation seen in Figure 7 and 8 is completely independent of RNA primer synthesis (40).

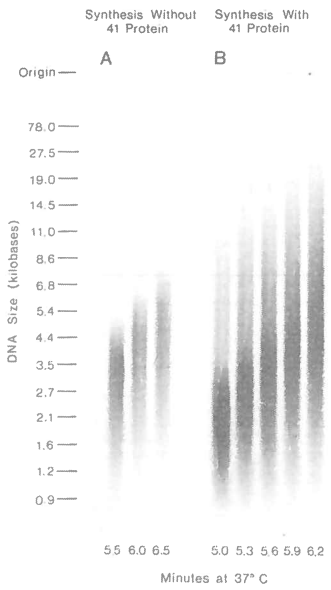

FIGURE 7. Autoradiogram showing an electrophoretic analysis of the restriction enzyme-resistant DNA products obtained according to Figure 5. Reactions were carried out as described for Figure 6, except that 32 protein was added to 21 μg/ml to start the reaction after 4.0 min at 37°C, and 41 protein (30 μg/ml) was included in the preincubation mixture in (B) only. After the indicated total times of incubation, the reaction products were processed according to Figure 5 (ref. 4). The labeled DNA products were then analyzed by autoradiography after being sized by electrophoresis through a 0.4% agarose gel (in 30 mM NaOH, 2 mM Na_3EDTA).

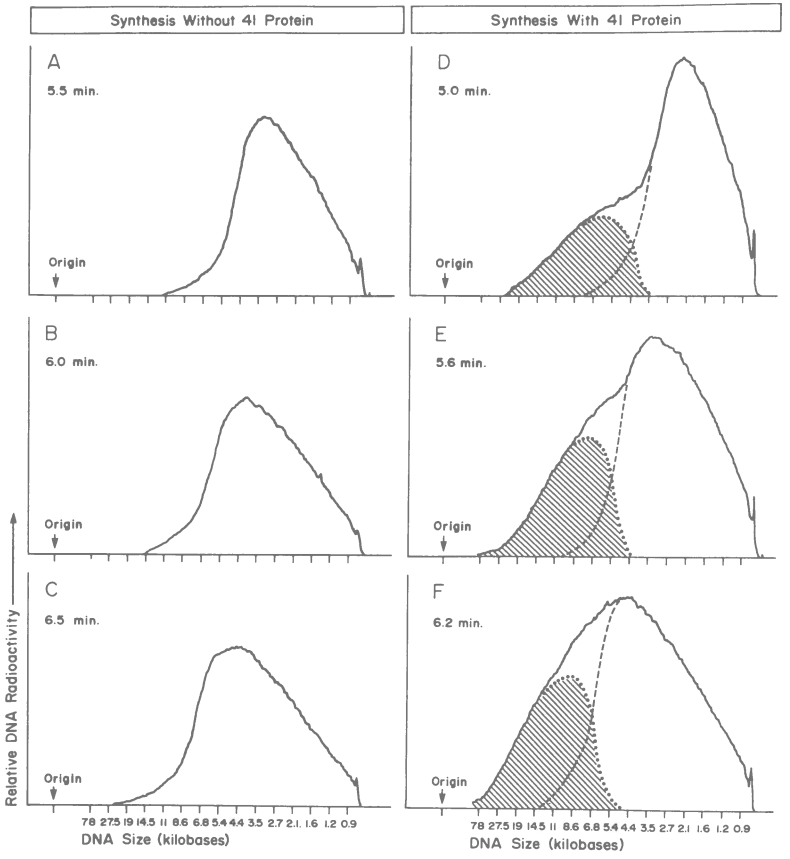

FIGURE 8. The T4 gene 41 protein speeds up the rate of DNA polymerase movement through a double-helical template. The autoradiogram in Figure 7 was scanned with a Joyce Loebl densitometer to obtain the graphs shown here. The size distribution obtained with 41 protein present was resolved as the indicated sum of two different size distributions, by reference to the parallel reaction without 41 protein.

In view of our previously mentioned evidence that the 41 protein GTPase runs rapidly along a DNA single-strand (26), we suspected that the 41 protein was utilizing some of its nucleotide hydrolysis energy to help push the fork ahead through double-helical templates, as schematically illustrated in Figure 9. As predicted by this scheme, addition of GTPγS to the six protein system [a synthetic nucleotide which specifically blocks nucleoside triphosphate hydrolysis by

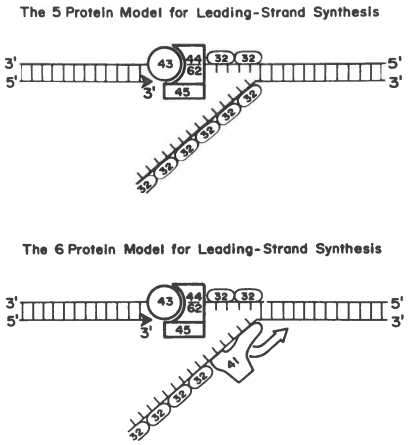

FIGURE 9. Schematic view of the six protein model for leading strand synthesis. The synthesis is believed to differ from that obtained with five proteins present only by the action of the gene 41 protein at the indicated position on the lagging strand (see text). The polymerase (gene 43 protein) and the polymerase accessory proteins (gene 44, 45 and 62 proteins) are thought to form a complex as indicated. The gene 32 protein is thought to help destabilize the helix by binding tightly and cooperatively to DNA single-strands.

the 41 protein (26)] immediately decreases the rate of movement of ongoing replication forks back to the slower five protein level (data not shown).

In conclusion, we propose that the 41 protein runs unidirectionally and rapidly along the single-stranded DNA template chain on the lagging side of a fork, where it performs both a RNA priming and a helix opening role. The 41 protein would thereby coordinate DNA synthesis on the leading and lagging sides of a fork, by functioning both as proposed for the *E. coli dna*B protein in RNA priming (32) and as proposed for a DNA-helicase in fork movement (31). [An analogy between the 41 protein and the T7 bacteriophage gene 4 protein (41) also seems likely].

The assay shown in Figure 5 has also proven to be of use for a variety of other purposes. For example, we find that the total DNA synthesis drops as the concentration of T4 DNA polymerase is lowered in the five protein reaction, while the fork rate remains unchanged. This finding suggests that the

polymerase accessory proteins allow the DNA polymerase to move in a completely processive fashion in this system, so that each enzyme molecule proceeds for 20,000 nucleotides or more on the same DNA chain without falling off (40).

The ability to measure changes in fork rates has also enabled us to use ATPγS addition to show that the continued movement of T4 DNA polymerase through a perfect double helix requires repeated hydrolysis of ATP by the polymerase accessory proteins, a result opposite to that observed when the polymerase is moving through the weak hairpin helices on an fd DNA template (see discussion above). While this difference is puzzling, it is possible that the accessory protein "clamp" is especially labilized when the polymerase is forcing its way through helical regions of the template, requiring frequent ATP hydrolysis by the 44/62 protein to prevent its complete disassembly. [Note that the observation that the polymerase nevertheless moves in a highly processive manner through these double-helical templates indicates that the clamp must require ATP hydrolysis for resetting before its actual dissociation]. In this view, the addition of ATPγS blocks the resetting process and thereby stops a polymerase molecule moving through a perfect double helix within a fraction of a minute; in contrast, since the helical "back pressure" is much less on a single-stranded DNA template, very little resetting of the polymerase accessory protein clamp is needed over the time intervals examined (5 to 20 min at 30°C).

We conclude that parental helix destabilization at a replication fork in the T4 replication system comes from the sum of 1) active DNA binding by the helix-destabilizing protein, 2) a push from the DNA polymerase, which moves ahead by polymerization while tied down by its associated accessory proteins, and 3) a special "DNA walking protein" which runs rapidly along the lagging strand DNA template and helps to push the adjacent helix open (Figure 9, above).

REPLICATION FIDELITY

A highly-sensitive assay for measuring the fidelity of *in vitro* DNA replication can be designed, based on measuring the reversion frequency of nonsense mutants of small bacteriophages (4,42). Our assay is schematically indicated in Figure 10; the T4 replication complex is first used to replicate a circular øx174 DNA template by a rolling-circle mode. After the long linear DNA tails are cut with a restriction endonuclease which makes one cut per genome, the linear DNA is circularized and used to transfect spheroplasts. Since nonsense mutant DNA is normally used as a template,

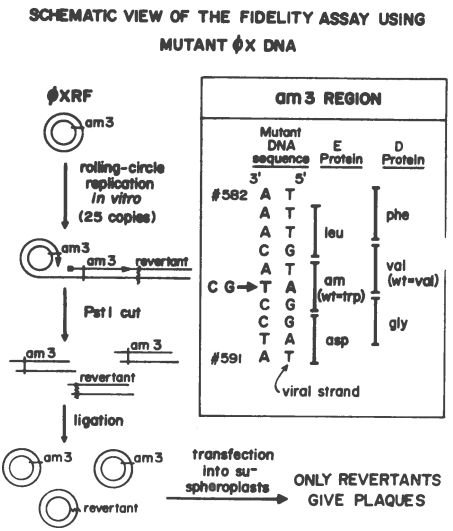

FIGURE 10. Schematic summary of the infectivity assay used to measure replication fidelity. The circular DNA from an amber mutant of bacteriophage øx174 is replicated *in vitro* and then processed as described in the text. For details, see refs. 4 and 22.

only revertants give plaques upon transfection of nonpermissive spheroplasts. The øx174 DNA template was chosen for these studies, since øx174 transfects with high efficiency, its complete DNA sequence is available (43), and nine of its nonsense mutants have been sequenced (see for example, the sequence for the am3 mutation in Figure 10).

Using this assay, we have been able to show that the DNA synthesized in our *in vitro* system is fully infectious, even after 50 template copies have been made (4,22). Reversion frequencies were measured for two different sequenced øx174 amber mutations. Biased deoxyribonucleoside triphosphate concentrations, which favor the errors leading to wild type revertants, were used in order to increase the sensitivity of the assay. Such a nucleotide bias also allows the direction of the misreading to be ascertained, and separate frequencies to be measured for both possible base mispairings leading back to wild type.

Table II presents a summary of the data obtained for these two mutant sites, with results in magnesium and manganese containing buffers compared. As for the *in vivo*

Table II

Site-specificity of mutation rates

Type of in vitro reaction		Fidelity (fraction of errors)	
dNTP bias	Divalent cation	Mutant am3 3'AAACATCCTA5' 5'TTTGTAGGAT3'	Mutant am18 3'CTGCATCTTC5' 5'GACGTAGAAG3'
C > T	Mg^{++}	$3.3 \pm 0.3 \times 10^{-7}$	$2.0 \pm 0.5 \times 10^{-6}$
G > A	Mg^{++}	$2.2 \pm 0.3 \times 10^{-6}$	$9.1 \pm 2.9 \times 10^{-6}$
C > T	Mn^{++}	$4.6 \pm 1.1 \times 10^{-6}$	$5.8 \pm 2.3 \times 10^{-6}$
G > A	Mn^{++}	$5.0 \pm 0.9 \times 10^{-5}$	$1.3 \pm 0.2 \times 10^{-5}$

Table II. DNA synthesis was performed under standard conditions, except that $MnCl_2$ was substituted for $MgCl_2$ where indicated. The fidelity assay used is that outlined in Figure 10. The use of strongly biased concentrations of deoxyribonucleoside triphosphates in the *in vitro* reaction was necessary for detection of these low error frequencies. The sequences of both amber mutants studied are shown. In both cases, the reversion to wild type requires a TA→CG transition at the base pair surrounded by a box. For details, see ref. 22.

process, manganese is seen to increase the frequency of some mispairs by as much as 20-fold. Most importantly, the normal fidelity of DNA synthesis in this *in vitro* system is impressive, being on the order of 10^{-6} errors per base pair replicated (22). Results analogous to these have been independently obtained in the laboratory of N. Sinha (44).

Although comparison to the *in vivo* error frequency is complicated by the fact that this frequency is known to vary widely depending on the exact site mutating (45), we believe that our *in vitro* fidelity is no more than about 100-fold worse than the replication fidelity seen *in vivo*. Any additional fidelity seen *in vivo* could come from a methyl-directed post-replication DNA repair system, similar to that recently discovered in the *E. coli* host (46). It is thus possible that the seven protein T4 *in vitro* system has the full fidelity of the comparable process which operates within the cell.

A base-pairing fidelity of one error in 10^6 is considerably greater than the fidelities reported for DNA polymerases alone on natural DNA templates (47). In order to explore possible sources for this increased accuracy, we have examined the effect of the individual replication proteins listed in Table I on the 3'→5' "proofreading exonuclease" (48) activity of the T4 DNA polymerase; an increase in this exonuclease activity in some "antimutator" T4 DNA polymerases has been demonstrated to increase fidelity both *in vivo* and *in vitro* (49). By using agarose gel electrophoresis to measure the rate of degradation of a linear double-stranded DNA fragment in the absence of dNTPs, we can show that the addition of the polymerase accessory proteins to the T4 DNA polymerase increases its 3'→5' degradation rate 5-fold, to about 100 nucleotides hydrolyzed per minute. This stimulation requires the presence of 44/62 protein, 45 protein, and ATP in a hydrolyzable form. Moreover, whereas addition of 100 μg/ml of 32 protein to the DNA polymerase alone inhibits its exonuclease (see also ref. 50), the addition of this concentration of 32 protein to the polymerase plus its accessory proteins increases the 3'→5' degradation rate by another 2-fold (i.e., to 10-fold the rate of the polymerase alone). It is therefore clear that the ratio of exonuclease to polymerase activities in the replication complex will depend on a variety of factors other than the DNA polymerase, and that this represents one of several ways in which the other replication proteins can be expected to increase DNA replication fidelity.

DISCUSSION

In view of the data discussed in this and other reports, we believe that the present seven protein T4 *in vitro* DNA replication system accurately mimics many aspects of the corresponding *in vivo* process of DNA replication fork movement. A cartoon schematically illustrating our current understanding of this process is presented in Figure 11. In summary:

1. The T4 DNA polymerase and its accessory protein complex (the gene 44/62 and 45 proteins) are viewed as a single unit, perhaps corresponding to the even larger "holoenzyme" complex which represents the functional DNA polymerase unit in the *E. coli* DNA replication apparatus (33). As illustrated, only a single DNA polymerase unit is likely to be operative on each DNA template strand at any one instant; whether or not the two units on each fork are

actually spatially coupled to create a fork with a convoluted tertiary structure (5) remains to be determined.

2. The gene 32 protein facilitates DNA synthesis by coating the single-stranded DNA template on the lagging strand, and also plays an important role in template helix-destabilization ahead of the leading strand DNA polymerase molecule (Figure 6). Additional evidence for the latter assertion has been obtained from studies in which the intact 32 protein has been substituted in the five protein reaction with its proteolytic cleavage product, the 32*I protein (51).

3. The 41 protein serves to couple the leading and lagging strand syntheses via its dual roles on the lagging strand (RNA priming and DNA helicase). Perhaps the most reasonable interpretation of the requirement for two different T4 proteins to make an RNA primer is that the 41 protein is primarily a helicase, which is also designed to mark the site where RNA primers are to be made. While we therefore suspect that the gene 61 protein is the actual RNA primase

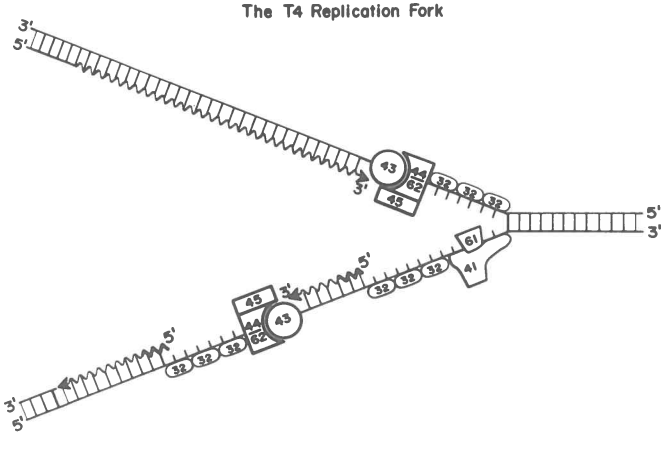

FIGURE 11. A schematic representation of a T4 replication fork, illustrating the proposed placement of the seven replication proteins listed in Table I. Discontinuous 5'→3' synthesis on the lagging strand is coupled to leading strand DNA synthesis by the action of the 41 protein "mobile promoter" complex, which aides in helix unwinding by moving along the lagging strand DNA template. The gene 32 helix-destabilizing protein also facilitates helix unwinding at the fork.

involved, our attempts to test this hypothesis have not been successful. Even though the assays employed would have detected 5% relative primer-synthesizing activity catalyzed by either the 41 or 61 protein alone, under none of the many conditions explored was any RNA product seen unless both proteins were simultaneously present (14).

4. RNA primers five nucleotides long are synthesized and somehow held in place by one or more of the RNA-priming proteins, until a DNA polymerase unit can pick up the primer and elongate it. As a result, the vast majority of the primers made will act to prime new DNA chain starts, despite their very short length.

There are two other multienzyme DNA replication systems for which detailed mechanisms have been elucidated: the *E. coli* and bacteriophage T7 systems (33,35). Table III lists some of the proteins thought to be involved in the movement of replication forks in all three well-characterized systems, grouped according to the functional classification scheme used in Table I. Despite possible gaps in our knowledge, it is clear that the complexity of the replication apparatus increases as one moves from T7 to T4 to *E. coli*. Thus, as organisms acquire larger genomes (and thus a need for increased replication fidelity (52)), they would appear to need more proteins to replicate their DNA.

Table III

An attempted comparison of three DNA replication systems of different complexities

	System		
Protein class	T7	T4	*E. coli*
Helix-destabilizing protein	gene 2.5	gene 32	*E. coli* ssb
DNA polymerase	gene 5	gene 43	pol III
Polymerase accessory proteins	none	genes 44, 62 & 45	six extra proteins in "holoenzyme"?
RNA priming proteins	gene 4	genes 41 & 61	*dna*B, *dna*G, plus others

Differences in Table III are especially marked in the class of proteins which we have called "polymerase accessory proteins." In this category, bacteriophage T7 has none, bacteriophage T4 has three, and *E. coli* has approximately six polypeptide chains (33). It is intriguing to speculate that the increased number of replication proteins associated with the polymerase is responsible for the decrease in base-pairing errors (estimated from mutation rates) as one moves from T7 to T4 to *E. coli*. It is also likely that the probability of a "geometrical error" (such as strand-switching by the DNA polymerase) decreases in the same order.

In our view, the T4 DNA replication system may be simple enough to allow detailed chemical and mechanistic studies on each of the replication proteins, including the determination of their tertiary structures by x-ray diffraction analyses. At the same time, this system is clearly sufficiently complex that a complete understanding of all of the protein-protein and protein-DNA interactions involved will be useful in the analysis of many other important genetic processes, for which the same molecular principles are certain to be used.

REFERENCES

1. Morris, C.F., Sinha, N.K., and Alberts, B.M. (1975). Proc. Natl. Acad. Sci., USA 72, 4800.
2. Alberts, B., Morris, C.F., Mace, D., Sinha, N., Bittner, M., and Moran, L. (1975). In "DNA Synthesis and Its Regulation", Vol. 3, ICN-UCLA Symposium on Molecular and Cellular Biology. (M. Goulian, P. Hanawalt and C.F. Fox, eds.), pp. 241-269. W.H. Benjamin, Inc., Menlo Park.
3. Alberts, B., Barry, J., Bittner, M., Davies, M., Hama-Inaba, H., Liu, C.-C., Mace, D., Moran, L.,Morris, C.F., Piperno, J., and Sinha, N.K. (1977). In "Nucleic Acid-Protein Recognition" (H.J. Vogel, ed.), pp. 31-63. Academic Press, New York.
4. Liu, C.-C., Burke, R.L., Hibner, U., Barry, J., and Alberts, B.M. (1979). Cold Spring Harbor Symp. Quant. Biol. 43, 469.
5. Sinha, N.K., Morris, C.F., and Alberts, B.M. (1980). J. Biol. Chem. 255, 4290-4303.
6. 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 Leilausis, A. (1964). Cold Spring Harbor Symp. Quant. Biol. 28, 375.
7. Warner, H.R., and Hobbs, M.D. (1967). Virology 33, 376.

8. Curtis, M.J., and Alberts, B. (1976). J. Mol. Biol. 102, 793.
9. Nossal, N.G., and Peterlin, B.M. (1979). J. Biol. Chem. 254, 6032.
10. Stahl, F.W. (1967). J. Cell Physiol. (suppl. 1) 70, 1.
11. Morris, C.F., Hama-Inaba, H., Mace, D., Sinha, N.K., and Alberts, B. (1979) J. Biol. Chem. 254, 6787.
12. Morris, C.F., Moran, L.A., and Alberts, B.M. (1979). J. Biol. Chem. 254, 6797.
13. Bittner, M., Burke, R.L., and Alberts, B.M. (1979). J. Biol. Chem. 254, 9565.
14. Burke, R.L., and Alberts, B.M., in preparation.
15. Wood, W.B., and Revel, H.R. (1976). Bacteriol. Rev. 40, 847.
16. Silver, L.L., and Nossal, N.G. (1979). Cold Spring Harb. Symp. Quant. Biol. 43, 489.
17. O'Farrell, P.Z., Goodman, H., and O'Farrell, P.H. (1977). Cell 12, 1133.
18. Yegian, C.D., Mueller, M., Selzer, G., Russo, V., and Stahl, F.W. (1971). Virology 46, 900.
19. Delius, H., Howe, C., and Kozinksi, A.W. (1971). Proc. Natl. Acad. Sci. USA 68, 3049.
20. Liu, L.F., Liu, C.-C., and Alberts, B.M. (1979). Nature, 281, 456.
21. Liu, L.F., Liu, C.-C., and Alberts, B.M. (1980). Cell, 19, 697.
22. Hibner, U., and Alberts, B.M. (1980). Nature, 285, 300.
23. Weiss, B., and Richardson, C.C. (1967). Proc. Natl. Acad. Sci. USA 57, 1021.
24. Nossal, N.G. (1980). J. Biol. Chem. 255, 2176.
25. Liu, C.-C., and Alberts, B.M. (1980). Proc. Natl. Acad. Sci. USA, submitted.
26. Liu, C.-C., and Alberts, B.M., in preparation.
27. Okazaki, T., Kurosawa, Y., Ogawa, T., Seki, T., Shinozaki, K., Hirose, S., Fujiyama, A.A., Kohara, Y., Machida, Y., Tamanoi, F., and Hozumi, T. (1979). Cold Spring Harbor Symp. Quant. Biol. 43, 203.
28. Donis-Keller, H., Maxam, A.M., and Gilbert, W. (1977). Nucleic Acids Res. 4, 2527.
29. Jay, E., Bambara, R., Padmanabhan, R., and Wu, R. (1974). Nucleic Acids Res. 1, 331.
30. Tu, C.-P.D., Jay, E., Bahl, C.P., and Wu, R. (1976). Anal. Biochem. 74, 73.
31. Kuhn, B., Abdel-Monem, M., and Hoffmann-Berling, H. (1979). Cold Spring Harb. Symp. Quant. Biol. 43, 63.
32. McMacken, R., Ueda, K., and Kornberg, A. (1977). Proc. Natl. Acad. Sci. USA 74, 4190.

33. Kornberg, A. (1980). "DNA Replication." W.H. Freeman and Co., San Francisco.
34. Mace, D., and Alberts, B.M., in preparation.
35. Alberts, B. and Sternglanz, R. (1977). Nature 269, 655.
36. Huang, C.-C., and Hearst, J.E., Anal. Biochem., in press.
37. Huang, C.-C., Hearst, J.E., and Alberts, B.M., manuscript in preparation.
38. Alberts, B.M., and Frey, L. (1970). Nature 227, 1313.
39. Newport, J.W., Kowalczykowski, S.C., Lonberg, N., Paul, L.S., and von Hippel, P.H., this volume.
40. Barry, J., and Alberts, B.M., manuscript in preparation.
41. Kolodner, R., and Richardson, C.C. (1977) Proc. Natl. Acad. Sci. USA 74, 1525.
42. Weymouth, L.A., and Loeb, L.A. (1978). Proc. Natl. Acad. Sci. USA 75, 1924.
43. Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, J.C., Hutchison, C.A., III, Slocombe, P.M., and Smith, M. (1977). Nature 265, 687.
44. Sinha, N.K., and Haimes, M.D., this volume.
45. Coulondre, C., and Miller, J.H. (1977). J. Mol. Biol. 117, 525.
46. Radman, M., Villani, G., Botieux, S., Kinsella, A.R., Glickman, B.W., and Spadari, S. (1979). Cold Spring Harbor Symp. Quant. Biol. 43, 937.
47. Loeb, L.A., Weymouth, L.A., Kunkel, T.A., Gopinathan, K.P., Beckman, R.A., Dube, D.K. (1979). Cold Spring Harb. Symp. Quant. Biol. 43, 921.
48. Brutlag, D., and Kornberg, A. (1972). J. Biol. Chem. 247, 241.
49. Lo, K.-Y., and Bessman, M.J. (1976). J. Biol. Chem. 251, 2480.
50. Huang, W.M., and Lehman, I.R. (1972). J. Biol. Chem. 247, 3139.
51. Burke, R.L., Alberts, B.M., and Hosoda, J., in preparation.
52. Drake, J.W. (1969). Nature 221, 1132.