THE ENZYMATIC REPLICATION OF DNA*

Author: **Arthur Kornberg**
Department of Biochemistry
Stanford University School of Medicine
Stanford, California

INTRODUCTION

The resolution and reconstitution of the pathways of sucrose metabolism in fermentation in yeast initiated modern biochemistry in this century. To this day, an understanding of the mechanics and regulation of fermentation and glycolysis in enzymatic terms continues to be a major focus for research. Our work on the enzymology of DNA replication has been patterned on these lines. It is now at a point where fermentation and glycolysis enzymology stood many years ago.

STRUCTURES OF GENE EXPRESSION

With regard to the enzymology of the three levels of gene expression—replication, transcription, and translation—that for replication has lagged behind. This is largely due to the ease of isolating relatively intact complex structures for transcription and translation, but not for replication.

Yet, the first discovered and purified constituent of gene expression was DNA polymerase I of *Escherichia coli*.[1] The enzyme is a single polypeptide of molecular weight 109,000. It is remarkable for its capacity to copy DNA templates with error-free fidelity. However, this enzyme alone cannot account for the physiological features of replication.

By contrast, the multisubunit RNA polymerase and the complex ribosome, augmented by a few dissociable factors, do give a remarkably good account of the processes of transcription and translation. These structures are readily purified from cell extracts because they are so compact; strong denaturing agents are needed to resolve them.

"THE REPLISOME"

Replication in *E. coli* requires at least a dozen proteins discovered in the past few years in addition to the original DNA polymerase. Very likely, still more will be identified as we search through the rubble of the cell lysate for the scattered remnants of the compact organelle responsible for replication, "the replisome". Unfortunately, the forces which hold this replisome together are dissipated by the gentlest means we have used to lyse the cell.

MAJOR QUESTIONS ABOUT REPLICATION IN 1971.

For some time now, the basic features of DNA chain growth have been clear.[2] This knowledge, based on studies of DNA polymerase I of *E. coli*, holds for all 30 or more DNA polymerases isolated from a variety of bacterial and animal cells. The same

* This review and the lecture on which it is based were prepared to celebrate Harland G. Wood's 70th birthday anniversary in Cleveland on September 9, 1977. They express my indebtedness for his inspirational contributions to biochemistry and my great personal affection for him.

mechanism applies to RNA polymerases as well. Whether the template is DNA or RNA, whether the chain synthesized is RNA or DNA, the mechanism is the same. The only major distinction between RNA and DNA polymerases is in the capacity to start a new chain. An RNA polymerase has the capacity to scan the DNA encyclopedia, select a given passage signalled by a phrase called a promoter, and transcribe only that message. Not so for DNA polymerases. None is yet known which can initiate a chain. Given a start or primer, a DNA polymerase copies a template with great fidelity from that start to as far as the template goes. How then is a DNA chain initiated?

Additional complexities in DNA replication were also inferred from genetic studies. In 1969 De Lucia and Cairns[3] isolated an *E. coli* mutant that was deficient in DNA polymerase I, the principal DNA-polymerizing activity in extracts of *E. coli*. Examination of the residual DNA-polymerizing activity in such mutants led to the isolation of two new DNA polymerases, named polymerase II[4,5,6] and polymerase III.[7] It was not clear which one or more of these polymerases was responsible for chromosome replication.

At about this time, genetic studies from several laboratories[8] showed there were at least six genes in the *E. coli* chromosome whose functions were essential for DNA replication. Assays were needed to identify the products of these replication genes, which were named *dna*B, *dna*C, etc. and to determine their functions.

Thus, in 1971, there were four major questions surrounding DNA replication.

1. How is a DNA chain initiated?
2. Which of several DNA polymerases in a single bacterial cell is responsible for synthesis of the chromosome?
3. What are the proteins coded by the multiple replication genes, and what do they do?
4. How is DNA replication regulated?

An answer to this last most important question requires answers to the first three. The parts of a machine must be identified before its regulation can be understood.

In a cartoon drawn at about that time, our ignorance of the molecular details of the growing fork of the *E. coli* chromosome was discreetly covered by a fig leaf. Direct knowledge of these details is still scant. However, inferences which can be drawn from replicative mechanisms of a phage ϕX174 that depends on *E. coli* for its enzymes now suggest a picture of the replicating fork with details that are agreeably explicit.

The twenty or more *E. coli* proteins employed in replication of ϕX174 can be assigned specific functions in the growth of the leading and lagging strands of the *E. coli* chromosome. The sizes, subunit structures, and relative abundances of the proteins will be important for understanding how they are assembled into replication complexes. Still unknown are the proteins whose function is initiation of replication of the *E. coli* chromosome at its origin.

SMALL DNA PHAGES AS PROBES OF CELLULAR REPLICATION

In trying to put the stattered pieces of the replication machinery together again, we had been frustrated in our attempts to obtain a molecularly dispersed system from a bacterium, a system that could sustain the replication of the bacterial chromosome. We turned then in 1971 to the tiniest DNA chromosomes, first to that of coliphage M13 (also known as phage fd and f1). Later it became clear how useful it would be to

study others, such as phages ϕX174 and G4. These phages rely on the host cell for replication enzymes. Their small chromosomes are only 1/1000 the size of that of *E. coli*. Each of these three phages has proved to be a remarkably effective probe for illuminating a distinctive piece of the machinery that the cell uses for replication of its own chromosome or of its extrachromosomal plasmids.

These phage chromosomes contain only about 5000 to 6000 nucleotides in a single-stranded circle (SS). They have several attractive features. They are relatively easy to isolate, characterize, and use as templates. They are converted into a double-stranded circular replicative form (RF), and they require initiation of a new strand, perhaps at a unique signal. And finally, they afford by their apparent simplicity, a strong psychological impetus; surely, one should be able to obtain a soluble-enzyme preparation from cell extracts capable of replicating so simple a chromosome.

CONVERSION OF PHAGE M13 VIRAL STRAND INTO A DUPLEX CIRCLE

We started our work on phage M13 and the initial stage of its replication, i.e., conversion of the single-stranded circle into the duplex circular replicative form.[9] First, we discovered a fundamental mechanism by which a DNA chain is initiated.[10] Second, we learned how to obtain the cellular replicative system in a soluble molecularly dispersed form.[10] Third, we found the complex multisubunit form of DNA polymerase III.[11,12] Incidentally, the technique of preparing lysates by gentle means in which the replicative enzymes for M13 are soluble and from which the host DNA has been removed,[13] also proved applicable to the replicative systems of phages ϕX174, G4, T7, and T4.

The knowledge that RNA polymerase can start chains and that DNA polymerases could extend them suggested a mechanism of DNA-chain initiation. A short transcriptional operation by RNA polymerase might provide a primer for DNA synthesis; this RNA start would later be erased and replaced by DNA. Therefore, we looked to see whether rifampicin, a specific inhibitor of RNA polymerase, would prevent the first step of phage M13 replication.

The life cycle of phage M13 is shown in Figure 1.[14] Rifampicin blocked the conversion of phage M13 single-stranded circular DNA into the duplex replicative form in a rifampicin-sensitive cell, but not in a resistant cell whose RNA polymerase failed to bind the inhibitor. These results were confirmed and extended in studies with the soluble enzyme system. Fractionation and resolution of the factors needed for the phage M13 single-strand conversion to duplex enabled us to reconstitute this operation in four stages (Figure 2).[15]

Stage 1—The DNA is bound by the DNA (single-strand) binding protein,[16,17] and the secondary structure within it is unwound.

Stage 2—A unique region containing a hairpin-like secondary structure[18] is bound and transcribed by RNA polymerase.

Stage 3—The RNA primer is extended by a complex form of DNA polymerase to produce the nearly complete replicative form (RFII); DNA polymerases I, II, and III are inactive. The active multisubunit form of DNA polymerase III (holoenzyme) will be considered below.

Stage 4—RNA is removed, and the gap is filled by DNA polymerase I; the synthetic complementary strand is sealed by ligase to form the final replicative form (RFI).

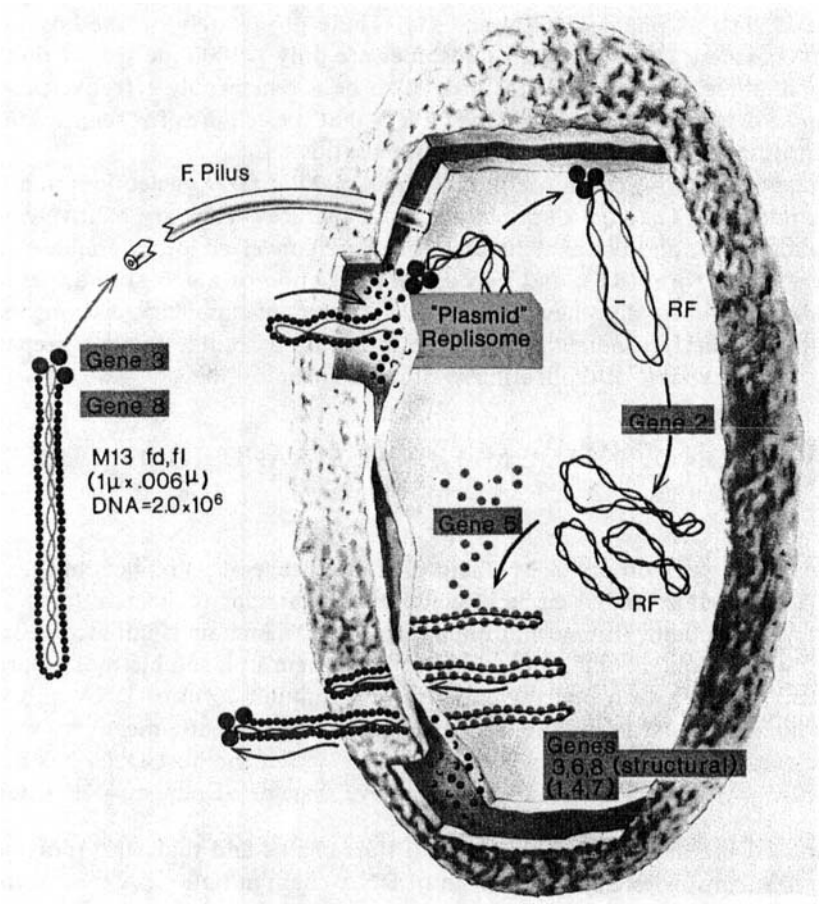

FIGURE 1. Life cycle of phage M13 (also called phage fd or fl). The phage is 1 μm long, 0.006 μm wide, and has 2 × 10⁶ DNA base pairs. RF is the circular, duplex, replicative form.

WHY DOES RNA PRIME DNA REPLICATION?

When we reflect on why RNA is the device for starting a DNA chain, we wonder why a DNA polymerase was not fashioned to do it. The reason may be to ensure the ultra-high fidelity of the DNA record. At or near the start of a DNA chain, the proof-reading and error-correcting devices built into DNA-replicative mechanisms may not function as well in removing base pairing errors as during chain growth. RNA messages are designed to be transient, whereas DNA is essentially indelible. Thus, the start of a DNA chain marked by a foreign RNA-DNA hybrid can be readily recognized and excised and the resulting gap filled by error-free DNA replication.

The priming mechanism for phage M13 DNA replication by RNA polymerase action was later shown to apply to certain extrachromosomal elements or plasmids, such as the colE1 factor.[19] However, it clearly could not apply to replication of the *E. coli* chromosome. In the Okazaki et al.[20] model for discontinuous chromosome growth, chains are initiated at the rate of one a second at the replication fork, even in the presence of rifampicin. Therefore, RNA priming of DNA replication is either not a general mechanism, or it may also be performed by a rifampicin-resistant RNA polymerase, distinct from the classic rifampicin-sensitive enzyme.

When it was learned that the first stage of phage ϕX174 replication (single-strand to

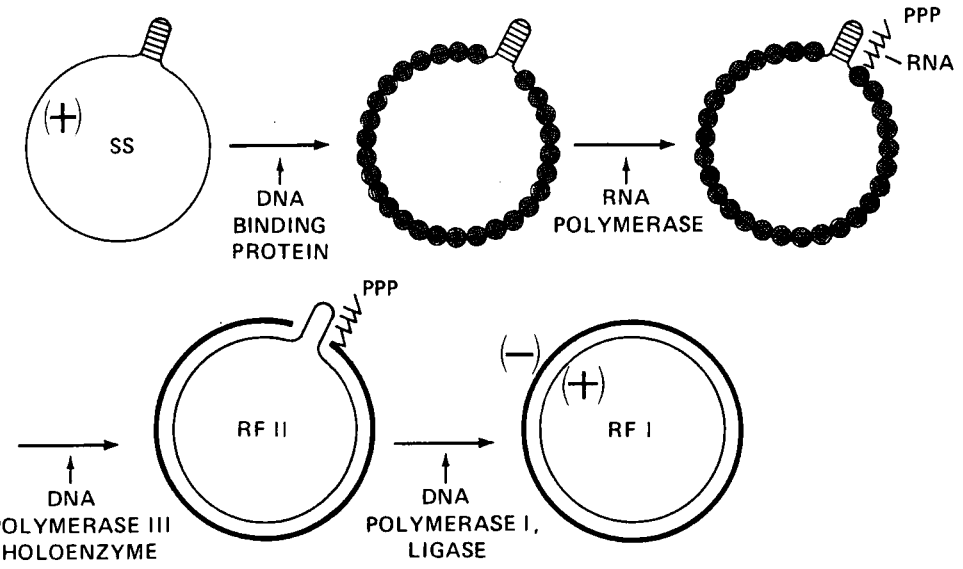

FIGURE 2. Scheme showing proteins and stages involved in reconstitution of conversion of M13 viral DNA into RFI DNA.

duplex conversion) was also insensitive to rifampicin,[21] we naturally turned to this system to find the alternative to the phage M13 mechanism.

CONVERSION OF ϕX174 VIRAL STRAND INTO A DUPLEX CIRCLE

The life cycle of phage ϕX174 is shown in Figure 3.[22] We have focused for the past 5 years on the resolution and reconstitution of the rifampicin-insensitive cellular machinery appropriated by phage ϕX174 DNA for its conversion into the replication form.[23] It is the very machinery used by the cell for replication of its own chromosome. The subsequent stage in which the phage ϕX174 replicative form is multiplied requires participation of additional components, the phage ϕX174 gene A (cistron A) product, and the *E. coli rep* gene product; these will be discussed later.

A combination of assays has been used to fractionate and identify the components of the system for the single-strand to duplex conversion of phage ϕX174 DNA.[24-27] One assay is the complementation of a crude fraction prepared from a thermosensitive replication mutant; the other is reconstitution of partially purified components resolved and separated by fractionation procedures. Separate purification procedures were devised for each component to obtain optimal yields and purity.

Four of the components (*dna*B protein, *dna*C protein, *dna*G protein, and the *dna*E subunit of polymerase III holoenzyme) could be assayed by complementation. However, four of the components (protein i, protein n, binding protein, and the β subunit of polymerase III holoenzyme) are not yet identified with any genetic locus, and their assay depends on reconstitution of a mixture of partially purified constituents.

Some believe that elucidating a complex biosynthetic pathway requires mutants at each step. However, in this instance, many of these replication proteins have been purified without having or using such mutants. It should also be clear that removal of a particular component as drastic as by a genetic deletion can be achieved by treatment of a fraction with $(NH_4)_2SO_4$, an ion exchanger, or a gel-filtration column.

Eight protein fractions incubated with the four ribonucleoside triphosphates and the four deoxyribonucleoside triphosphates, spermidine, and Mg^{2+} sustained the replica-

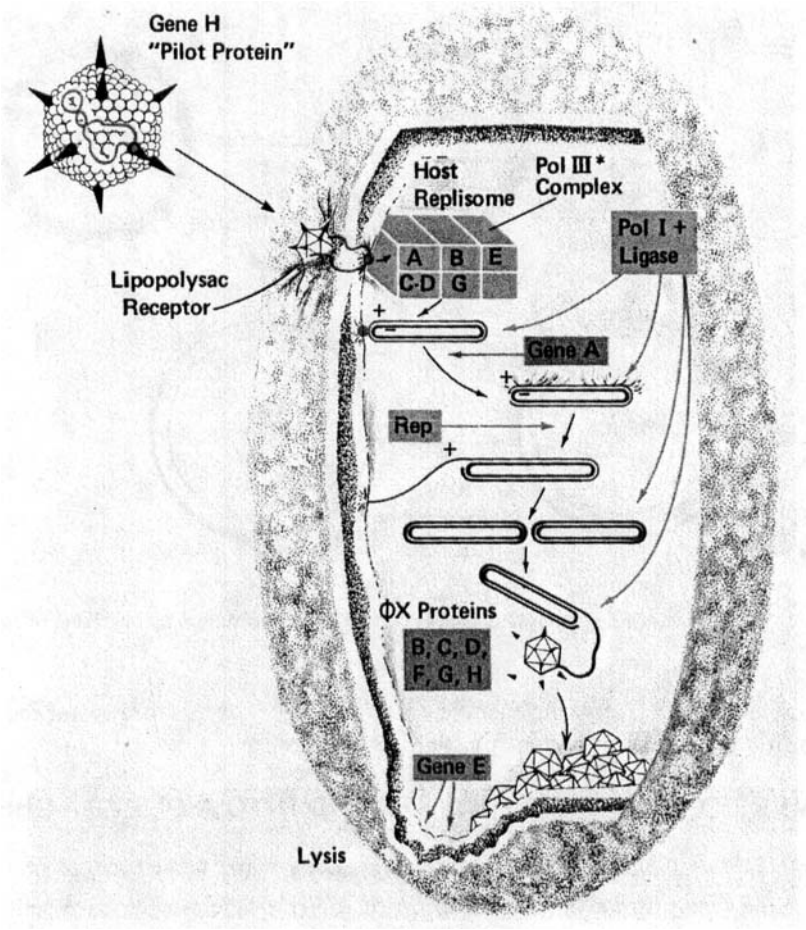

FIGURE 3. Life cycle of phage ϕX174.

tion of the phage ϕX174 viral template (Table 1); omission of any one component decreased synthesis by tenfold or more. Closure of the RFII DNA requires, as with phage M13, DNA polymerase I and ligase. For lack of RNA polymerase, no combination of proteins in Table 1 was able to utilize phage M13 DNA as a template.

What are the functions of the many proteins needed for converting the single-stranded form into the RFII form? Only the functions of the binding protein and the DNA polymerase III holoenzyme could be surmised. Fortunately, Nigel Godson's curiosity about ϕX174-like phages led him to discover phage G4 which is grossly similar to phage ϕX174, but differs considerably in nucleotide sequence.[28] It was phage G4 that enabled us to examine the function of *dna*G protein. Although conversion of phage G4 DNA into the RFII form is like that of phage ϕX174, a rifampicin-insensitive operation, the requirements are strikingly simpler: only binding protein, *dna*G protein, and the DNA polymerase III holoenzyme are needed; four proteins essential for phage ϕX174 are dispensable for the phage G4 conversion.[29]

CONVERSION OF PHAGE G4 VIRAL STRAND INTO A DUPLEX CIRCLE PRIMED BY *dna*G PROTEIN: ACTION OF PRIMASE

The facility and simplicity of the phage G4 conversion into the replicative form provided an assay for the isolation of homogeneous binding protein,[17] *dna*G protein,[30]

TABLE I

Proteins Required for the Reconstitution of the Replication of ϕX174, G4, and M13 Viral DNA Molecules

Component omitted	Template for synthesis (amount required [pmo])		
	ϕX174	G4	M13
None	64.0	102	1.8
Binding protein	0.7	0.6	1.2
*dna*B Protein	2.0	94	0.6
*dna*C Protein	2.3	99	0.8
Protein i	2.7	103	0.9
Protein n	1.8	86	0.4
*dna*G Protein (primase)	3.3	6.7	0.5
Polymerase III holoenzyme	0.9	8.6	0.5

and the subunits of DNA polymerase III holoenzyme.[31] Their actions can be separated into three stages (Figure 4): (1) DNA binding by binding protein; (2) RNA synthesis by *dna*G protein to yield a primed single-strand form as an intermediate; (3) DNA synthesis by the holoenzyme through covalent extension of the short RNA primer.

In earlier studies (in the absence of the RFII closure by DNA polymerase I and ligase),[29] the small gap in the RFII DNA was located about 5% of the circle length from the single EcoR1 site; the 3′ strand end was located at the left end of the gap, the 5′ origin, at the right end. The synthetic strand was dissociated from the template by formamide rather than alkali in order to preserve the RNA.[29] The synthetic strand was examined in formamide gradients before and after cleavage near the EcoR1 site of the RFII DNA. The RNA (labeled with ^{32}P-labeled ribonucleotides) sedimented with the DNA (labeled by ^{3}H-labeled deoxyribonucleotides) to a 14S position, as expected of a nearly full length linear strand. By contrast, after cleavage at the EcoR1 site of the replicative form, all the RNA was shifted to a position of a small fragment of about 5S (about 200 nucleotides long) accompanied by about 5% of the synthetic DNA. This is the size of DNA expected from the distance between the origin and the EcoR1 cleavage site.

The RNA transcript was 29 nucleotides long and was utilized as a primer in the DNA synthesis stage. RNAase (ribonuclease) digestion and analysis of the EcoR1 fragment yielded a unique RNA sequence starting from the origin (Figure 5).[32]

A duplex (hairpin) structure containing seven G·C base pairs and one A·U pair in this sequence represents a complementary structure in the G4 viral strand which has recently been located by direct sequence analysis.[53] Presumably, this is the structure that resists unwinding by the binding protein and serves as the promoter for polymerase action by the *dna*G protein.

We imagine that a secondary structure in the *E. coli* chromosome, analogous to that in phage G4 DNA, serves as the promoter for *dna*G-protein action in starting nascent fragments.[33] Creating such a structure in certain regions of phage ϕX174 DNA may be the job of the several proteins needed for phage ϕX174, but not for phage G4, replication.

Deoxyribonucleotides can substitute in part for ribonucleotides, and priming coupled to DNA synthesis produces a hybrid primer only about 6 residues long rather than the 29-residue RNA transcript.[34] The sharp specificity of *dna*G protein in choice of template and its capacity to utilize deoxyribonucleotides as well as ribonucleotides to produce effective primers for DNA synthesis suggests to us that *dna*G protein, previously named RNA polymerase, be renamed primase.

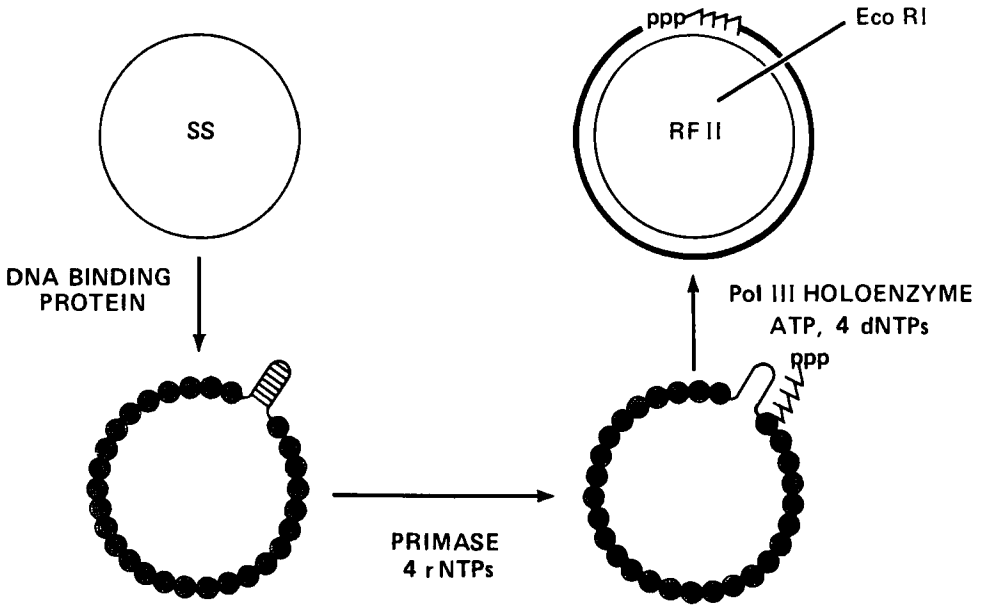

FIGURE 4. Scheme showing proteins and stages involved in reconstitution of conversion of G4 viral DNA into RFII DNA.

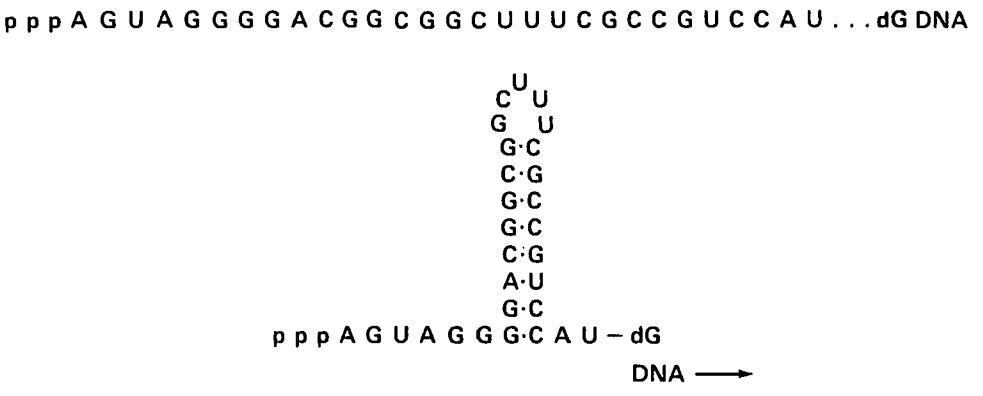

FIGURE 5. Sequence for the RNA primer initiating replication on G4 viral DNA.

MULTIPLE STAGES IN THE ϕX SINGLE-STRAND CONVERSION TO A DUPLEX

In attempting to separate the conversion of a single strand to a duplex into its component stages, it was learned that prior incubation of the phage ϕX174 circular DNA with only five of the required proteins, but in the absence of any RNA or DNA synthesis, permitted subsequent replication to proceed 20 times faster.[35,36] The synthetic reaction was complete in less than 1 min, rather than in 20 min. Formation of this intermediate required incubation at 30°C. The intermediate was isolated quantitatively by filtration or by sedimentation. Among the requirements for forming the intermediate,[36] the *dna*G protein, holoenzyme and nucleotide substrates were dispensable. ATP was essential and spermidine was stimulatory. Kinetic studies suggested that of the five proteins required, the binding protein, *dna*B protein, and *dna*C protein were used stoichiometrically, whereas proteins i and n served catalytically.

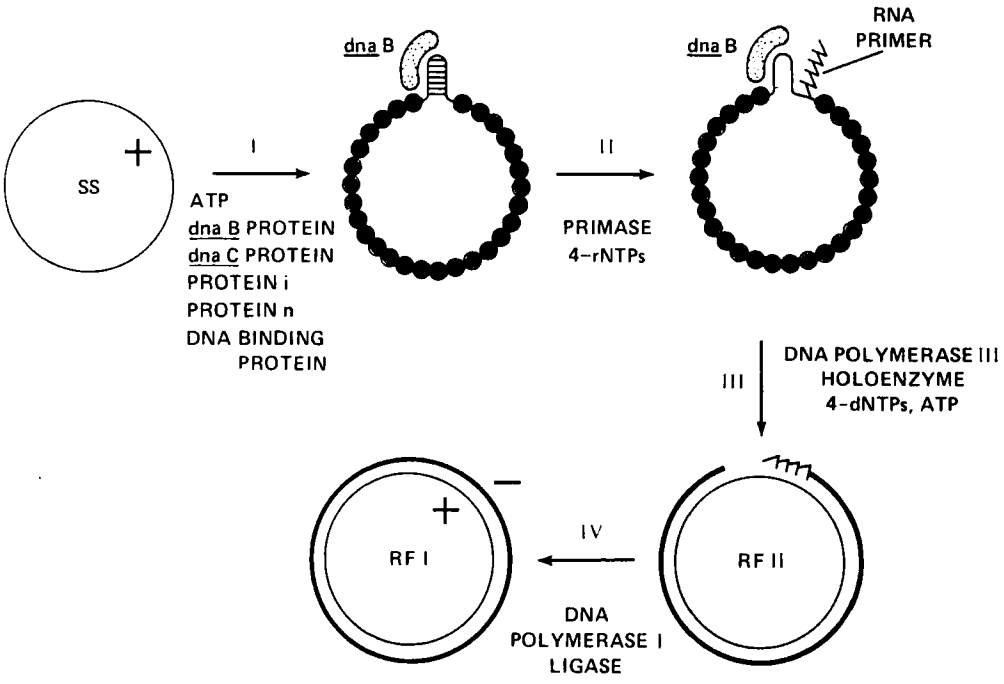

FIGURE 6. Scheme showing proteins and stages involved in reconstitution of conversion of ϕX174 viral DNA into RFI DNA.

Antibodies directed against the individual proteins were helpful in analyzing the sequence in which each of these proteins acted and in determining their presence in replication intermediates. Rabbit immunoglobulins against binding protein, *dna*B protein, protein i, and protein n, each prevented formation of the intermediate. When added after the intermediate had been formed, antibodies against the first two blocked the subsequent functioning of the intermediate, implying the presence of these components in it. We attribute the failure of antibodies against protein i and protein n to inhibit the action of the intermediate to the absence of these proteins from the intermediate, although their presence in an inaccessible form cannot be dismissed. The inhibitory action of the antibody to *dna*G protein simply reflects the need for the action of the protein in utilizing the intermediate for replication.

A hypothetical scheme for the stages in which the many proteins participate in the phage ϕX174 single-strand → duplex conversion is shown in Figure 6.

*dna*B PROTEIN AS A "MOBILE REPLICATION PROMOTER"

In order for primase to act on ϕX DNA, a replication intermediate must first be produced. The intermediate contains one molecule of *dna*B protein as an essential component for the synthesis of primer by primase (Figure 6). Because priming of the replication intermediate, uncoupled from replication, gives rise to many primers on a single circle, we propose that processive movement of *dna*B protein on the template strand permits it to act as a "mobile replication promoter" (Figure 7). This mechanism for *dna*B protein action may also apply to the initiation of a cycle of replication at the *E. coli* origin and to the initiation of primer transcripts for nascent (Okazaki) fragments in discontinuous replication at the replicating fork of the chromosome.

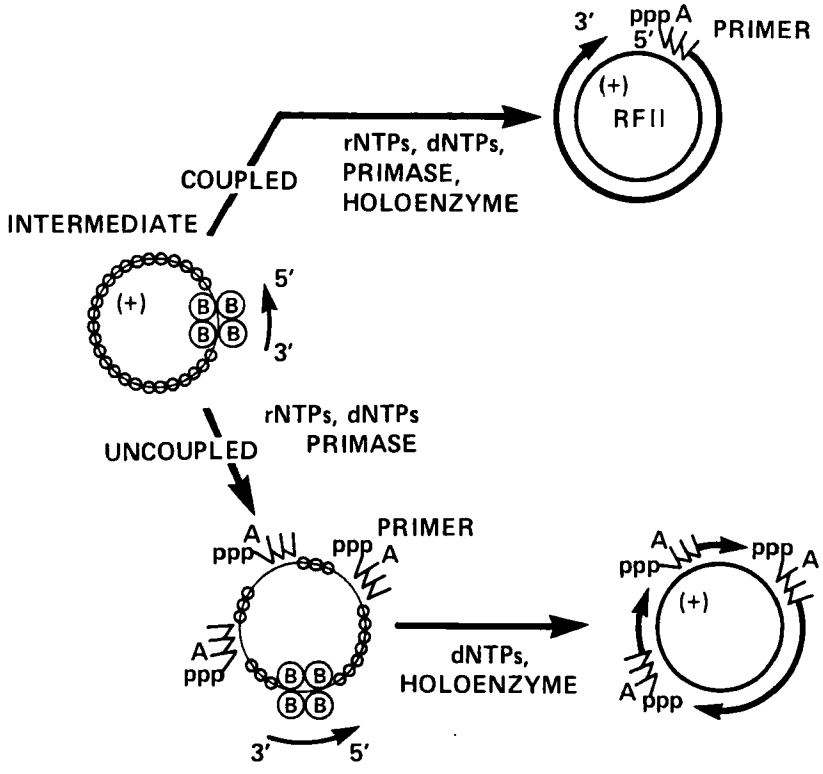

FIGURE 7. Model for role of *dna*B protein in priming bidirectional DNA replication at chromosomal origin of replication.

INITIATION OF REPLICATION: MULTIPLE MECHANISMS

The first stage of replication of phages M13, ϕX174, and G4 has disclosed in each instance a distinctive replicative system. The principal distinction is in the initiation mechanism. The rifampicin-sensitive (RNA polymerase) system used by phage M13 probes the cellular mechanism for replicating extrachromosomal DNA; the rifampicin-insensitive (*dna*G protein) system used by phage ϕX174 elucidates the cellular mechanism for nascent-chain initiation and growth of the chromosome itself; the simpler phage G4 system which does not require the *dna*B or *dna*C proteins may be revealing an especially available promoter for *dna*G-protein action. Would a survey of other small DNA coliphages disclose any novel replicative systems beyond the three mentioned? So far, analysis of nine additional phages shows them all falling into one or another of these three categories: phages fd and PF1 are M13-like; phages ST-1 and ϕX-tb are G4-like; and phages S13, G6, G13, G14, and ϕXahb are ϕX174-like.[54]

ELONGATION: DNA POLYMERASE III HOLOENZYME

The composition and structure of DNA polymerase III holoenzyme remains a complex issue. In addition to DNA polymerase III (pol III), the *dna*E-gene product (determined to have a molecular weight of 140,000 in one strain and identified previously with a subunit of molecular weight 90,000 in another), there are at least four and very likely more subunits.[31] Pol III is released from holoenzyme by dilution or aging. A "conversion factor" which converts pol III back into an active form has been purified from holoenzyme by inactiving pol III with the Zn^{2+} chelator, *o*-phenanthroline; this

component is approaching homogeneity after a 25,000-fold enrichment. It contains two polypeptides of molecular weights 52,000 and 32,000. One of them is the *dnaZ*-gene product, as judged by complementation of deficient mutants and by its enhancement in strains carrying a plasmid with this gene.[27,31]

MULTIPLICATION OF PHAGE ϕX174 REPLICATIVE-FORM DNA: REPLICATION OF A DUPLEX DNA

Replication of phage ϕX174 replicative-form DNA (RFI) is known from genetic studies to require at least two additional proteins. One is phage-induced *cis*A protein, whose function is to nick the twisted circular replicative-form DNA at one particular place in the (+) strand.[37,38] The other is the product of the *E. coli rep* gene; mutants in this gene fail to sustain phage ϕX174 replicative-form DNA replication and slow down the fork movement of the host chromosome.[39]

A soluble enzyme preparation was observed to replicate RF. Both (+) and (−) strands of the RFI and RFII DNA products were synthesized. The *rep* protein (65,000 daltons) was purified from wild-type cells by complementation of extracts from *rep* mutants; *cis*A protein (60,000 daltons) was purified from infected cells by complementation of extracts from uninfected cells.

These two homogeneous proteins, together with binding protein and DNA polymerase holoenzyme, sustained rapid and extensive synthesis on the RF template. The product consists exclusively of viral (+) circles generated by a rolling circle mechanism (Figure 8).[40] The (+) strand product, like DNA extracted from phage, served as a template for (−) strand synthesis by the multiprotein system described in Figure 6. This (+) strand synthesis, once initiated, is continuous, whereas (−) strand synthesis, requiring a fresh initiation on each circle, is discontinuous.

These two separate mechanisms for (+) strand and (−) strand synthesis suggest at once that all the steps in the phage ϕX174 replicative cycle may be accounted for by one or the other of these mechanisms. Should this generalization hold for phage ϕX174, we suggest that the replication patterns of other circular DNA and even linear DNA molecules are all based on two mechanisms: a relatively continuous replication of the leading strand and a separate discontinuous mechanism, depending on initiation of DNA chains of the lagging complementary strand.

ϕX174 CISTRON A PROTEIN IS A MULTIFUNCTIONAL ENZYME IN DNA REPLICATION

The *cis*A protein of ϕ174 nicks the viral strand of the superhelical ϕX174 duplex DNA, at the same time forming a complex with the DNA. The protein, seen bound to the DNA in the electron microscope (Figure 9), was found in the restriction-endonuclease fragment between nucleotides 4290 and 4330 on the ϕX174 map.[41] Replication was also observed to initiate at this point, thus identifying the site of *cis*A nicking and binding as the origin of replication.

The *cis*A-DNA complex (separated from free *cis*A), upon the addition of *rep* protein, ATP, and binding protein, is unwound to generate a single-stranded linear (presumably the nicked (+) strand) and a circular molecule (presumably the (−) strand) (Figure 10). The *cis*A-DNA complex, upon addition of DNA polymerase III holoenzyme and deoxynucleoside triphosphates, supports replication to generate viral, single-stranded circles, as many as 15 circles per *cis*A-DNA complex (Figure 10).

Replicating intermediates visualized in the electron microscope show a thickened single-strand loop, coated with binding protein, and attached to a duplex circle. Contour lengths of the loops in Figure 11 show that 20, 30, and 60%, respectively, of the

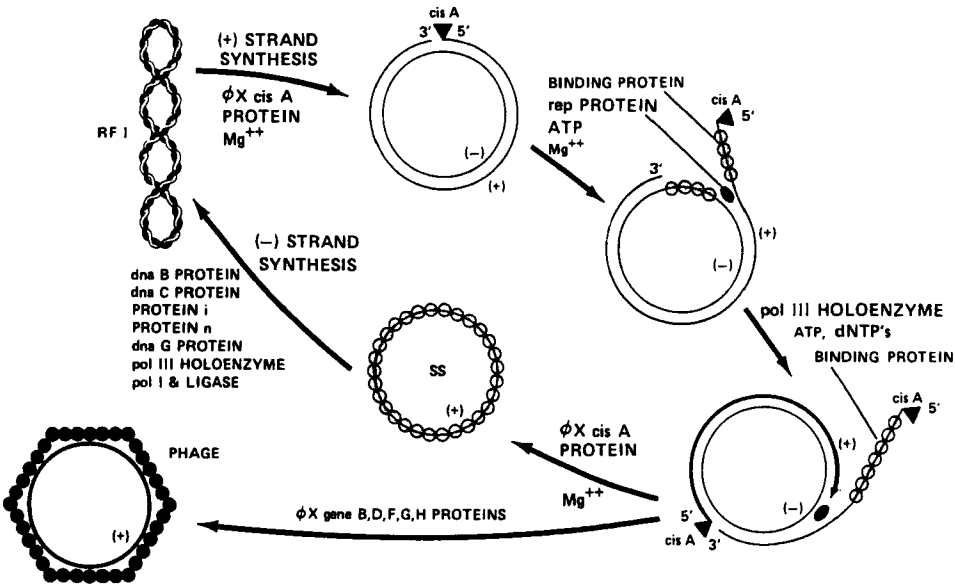

FIGURE 8. Hypothetical scheme for some of the replicative events in the life cycle of ϕX174.

duplex circles had been replicated. (A factor of 2.56 was applied to correct for condensation of the single strand by binding protein). No intermediates with free ends were found. Rolling circles with free single-stranded tails were found when the purified DNA was examined by electron microscopy (40% formamide) without prior crosslinking of protein to DNA.

Thus, the 5′ end of the displaced strand, presumably with *cis*A protein attached, is at the replicating fork, moving around the circular template, and probably complexed with *rep* protein.

The roles of *cis*A protein and *rep* protein in DNA replication are described schematically in Figure 11. The *cis*A protein fulfills multiple functions: (1) nicking of the viral (+) strand at a specific sequence, the origin, to initiate a round of replication, (2) participating in the unwinding of the two strands to create the replication fork, presumably by interacting with the *rep* protein, (3) excising a ϕX genome-length DNA after completion of a full round of replication, and (4) ligating the excised DNA to form a covalently closed, circular molecule.

The unique nucleotide sequence in a single-stranded region, exposed in superhelical RFI DNA, and nicked by the *cis*A protein is located in the viral (+) strand, between nucleotides 4290 and 4330 on the ϕX DNA sequence map. The nicking results in what we judge to be a covalent attachment of the *cis*A protein to the 5′ end of the viral strand. This linkage, as in the case of the untwisting enzymes (which carry out repeated nicking and ligating steps without an energy requirement), conserves the energy of the cleaved phosphodiester bond. To explain two successive nicking steps followed by ligation carried out by the bound *cis*A protein, we propose two active sites or subunits in the bound *cis*A protein, only one of which (open circle in Figure 11) is bound covalently to the 5′ end in a given complex.

We suggest that *rep* protein recognizes the origin by its interaction with bound *cis*A protein. Unwinding of the two strands, sustained by ATP hydrolysis and binding protein, then ensues. Movement of the 5′ end into the duplex would displace the viral (+) strand from its complementary (−) circular template. The 3′ hydroxyl terminus, at the origin, serves as a primer and is extended by holoenzyme to regenerate the origin.

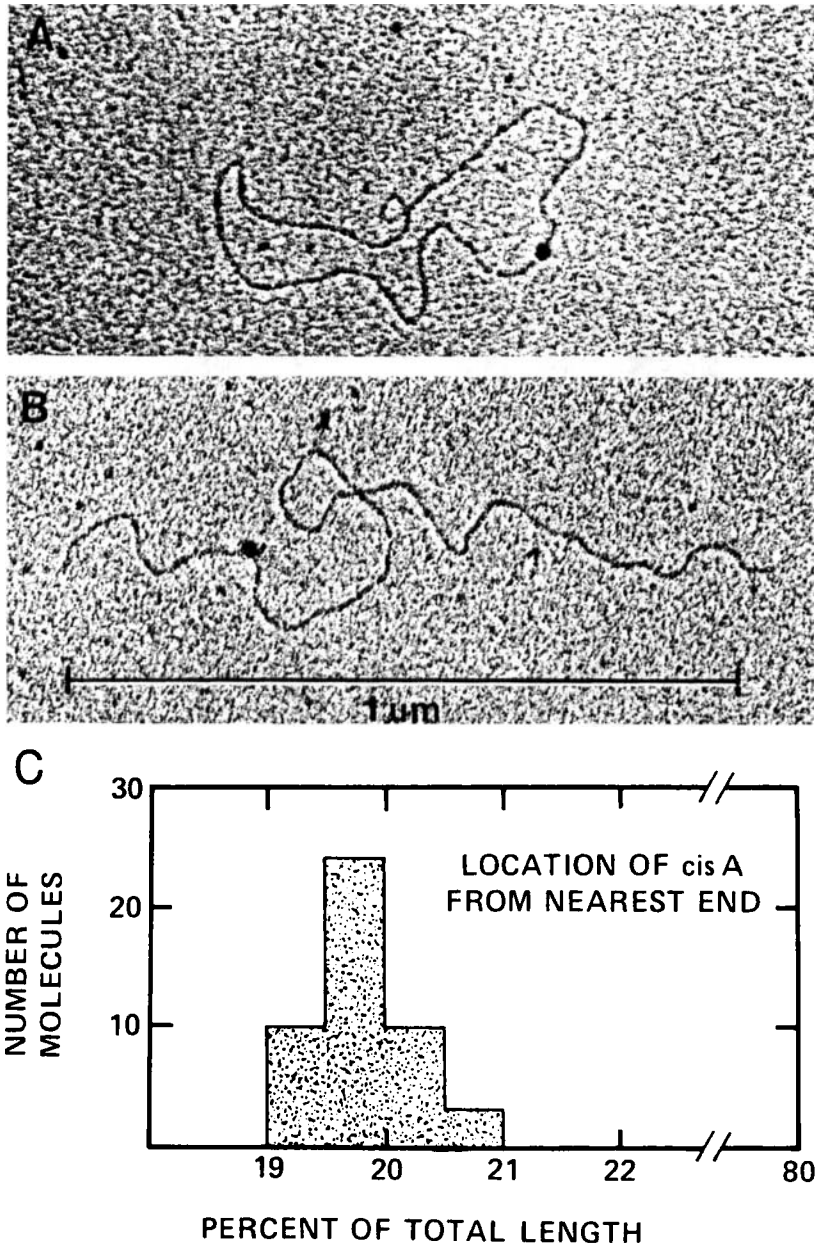

FIGURE 9. Specific location of *cis*A protein on ϕX RFII.

When the *cis*A protein-*rep* protein complex traverses the full length of the template circle, a genome-length ϕX DNA would be excised by the unoccupied active site of the *cis*A protein (solid circle in Figure 11) by displacing and attacking the regenerated origin in the viral (+) strand. Concurrently with the nucleolytic scission carried out by the second *cis*A subunit (solid circle in Figure 11), the first subunit of the *cis*A protein (open circle) ligates the newly created 3′ hydroxyl with the bound 5′ end, using the conserved energy of the covalent protein-DNA bond. In this mechanism using two active sites or subunits of the *cis*A protein alternatively, the *cis*A protein-RFII complex

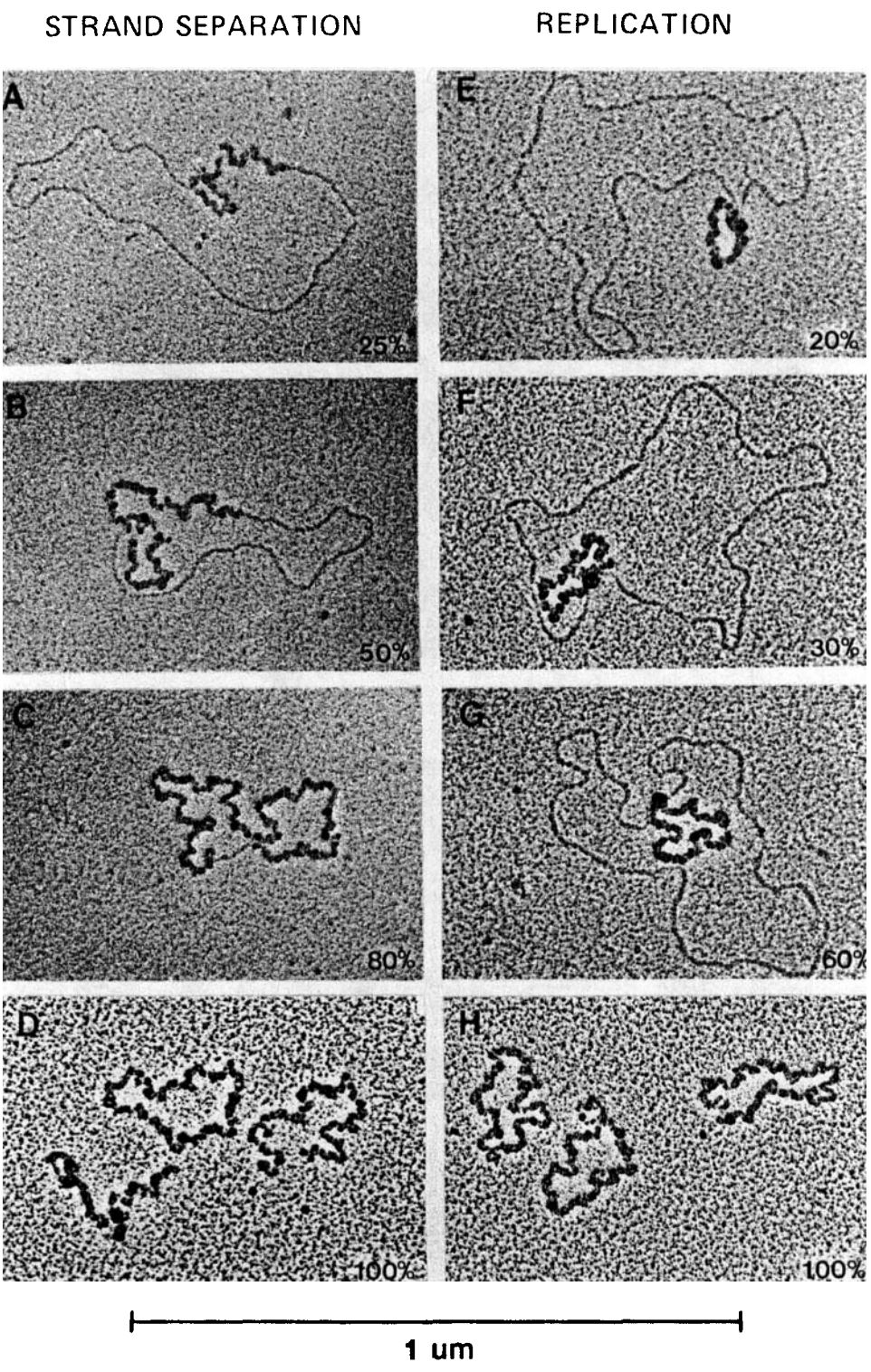

FIGURE 10. Intermediates in RFII strand separation and replication reactions.

is restored with the completion and release of each viral circle, and the catalytic role of the complex is assured. This mechanism provides for efficient use of *cis*A protein

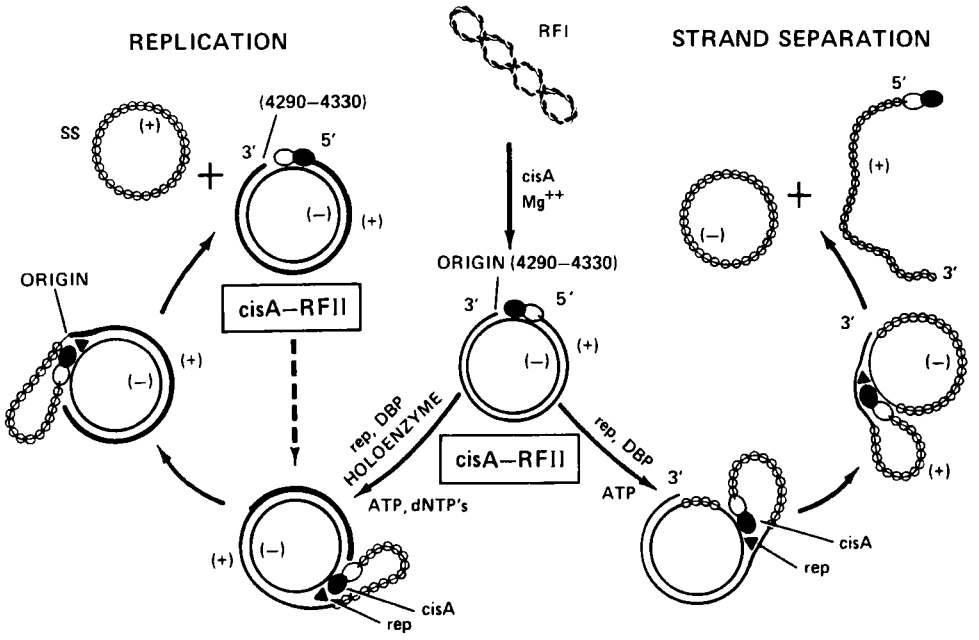

FIGURE 11. Scheme for *cis*A protein action illustrating its multiple functions.

and its proper orientation for the continuous synthesis of viral (+) DNA on a single template.

Initiation of complementary (−) strand synthesis by host proteins (including *dna*B, *dna*C, i, and n proteins, primase, holoenzyme, and other proteins) takes place either on the displaced viral loop of the replicating intermediate or on the released viral (+) circle. Further studies in which the purified (+) strand synthesizing system is coupled with that for the (−) strand are now required to answer this question.

A mechanism analogous to ϕX RF replication may apply in the replication of other chromosomes. Initiation of synthesis of one strand by extension of a nick introduced by *cis*A-like protein represents an origin of replication. (In chromosomes that replicate bidirectionally, nicking at the origin would involve both strands.) This "leading" strand, synthesized continuously, is analogous to the ϕX viral (+) DNA. The other "lagging" strand, analogous to the ϕX complementary (−) DNA, would be initiated repeatedly on the exposed parental (+) template.

THE *rep* PROTEIN OF *E. coli*: AN ATPase WHICH SEPARATES DUPLEX DNA STRANDS IN ADVANCE OF REPLICATION

The product of rep gene of *E. coli* catalytically separates ϕX174 duplex DNA strands in advance of their replication, utilizing ATP in the process. The enzyme has now been purified to homogeneity. Relatively large quantities of the homogeneous enzyme were obtained from colE1-plasmid-containing cells in which the level was 7 to 10 times above wild type.[42]

Hydrolysis of ATP by *rep* requires a single-stranded region of DNA four residues long.[43] The true effector for *rep* ATPase appears to be a replicating fork rather than a random coil. At or near a fork in duplex DNA, *rep* ATPase action is different from what it is on DNA lacking secondary structure (single-stranded). At or near a fork: (1) Km for ATP is tenfold lower (about 25 μM), (2) specificity is for ATP and dATP with no action on other nucleoside triphosphates, (3) sensitivity to certain ATP analogs

TABLE 2

Escherichia coli Replication Proteins

Polypeptide	Size (K)	Subunits	Function	Unamplified yield: Molecules per cell	Unamplified yield: mg/kg[a]	Amplification
DBP	74	4	Binding DNA	300	20	—
*dna*B	250	4	Mobile promoter	20	0.3	5
*dna*C[b]	29	1	Prepriming	200	—	—
protein i	80	4	Prepriming	150	0.5	—
Protein n[b]	—	—	Prepriming	—	—	—
Primase	60	1	Primer formation	100	0.2	—
Pol III holo	300	4	DNA synthesis	20	0.2	—
Pol III (α)	140	1	DNA synthesis	—	—	—
Copol III (β)	40	1	DNA synthesis	—	—	—
52K (γ)	52	1	DNA synthesis	—	—	10
32K (δ)	32	1	DNA synthesis	—	—	
Pol I	109	1	DNA synthesis	250	10	70
Ligase	74	1	Ligation	300	11	500
Gyrase	140	—	Supertwisting	—	—	—
rep	65	1	Strand separation	30	0.6	10
*cis*A (ϕX174)	60	1	Initiation	3000	20	—
dUTPase	64	4	dUTPase	350	3	—

[a] Wet weight.
[b] dnaC, 30% pure; protein n, 50% pure.

is reduced, (4) presence of a DNA-nicking enzyme (e.g., *cis*A induced by ϕX174) is required, and (5) *E. coli* DNA-binding protein facilitates rather than inhibits. During the separation of strands accompanying replication, two molecules of nucleoside triphosphate (ATP or dATP) are hydrolyzed for every nucleotide polymerized. Utilization of ATP by *rep* protein may provide energy for catalytic strand separation at a fork in advance of replication.

SUMMARY OF THE REPLICATION PROTEINS

Table 2 lists the proteins that we believe are needed to convert phage ϕX174 single-stranded DNA into the duplex replicative form; DNA polymerase I and ligase would be needed to complete and seal the RFII form to the RFI form and gyrase to supercoil it. Some very tentative estimates are given of their abundance, size, and subunit structure. With respect to their functions, the binding protein binds the DNA except at a promoter region. Primase recognizes that region to make a primer, which is extended by the DNA polymerase III holoenzyme, possibly assisted by protein u. This leaves four proteins without an assigned function: *dna*B protein, *dna*C protein, and proteins i and n. These proteins are needed to prepare the phage ϕX174 circular DNA for the transcriptional (priming) step by primase.

The low concentrations of the replication proteins makes their isolation and characterization difficult. In order to increase their concentrations, we have used recombinant DNA molecules to obtain strains in which colE1 plasmids incorporate one of the replication genes; examples are genes *dna*B, *dna*C, and *dna*Z. So far, large-scale cultures of strains containing the *dna*B and *dna*Z genes in colE1 plasmids have yielded extracts in which the concentrations of these proteins are increased fivefold or more. Lysogenic phages are successful vectors for amplifying the genes of ligase[44] and DNA polymerase I.[45]

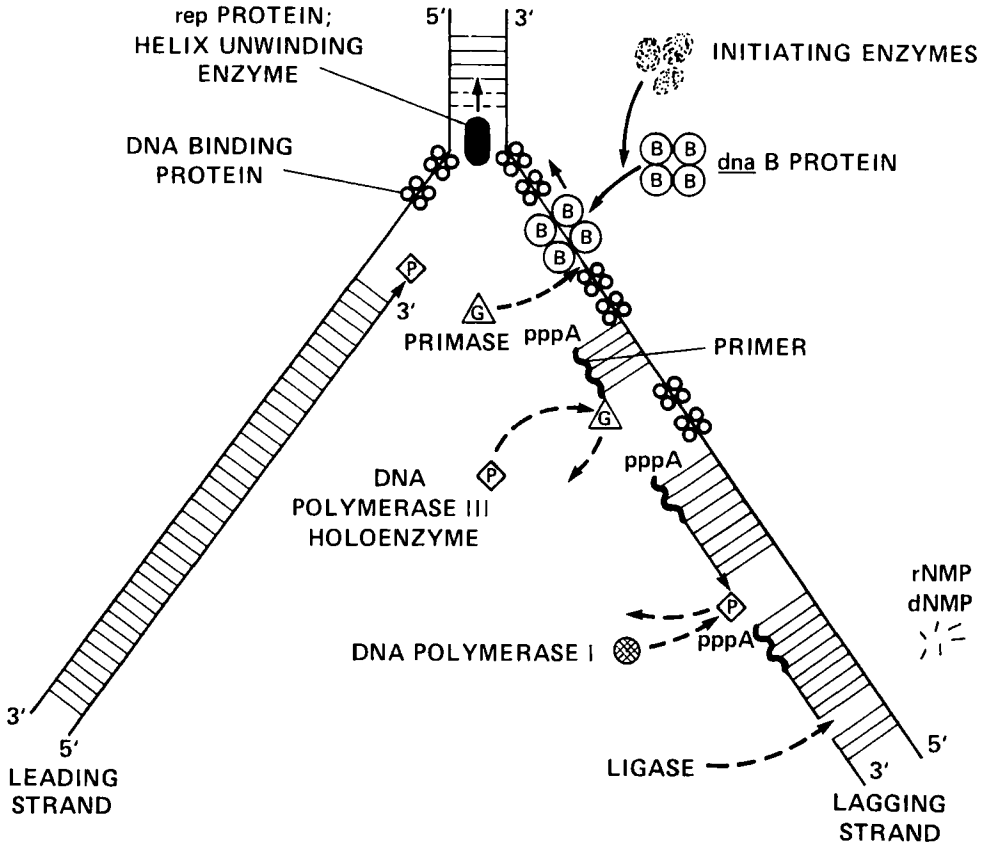

FIGURE 12. Hypothetical scheme for a replication fork of the *Escherichia coli* chromosome.

THE REPLICATING FORK

Some translucency has now been introduced into the fig leaf which previously covered the anatomy and dynamics of the replicating fork. From studies of phage replication, we begin to discern some of the components and their functions (Figure 12).

A leading strand, started at the origin of replication of the chromosome, advances continuously. Its progress is made possible by the helicase action of *rep* protein. Binding protein stabilizes the single-stranded DNA for template function. The bared opposite strand becomes the template for discontinuous synthesis of the lagging strand. Delivery of a *dan*B protein molecule to this strand by prepriming proteins sets the stage for initiation of a synthetic strand, an Okazaki fragment. Primase transcribes a very brief segment with ribonucleotides and deoxyribonucleotides to serve as a primer for covalent extension by DNA polymerase III holoenzyme. Elongation may average 1000 or so nucleotides.

Processive movement of the *dna*B protein in the 5′ → 3′ direction of the strand to which it is bound enables it to follow the progress of the replicating fork. It enables primase to make successive discontinuous initiations of the lagging strand. Removal of the RNA-DNA hybrid primers at the 5′ end of each of the Okazaki fragments and gap-filling by DNA polymerase I brings 3′-hydroxyl and 5′-phosphoryl ends into juxtaposition. DNA ligase can then seal these ends, thus uniting the nascent fragments into a covalently intact progeny duplex.

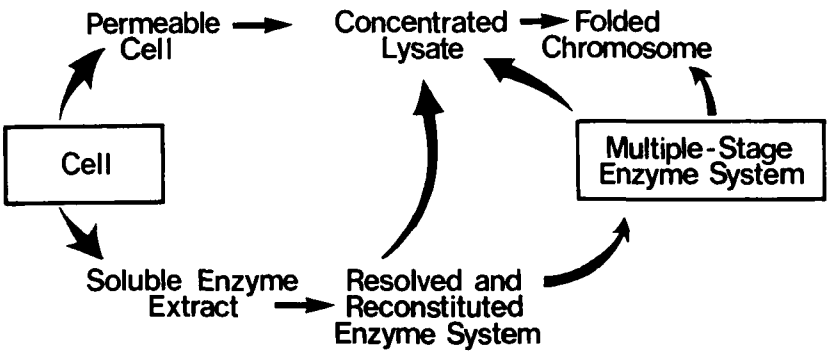

FIGURE 13. Approaches to DNA replication in vitro.

APPROACHES TO DNA REPLICATION IN VITRO

Because of the instability of the DNA-replication enzyme system, earlier studies had to rely on cells made permeable to substrate molecules[46] or on an organized lysate (immobilized on cellophane discs)[47] permeable to small and large molecules (Figure 13). However, such systems are not amenable to refined analysis of molecular events.

We have attempted by classical enzymatic techniques to resolve and reconstitute the enzymes of DNA replication. We have learned how to obtain at least part of this replication apparatus in a molecularly dispersed form and to resolve it into at least a dozen discrete components. This review has dwelt on our attempts to examine the functions of this reconstituted system in its multiple stages. However, there is an inherent and serious deficiency in this approach. Our assays cannot assure us that we have reproduced the physiological replication event. We may be missing important parts of the replication machinery, or we may have substituted foreign parts derived from other cell machinery. It is therefore essential to test the validity of isolated components and reaction sequences and mechanisms. One approach will be to introduce them into the cruder and more organized systems, such as the gently lysed soluble cell extract or to the Bonhoeffer cellophane disc lysate or to the folded chromosome.[48]

LESSONS ABOUT DNA REPLICATION

From these studies of DNA replication, we have learned the value of certain approaches for enzyme studies.

1. Use an intact chromosome as a template for replication rather than bits and pieces of commercially supplied DNA. Template primers, such as salmon sperm and calf thymus DNA, synthetic homopolymers, and copolymers, have all been useful, but they are unlikely to be reliable as guides to the isolation of the replicating machinery.
2. Use a simple chromosome because of the ease of its isolation, characterization, and performance as template.
3. Use viral DNA molecules as probes to illuminate the host replicative systems that they appropriate. The value of viral probes for study of animal chromosomal replication should also be clear; the viral SV40 and polyoma chromosomes are the same size as phage ϕX174, and that of adenovirus is near the size of phage T7. These animal viruses multiply in the nucleus and use nuclear enzymes for their replication.

4. Use a variety of phage probes. We and the Hurwitz laboratory have used the small phages M13 and ϕX174; Richardson and Scherzinger use the medium-sized phage T7;[49-51] and Alberts et al. used the still larger phage T4.[52] Each opens a window on the basic patterns of replication. In gauging which of these phages to choose, we are aware of the universal law, that the product of the virtue of a system and its limitations is a constant. We often observe that the product of the importance of a protein and its abundance is usually a constant, as is the product of the biological attractiveness of a system and biochemical knowledge of a system.

SIGNIFICANCE OF PROKARYOTIC STUDIES

Insights into the basic mechanisms and control of DNA replication derived from studies of accessible prokaryotic systems are likely to be of general significance for DNA replication throughout nature. Evolutionary conservation of ingenious metabolic designs, which accounts for the unity of biochemical mechanisms, including gene expression, is not likely to have excluded gene replication. The techniques and insights obtained from examining small chromosomes can now be applied to studies of cellular chromosomes. Why is replication active in an embryonic cell and quiescent in an adult one? What controls are altered to transform an adult cell to proliferative growth? Not knowing the working parts of the replication machinery, it is unlikely that we would be able to answer these questions.

Although we know the chemical mechanism of DNA chain growth by polymerase action in general outline, we know little about the factors that insure its rapidity, processivity, and ultrahigh fidelty. Do these factors reside in the multisubunit structure of the enzyme? How are they influenced by interaction with other molecules, large and small? Understanding polymerase action is fundamental to understanding DNA repair and recombination, as well as replication. The "3 R's", repair, recombination, and replication, have a profound importance for the growth and aging of individuals and for the evolution of species.

As little is known about termination of a cycle of replication as of its initiation. How is termination integrated with septation and cell division so as to insure full and equal chromosomal partition between the daughter cells? An association of replicating DNA with the plasma membrane is plausible and has been alleged repeatedly, but there is as yet no firm structural and functional evidence for it.

The massive biomedical research efforts to answer these and related questions about DNA replication are sharply limited because their biochemical foundations are so fragile. Biological edifices built on insecure chemical footings inevitably crumble. The enzymology of DNA replication and its spatial orientation in the cell are compelling in importance for basic biology and the medical and agricultural sciences which rest on it.

ACKNOWLEDGMENTS

I acknowledge with gratitude the contributions of many gifted and motivated students and postdoctoral fellows in recent years: D. Bates, J. -P. Bouché, D. Brutlag, S. Eisenberg, K. Geider, C. McHenry, R. McMacken, L. Rowen, R. Schekman, J. Scott, J. Shlomai, H. Tabak, K. Ueda, A. Weiner, J. Weiner, O. Westergaard, W. Wickner, J. Wiseman, and K. Zechel. I am also indebted to my research associate L. Bertsch. The work has been supported by grants from the National Institutes of Health and the National Science Foundation.

REFERENCES

1. Kornberg, A., Active center of DNA polymerase, *Science*, 163, 1410, 1969.
2. Kornberg, A., *DNA Synthesis*, W. H. Freeman, San Francisco, 1974.
3. DeLucia, P. and Cairns, J., Isolation of an *E. coli* strain with a mutation affecting DNA polymerase, *Nature (London)*, 224, 1164, 1969.
4. Kornberg, T. B. and Gefter, M. L., DNA synthesis in cell-free extracts of a DNA polymerase-defective mutant, *Biochem. Biophys. Res. Commun.*, 40, 1348, 1970.
5. Moses, R. E. and Richardson, C. C., A new DNA polymerase activity of *Escherichia coli*, BBRC, 41, 1557, 1970.
6. Knippers, R., DNA polymerase II, *Nature (London)*, 228, 1050, 1970.
7. Kornberg, T. B. and Gefter, M. L., Purification and DNA synthesis in cell-free extracts: properties of DNA polymerase II, *Proc. Natl. Acad. Sci. U.S.A.*, 68, 761, 1971.
8. Gross, J. D., DNA replication in bacteria, in *Curr. Top. Microbiol. Immunol.* 57, 39, 1972.
9. Brutlag, D., Schekman, R., and Kornberg, A., A possible role for RNA polymerase in the initiation of M13 DNA synthesis, *Proc. Natl. Acad. Sci. U.S.A.*, 68, 2826, 1971.
10. Wickner, W., Brutlag, D., Schekman, R., and Kornberg, A., RNA synthesis initiates *in vitro* conversion of M13 DNA to its replicative form, PNAS, 69, 965, 1972.
11. Wickner, W., Schekman, R., Geider, K., and Kornberg, A., A new form of DNA polymerase III and a copolymerase replicate a long, single-stranded primer-template, PNAS, 70, 1764, 1973.
12. Wickner, W. and Kornberg, A., Holoenzyme form of DNA polymerase III: isolation and properties, *J. Biol. Chem.*, 249, 6244, 1974.
13. Weiner, J., Brutlag, D., Geider, K., Schekman, R., Wickner, W., and Kornberg, A., in *DNA Replication*, Wickner, R. B., Ed., Marcel Dekker, New York, 1974, 187.
14. Marvin, D. A. and Hohn, B., Filamentous bacterial viruses, *Bacteriol. Rev.*, 33, 172, 1969.
15. Geider, K. and Kornberg, A., Initiation of DNA synthesis. VIII. Conversion of the M13 viral single strand to the double-stranded replicative forms by purified proteins, *J. Biol. Chem.*, 249, 3999, 1974.
16. Sigal, N., Delius, H., Kornberg, T., Gefter, M., and Alberts, B., A DNA-unwinding protein isolated from *Escherichia coli*: its interaction with DNA and with DNA polymerases, PNAS, 69, 3537, 1972.
17. Weiner, J. H., Bertsch, L. L., and Kornberg, A., The DNA unwinding protein of *Escherichia coli*: properties and functions in replication, *J. Biol. Chem.*, 250, 1972, 1975.
18. Schaller, H., Uhlmann, A., and Geider, K., A DNA fragment from the origin of single-strand to double-strand DNA replication of bacteriophage fd, *Proc. Natl. Acad. Sci. U.S.A.*, 73, 49, 1976.
19. Williams, P. H., Boyer, H. W., and Helinski, D. R., Size and base composition of RNA in supercoiled plasmid DNA, PNAS, 70, 3744, 1973.
20. Okazaki, R., Sugino, A., Hirose, S., Okazaki, T., Imae, Y., Kainuma-Kuroda, R., Ogawa, T., Arisawa, M., and Kurosawa, Y., in *DNA Synthesis in vitro*, Wells, R. and Inman, R., Eds., University Park Press, Baltimore, 1973, 83.
21. Silverstein, S. and Billen, D., Transcription: role in the initiation and replication of DNA synthesis in *E. coli* and ϕX174, *Biochim. Biophys. Acta*, 247, 383, 1971.
22. Sinsheimer, R. L., in *Progress in Nucleic Acid Research and Molecular Biology*, Vol. 8, Davidson, J. N. and Cohn, W. E., Eds., Academic Press, New York, 1968, 115.
23. Schekman, R., Wickner, W., Westergaard, O., Brutlag, D., Geider, K., Bertsch, L. L., and Kornberg, A., Initiation of DNA synthesis: synthesis of ϕX174 replicative form requires RNA synthesis resistant to rifampicin, PNAS, 69, 2691, 1972.
24. Schekman, R., Weiner, A., and Kornberg, A., Multienzyme systems of DNA replication, *Science*, 186, 987, 1974.
25. Schekman, R., Weiner, J. H., Weiner, A., and Kornberg, A., Ten proteins required for conversion of ϕX174 single-stranded DNA to duplex form *in vitro*, *J. Biol. Chem.*, 250, 5859, 1975.
26. Wickner, S. and Hurwitz, J., in *DNA synthesis and its regulation*, Goulian, M. and Hanawalt, P., Eds., W. A. Benjamin, Menlo Park, Calif., 1975, 227.
27. Wickner, S. and Hurwitz, J., Involvement of *Escherichia coli dnaZ* gene product in DNA elongation *in vitro*, *Proc. Natl. Acad. Sci. U.S.A.*, 73, 1053, 1976.
28. Godson, N., Evolution of ϕX174. Isolation of four new ϕX-like phages and comparison with ϕX174, *Virology*, 58, 272, 1974.
29. Zechel, K., Bouché, J. -P., and Kornberg, A., Replication of phage G4: a novel and simple system for the initiation of DNA synthesis, *J. Biol. Chem.*, 250, 4684, 1975., Bouché, J. -P., Zechel, K., and Kornberg, A., *dnaG* Gene product, a rifampicin-resistant RNA polymerase, initiates the conversion of a single-stranded coliphage DNA to its duplex replicative form, *J. Biol. Chem.*, 250, 5995, 1975.

30. Rowen, L. and Kornberg, A., Primase, the *dna*G protein of *Escherichia coli:* an enzyme which starts DNA chains, *J. Biol. Chem.*, in press.
31. McHenry, C. and Kornberg, A., DNA polymerase III holoenzyme of *Escherichia coli:* purification and resolution into subunits, *J. Biol. Chem.*, 252, 6478, 1977.
32. Bouché, J. -P., Rowen, L., and Kornberg, A., The RNA primer synthesized by primase to initiate phage G4 DNA replication, *J. Biol. Chem.*, in press.
33. Olivera, B. M., Lark, K. G., Herrmann, R., and Bonhoeffer, F., in *DNA synthesis in vitro*, Wells, R. and Inman, R., Eds., University Park Press, Baltimore, 1973, 215.
34. Rowen, L. and Kornberg, A., A ribo-deoxyribonucleotide primer synthesized by primase, *J. Biol. Chem.*, 253, 758, 1978.
35. Wickner, S., Wright, M., and Hurwitz, J., Association of DNA-dependent and independent ribonucleoside triphosphatase activities with *dna*B gene product of *E. coli*, *Proc. Natl. Acad. Sci. U.S.A.*, 71, 783, 1974.
36. Weiner, J. H., McMacken, R., and Kornberg, A., Isolation of an intermediate which precedes *dna*G RNA polymerase participation in enzymatic replication of bacteriophage ϕX174 DNA, *Proc. Natl. Acad. Sci. U.S.A.*, 73, 752, 1976.
37. Henry, T. J. and Knippers, R., Isolation and function of the gene A initiator of bacteriophage ϕX174: a highly specific DNA endonuclease, *Proc. Natl. Acad. Sci. U.S.A.*, 71, 1549, 1974.
38. Singh, S. and Ray, D. S., A novel single strand endonuclease specific for ϕX174 DNA, *Biochem. Biophys. Res. Commun.*, 67, 1429, 1975.
39. Lane, H. E. D. and Denhardt, D. T., The *rep* mutation, *J. Mol. Biol.*, 97, 99, 1975.
40. Eisenberg, S., Scott, J. F., and Kornberg, A., Enzymatic replication of viral and complementary strands of duplex phage ϕX174 DNA proceeds by separate mechanisms, *Proc. Natl. Acad. Sci. U.S.A.*, 73, 1594, 3151, 1976.
41. Eisenberg, S., Griffith, J., and Kornberg, A., The ϕX174 *cistron.* A protein is a multifunctional enzyme in DNA replication, *Proc. Natl. Acad. Sci. U.S.A.*, 74, 3198, 1977.
42. Scott, J. F. and Kornberg, A., Purification of the *rep* protein of *Escherichia coli:* an ATPase which separates duplex DNA strands in advance of replication, *J. Biol. Chem.*, 253, 3292, 1978.
43. Kornberg, A., Scott, J. F., and Bertsch., L L., ATP utilization by *rep* protein in catalytic separation of DNA strands at a replicating fork, *J. Biol. Chem.*, 253, 3298, 1978.
44. Panasenko, S. M., Cameron, J. R., Davis, R. W., and Lehman, I. R., Five-hundred-fold overproduction of DNA ligase after induction of a hybrid lambda lysogen constructed *in vitro*, *Science*, 196, 188, 1977.
45. Kelley, W. S., Chalmers, K., and Murray, N. E., Isolation and characterization of a γ*pol*A transducing phage, *Proc. Natl. Acad. Sci. U.S.A.*, 74, 5632, 1977.
46. Vosberg, H. -P. and Hoffmann-Berling, H., DNA synthesis in nucleotide-permeable *Escherichia coli* cells. I. Preparation and properties of ether-treated cells, *J. Mol. Biol.*, 58, 739, 1971.
47. Schaller, H., Otto, B., Nusslein, V., Huf, J., Hermann, R., and Bonhoeffer, F., Deoxyribonucleic acid replication *in vitro*, *J. Mol. Biol.*, 63, 183, 1972.
48. Kornberg, T. B., Lockwood, A., and Worcel, A., Replication of the *Escherichia coli* chromosome with a soluble enzyme system, *Proc. Natl. Acad. Sci. U.S.A.*, 71, 3189, 1974.
49. Modrich, P. and Richardson, C. C., Bacteriophage of *Escherichia coli* required for bacteriophage T7 DNA polymerase activity, *J. Biol. Chem.*, 250, 5508, 1975.
50. Mark, D. F. and Richardson, C. C., *Escherichia coli* thioredoxin: a subunit of bacteriophage T7 NA polymerase, *Proc. Natl. Acad. Sci. U.S.A.*, 73, 780, 1976.
51. Scherzinger, E., Lanka, E., Morelli, G., Seiffert, D., and Yuki, A., Bacteriophage T7-induced DNA-priming protein. A novel enzyme involved in DNA replication, *Eur. J. Biochem.*, 72, 543, 1977.
52. Alberts, B., Morris, C. F., Mace, D., Sinha, N., Bittner, M., and Moran, L., in *DNA synthesis and its regulation*, Goulian, M. and Hanawalt, P., Eds., W. A. Benjamin, Menlo Park, Calif., 1975, 241.
53. Godson, N., unpublished work.
54. Scott, J. and Kornberg, A., unpublished work.