Reprinted from:
Nucleic acid-protein interactions – Nucleic acid synthesis in viral infection
D.W. Ribbons, J.F. Woessner and J. Schultz, Editors
North-Holland Publishing Company – Amsterdam · London

ON THE STRUCTURE OF THE REPLICATION APPARATUS

BRUCE ALBERTS
Dept. of Biochemical Sciences
Princeton University, Princeton, N.J. 08540

Abstract: T4 bacteriophage gene 32 codes for a DNA-binding protein which is essential for DNA replication. The properties of "32-protein", together with the genetics of T4 DNA replication, suggest that the DNA, T4 DNA polymerase, 32-protein and at least 2 to 4 additional proteins combine to produce a convoluted structure through which unwound template DNA strands pass as replication proceeds.

INTRODUCTION

Although the first DNA polymerase was discovered over 15 years ago (1), and a large amount of outstanding work has followed, the detailed enzymology of DNA replication remains one of the major unsolved problems in molecular biology today. One reason why progress in this field has been limited is that only within the last year has an *in vitro* system which replicates double-stranded DNA at *in vivo rates* been developed; even in this crude system, replication continues for only about one minute at 37° (2). This appears to reflect a basic instability of the replication apparatus which is not shared by either the ribosome (and its associated factors) or by RNA polymerase, since purified *in vitro* systems for both protein synthesis and RNA transcription which operate at close to *in vivo* rates can be obtained with relative ease.

Recently, a new type of DNA-binding protein has been isolated which seems to be the main structural protein of the replication apparatus in the T4 bacteriophage system (3, 4, 5). This protein is the product of T4 gene 32. The properties and mode of action of "32-protein", together with the genetics of T4 DNA replication, appear to allow the general nature of the T4 replication apparatus to be delineated. The model which results helps to explain why the system that replicates DNA is so unstable *in vitro* and, if correct, may point the way to an eventual solution of the replication problem.

PROPERTIES OF PURIFIED 32-PROTEIN AND ITS BINDING TO DNA

Through the use of appropriate mutant bacteriophages, one of the major DNA-binding proteins synthesized after T4 bacteriophage infection of E. coli can be identified as the product of T4 gene 32 (3, 4). Elution from a single-stranded DNA-cellulose column with 2M NaCl followed by DEAE-cellulose chromatography yields 32-protein which is electrophoretically homogeneous (5). As detailed in previous publications (3, 4, 5), the native protein consists of a single polypeptide chain, with a molecular weight of about 35,000 daltons. The 32-protein carries a net negative charge at pH 7, yet shows tight, salt-sensitive binding to single-stranded DNA.

The stoichiometry of the complex which 32-protein forms with single-stranded DNA at low salt concentrations has been examined by sucrose gradient sedimentation of a fixed quantity of ^{3}H-labelled 32-protein in the presence of varying amounts of the circular, single-stranded DNA from bacteriophage fd. The results are shown in Fig. 1. At lower concentrations of DNA, two distinct peaks of radioactive protein are seen; one sediments rapidly with the DNA, the second at the slow rate characteristic of the free protein. The free protein peak is absent above a weight ratio of DNA to protein of 1 : 12. The complex therefore contains about one protein molecule of 35,000 daltons for every 10 single-stranded DNA nucleotides.

Although 32-protein binds tightly to all single-stranded DNA's tested, including the synthetic polynucleotides poly dA, poly dI, and poly dT, no binding of the purified protein to double-stranded DNA or to R17 RNA was detected by sucrose gradient sedimentation at 4°.

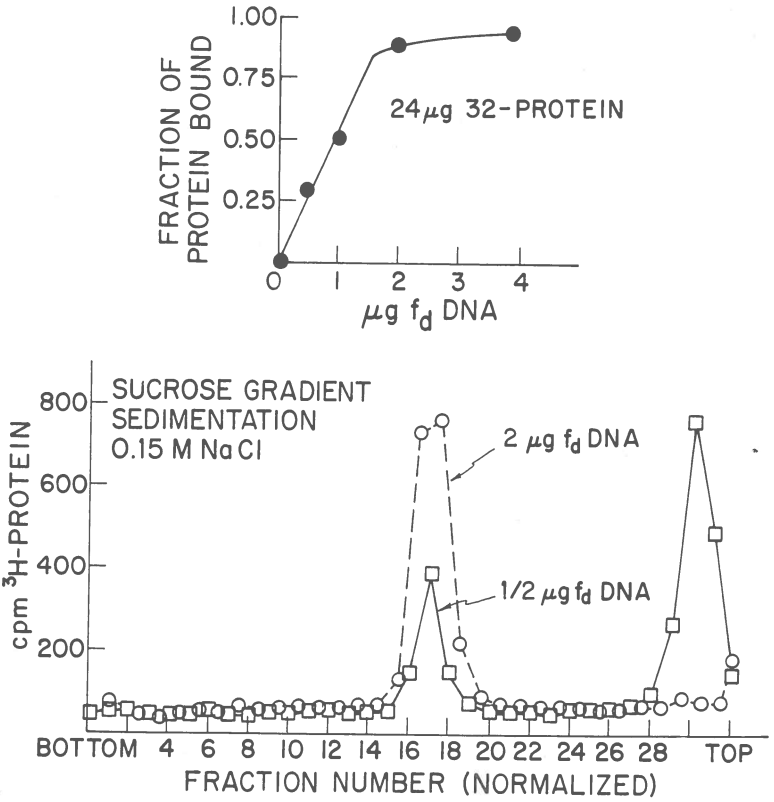
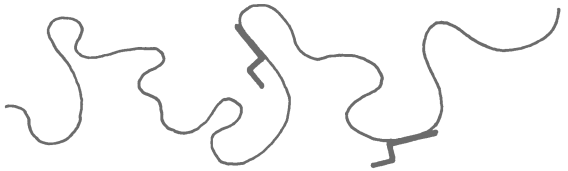


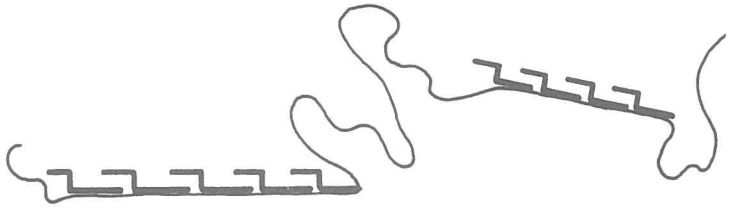
Fig. 1. Stoichiometry of 32-protein interaction with single-stranded DNA. The indicated quantity of purified fd DNA was mixed with 24 μg of ^{3}H-32-protein in 0.2 ml of 0.02 M Tris-HCl, pH 8.1, 0.5 mM Na_3EDTA, 0.15 M NaCl, 100 μg/ml bovine serum albumin (BSA), 10% glycerol, 1 mM β-mercaptoethanol at 4°. After 20 min., the mixture was layered at 4° onto a 5 ml, 5 - 30% sucrose gradient containing the same buffer. Following centrifugation, fractions were collected and monitored for radioactivity by standard techniques. Recoveries of ^{3}H-protein added averaged about 75%. The stoichiometry of the complex appears to be unchanged with 0.01 M $MgCl_2$ added, or with single-stranded T4 DNA substituted for fd DNA under the above conditions.

The binding of 32-protein to single-stranded DNA is highly cooperative. As illustrated by the model in Fig. 2, this has been explained by proposing that protein monomers line up in close juxtaposition along a DNA strand. Due to favorable protein-protein interactions, an incoming 32-protein monomer strongly prefers to bind to the 10 nucleotides adjacent to a previously bound molecule of 32-protein rather than to an isolated stretch of 10 nucleotides. At least three different types of observations support the validity of this model (for details see Ref. 5):

1) The 32-protein monomers self-aggregate in the absence of DNA at a concentration of 0.5 mg/ml or higher (see, for example, Fig. 6, below), suggesting that the cooperative binding to DNA is the result of direct stabilizing interactions between adjacent protein molecules.
2) As predicted (Fig. 2), large clusters of bound 32-protein molecules are observed to be interspersed with long stretches of bare DNA under conditions of DNA excess.
3) As judged either by translational frictional coefficients or by electron microscopy, the complexing of 32-protein with a single-stranded DNA molecule greatly extends its otherwise highly-folded conformation.



WEAK BINDING AT LOW 32-PROTEIN CONCENTRATION



STRONG BINDING AT HIGH 32-PROTEIN CONCENTRATION

Fig. 2. Model for the cooperative binding of 32-protein to single-stranded DNA.

DENATURATION AND RENATURATION OF DNA WITH 32-PROTEIN

Histones and polyamines bind to double-helical DNA more tightly than to single strands and thereby raise the temperature for DNA denaturation. Conversely, the strong selective affinity of 32-protein for single-stranded DNA drastically lowers the thermal denaturation temperature of double-stranded DNA. At 37°, with concentrations of 32-protein near the physiological level (200 μg/ml), poly dAT is rapidly denatured, while T4 DNA retains its native state due to the added stability conferred by G-C base-pairs (5). Nevertheless, the cooperative binding of 32-protein can be shown by electron microscopy to generate transient regions of local denaturation in the T4 double helix under similar conditions (H. Delius and B.A., ms in preparation).

At elevated ionic strengths, complementary strands of denatured DNA can pair rapidly with each other in vitro to reform a double helix. This renaturation process is optimal at about 65°, and sharply decreases in rate at lower temperatures (6). This decline in rate is attributed to the tight intrastrand folding of denatured DNA, which begins when the temperature is lowered more than 25° below the Tm of the double helix (7). These folds of hair-pin base-paired loops make the DNA bases relatively inaccessible, and thereby prevent complementary single strands from finding satisfactory interstrand pairings. This is purely a kinetic effect, since the equilibrium stability of the double helix relative to single strands is greater at the lower temperature.

In addition to destabilizing the double helix, the binding of 32-protein to DNA single strands forces them to adopt a much more open conformation, and thereby increases their rate of renaturation under physiological conditions by as much as 1000-fold (5). A plot of the effect of 32-protein on renaturation rates is given in Fig. 3, in order to emphasize that a decrease in helix stability can accelerate the renaturation, as well as the denaturation, of DNA.

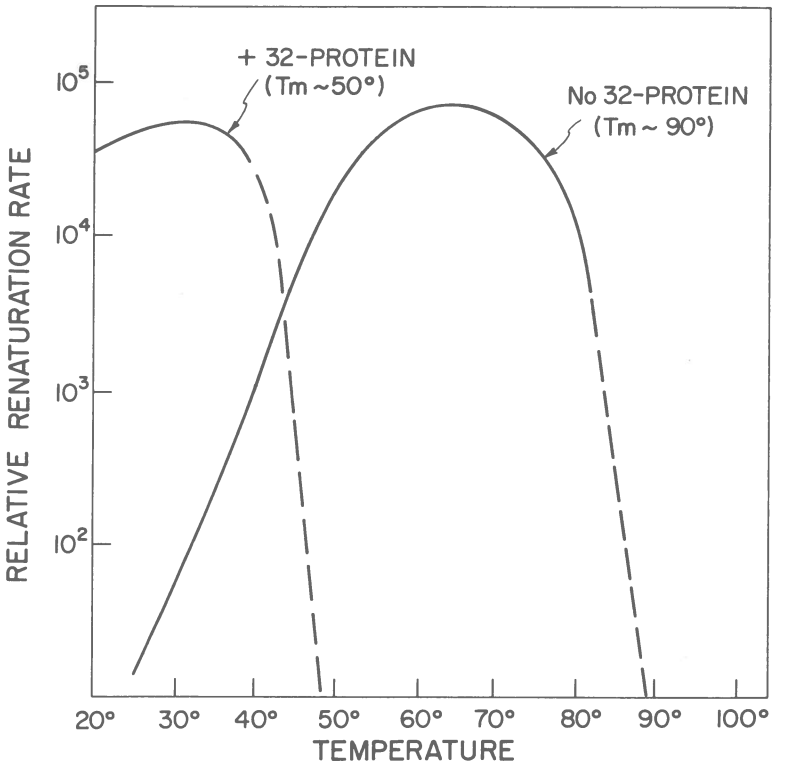


Fig. 3. Approximate effect of 32-protein on the temperature dependence of DNA renaturation rates. The "no 32-protein" curve has been estimated from the results of Studier (7).

THE PROTEINS ESSENTIAL FOR T4 DNA REPLICATION

As first shown by Tomizawa and coworkers (8), 32-protein is required for an early step in genetic recombination in the T4 bacteriophage system. This can be explained by the dual action of 32-protein in transiently opening up local regions of native DNA while at the same time facilitating helix formation between matching single strands, both of which may be essential for the recombination process (see Ref. 5).

Of concern in the present context is that 32-protein is also one of several gene products required for T4 DNA replication (9, 10). That the role of this protein in replication is a direct one, independent of its role in genetic recombination, is clearly demonstrated by studies with T4 mutant ts P7, which makes a temperature-sensitive 32-protein. As first shown by Riva, Cascino and Geiduschek (11), cells infected with this mutant bacteriophage synthesize T4 DNA at normal rates at 25°, but stop replication completely within 1 minute after a shift to 42° (Fig. 4, M. Curtis, Senior Thesis, Princeton University, 1970).

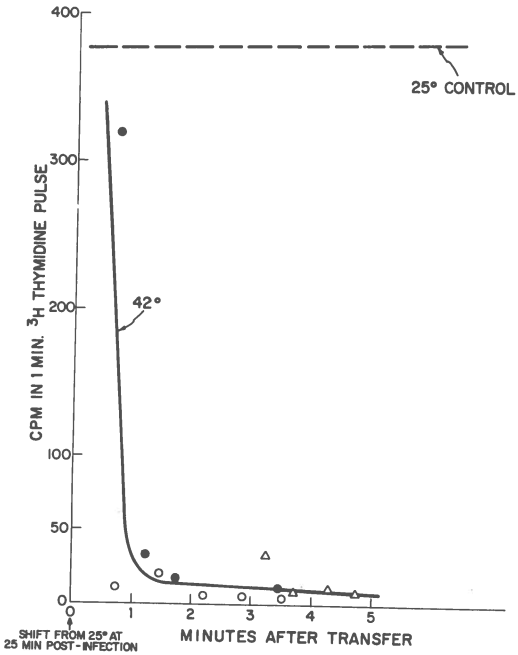


Fig. 4. Dependence of the rate of T4 DNA synthesis on the continued function of gene 32-protein. E. coli B3 at 3 x 10^8 cells/ml in M_9 minimal medium containing 0.3% casein hydrolysate and 20 μg/ml thymidine were infected at 25° with a multiplicity of 5 T4 ts P7 phage per bacterium. After 25 min., the infected cells were diluted fivefold into prewarmed media (without additional thymidine). At various times, thymidine-^{3}H was added for 1 min. to an aliquot. At the indicated time, incorporation was stopped with an equal volume of cold 10% trichloroacetic acid, and the precipitate collected and counted.

Similar experiments indicate a direct role in T4 DNA replication for at least three other gene products: genes 43, 41, and 45 (11). Amber mutants in any of these genes or in genes 62 and 44 synthesize very little or no DNA, even though all deoxyribonucleotide triphosphate precursors are present (9, 12). Therefore, in addition to 32-protein, it appears that at least five T4-induced proteins play a central role in building an active T4 replication apparatus. The function of only one of these proteins is known: gene 43-protein is the T4 DNA polymerase (13). This polymerase has an absolute requirement for a single-stranded DNA template, and needs a pre-existing 3' hydroxyl-terminated primer chain onto which it polymerizes 5' deoxyribonucleotide triphosphate precursors in the 5' to 3' direction (13, 14).

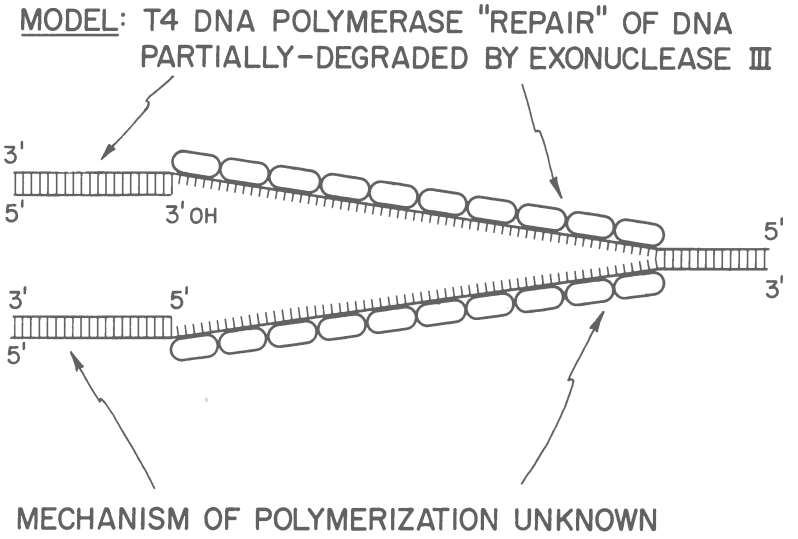


Fig. 5. Schematic view of a simple replication fork containing 32-protein.

THE INTERACTION OF T4 DNA POLYMERASE AND 32-PROTEIN

How might the products of genes 43 and 32 function together in the replication apparatus? The requirement of T4 DNA polymerase for a single-stranded DNA template and the ability of 32-protein to unwind the double helix suggest that 32-protein might precede the polymerase, exposing the template and aligning the bases in a conformation optimal for polymerase action.

In this case, the "repair" of DNA, partially degraded by exonuclease-III, by T4 DNA polymerase in the presence of 32-protein should provide a system for study of the synthetic events on at least one side of the replication fork (Fig. 5). A detailed study of this system (J. Huberman, A. Kornberg, and B.A., ms submitted for publication) has revealed that:

1) 32-protein stimulates the *in vitro* rate of polymerization of T4 DNA polymerase on an exonuclease-III-degraded DNA template 5 - 10 fold. The rate of stimulated synthesis is about 10% of that measured for growing point movement *in vivo*, and is greatest with a level of 32-protein sufficient to bind most or all of the template strand present.
2) The 32-protein from a temperature-sensitive mutant in gene 32 (*ts* P7) is also temperature-sensitive in the stimulation of T4 DNA polymerase *in vitro*, suggesting that this stimulation is related to the intracellular function of 32-protein in DNA replication.
3) The activity of DNA polymerase from *E. coli* on an exonuclease-III-degraded DNA template is not significantly affected by 32-protein. In addition, in the absence of DNA, a weak complex is formed between 32-protein and T4 DNA polymerase, while no such complex is formed with *E. coli* DNA polymerase (Fig. 6). These observations could be explained if some specific fit between polymerase and template strands complexed with 32-protein is essential for the stimulation of DNA synthesis observed.

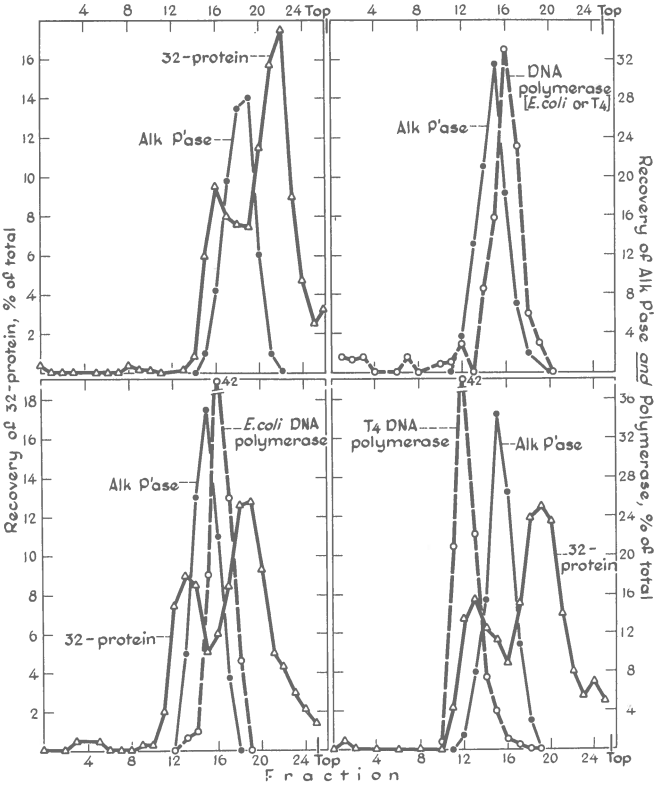


Fig. 6. Sucrose gradient sedimentation of DNA polymerases in the presence of a high concentration of 32-protein. Sedimentation was through 5 ml, 5 - 30% sucrose gradients containing 10 mM MOPS, pH 7.4, 50 mM KCl, 0.5 mM Na_3 EDTA, 10 mM $MgCl_2$, 1 mM β-mercaptoethanol, 100 μg/ml bovine serum albumin, and 10% glycerol. The following proteins, in 0.1 ml of sucrose-gradient buffer containing 10 μg of E. coli alkaline phosphatase, were layered onto the respective gradients: a) 200 μg of ^{3}H-32-protein; b) either 1.5 μg of T4 DNA polymerase, or 1 μg of E. coli DNA polymerase; c) 200 μg of ^{3}H-32-protein plus 1 μg of E. coli DNA polymerase; d) 200 μg of ^{3}H-32-protein plus 1.5 μg of T4 polymerase. If, in an experiment similar to d), only 30 μg of 32-protein is used, the T4 DNA polymerase peak sediments behind the alkaline phosphatase marker, as in b). Thus, the complex of 32-protein and T4 polymerase is a loose one. Note that, at the concentration used, an uniquely stable aggregate of 32-protein appears. This species is believed to arise when end to end aggregates form which are long enough to circularize.

IMPLICATIONS FOR THE STRUCTURE OF THE REPLICATION APPARATUS

The notion that 32-protein and T4 DNA polymerase function together in DNA replication does not at first sight drastically alter the conventional view of the replication process. However, gene dosage experiments carried out by Snustad have revealed that gene 32 is unique among the T4 genes affecting DNA metabolism in that the amount of 32-protein synthesized directly limits the number of progeny phage produced (15). More recently, it has been shown that the total amount of DNA made in these cells is roughly proportional to the amount of 32-protein synthesized, at least over the six-fold range explored (N. Sinha and D.P. Snustad, ms submitted for publication). Yet about 10,000 molecules of 32-protein are made per cell in a normal infection, and very few, if any, are destroyed (4). The fact that 32-protein is needed in such large amounts for DNA replication suggests that it plays a major structural role in this process.

In the gene dosage experiments, the level of 32-protein could be controlling the rate at which replication forks move. At the other extreme, 32-protein could be determining the number of replication forks made but not their rate of movement. Since the intracellular level of 32-protein increases continuously throughout infection (4), analysis of how replication rates change during a normal infectious cycle can be used to evaluate these two extreme possibilities.

According to Werner, during a normal T4 bacteriophage infection at 25°, DNA replication begins at about 10 minutes post-infection and continues at a rate which increases linearly until 30 minutes post-infection, when the rate stabilizes at a constant maximum (16). This increase in the rate of replication is caused by a parallel increase in the number of replication forks from zero to 60 per infected cell. The crucial point is that each replication fork moves at a constant rate which is independent of the time after infection, and therefore independent of the amount of 32-protein present (16). Moreover, if the protein synthesis inhibitor chloramphenicol is added to T4-infected cells during the period when the rate of DNA synthesis (and therefore the number of replication forks) is increasing, DNA synthesis continues, but at a rate which gradually decreases from that established just prior to the moment of chloramphenicol addition (17). The simplest interpretation of this result is that generation of each new replication fork requires at least one protein present in limiting amounts; it is therefore the rate of synthesis of this protein which determines how rapidly new replication forks are made. Among known genes, this protein could only be the product of gene 32, since the products of the other T4 genes which affect

DNA replication are clearly synthesized in large excess (15).

Thus, we arrive at the following general model for the structure of the T4 replication apparatus:

> A replication unit capable of synthesizing DNA contains a fixed number, "X", of 32-protein molecules. (As 10,000 molecules of 32-protein are enough to make at least 60 replication forks, X cannot be more than 170.) During the course of infection, as X molecules of 32-protein become available, they are incorporated together into a replication unit ("replication apparatus") which travels along its DNA template at a rate independent of the concentration of free 32-protein. In principle, if there are only X-1 molecules of 32-protein in a cell, DNA replication should either not begin, or it should be aberrant.

A replicating apparatus which demands exactly X molecules of 32-protein should of course have a unique three-dimensional structure and also require a fixed number of every other protein which functions in it (i.e., the products of T4 genes 41, 43, 45, and perhaps 44 and 62).

SPECIFIC MODELS

With an in vivo polymerization rate of about 1500 nucleotides sec^{-1} at 37° (2), 32-protein must be added to each template strand at a rate of about 150 molecules sec^{-1} in order to keep ahead of the polymerase. As the association constant for 32-protein invasion of a double-helical region is not yet known, the possibility that this is accomplished by random collision with a free intracellular pool of 32-protein molecules cannot yet be ruled out. Nevertheless, it is intriguing to speculate that the X molecules of 32-protein in each replication unit actually remain within that unit as replication proceeds. Regardless of whether DNA replication proceeds continuously on both strands (Fig. 7A) or discontinuously on the second strand (Fig. 7B, Ref. 18), a requirement for circulation of 32-protein within each replicating structure not only puts interesting constraints on the types of replicating structures which are possible (see heading to Fig. 7), but provides a simple rationale for the proposal that replication only proceeds when one particular conformation of the replication complex is attained.

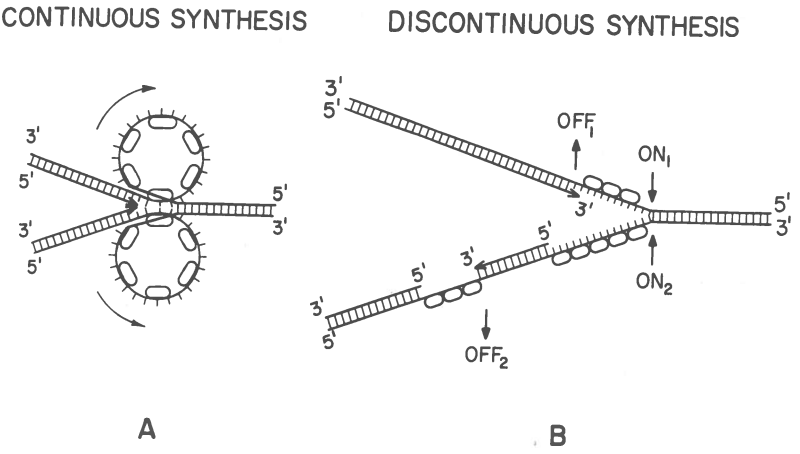


Fig. 7. Schematic models for circulation of 32-protein within the replication apparatus.
A. Continuous synthesis of DNA on both template strands. This model would require a second type of DNA polymerase, permitting polymerization of triphosphates in the 3' to 5' direction on one strand.
B. Discontinuous synthesis of DNA on one template strand (18). Only one type of polymerase is needed. However, the circulation of 32-protein would be more complicated, most likely joining OFF_1 with ON_1, and OFF_2 with ON_2 in a convoluted structure. Note that in both models the total length of single-stranded DNA must be kept constant at all times.

With this viewpoint, it also becomes easier to understand why polymerase and 32-protein are not sufficient by themselves to replicate DNA in the T4 system. The additional gene products needed, thus far identified only by genetic experiments, might be expected to interact with 32-protein, T4 DNA polymerase and/or each other in building the correct spatial relationships between the different regions of the replicating unit. They would also be required for stabilization of the complementary template strands in a complex with 32-protein, since in the absence of other factors, the double helix should be so highly favored under physiological conditions as to allow a structure such as that illustrated in Fig. 5 only a very transient existence.

In concluding, it is instructive to compare this view of the

replication apparatus with the ribosome, where an unique three-dimensional structure is also built with proteins around a framework of nucleic acid. As with the ribosome (19), some of the proteins of the replication apparatus are not expected to be bound directly to the nucleic acid component. A crucial difference is, of course, that the ribosome is a stable structure which contains its own RNA while the replication apparatus must be free to slide rapidly along the DNA threaded through it, pushed by the cleavage of triphosphates behind. In such a structure, rather weak protein-protein interactions, such as that observed between T4 DNA polymerase and 32-protein (Fig. 6), might be expected to prevail in place of the much stronger forces which stabilize the ribosome. This could explain why all attempts to isolate a replication apparatus thus far have failed.

OTHER BIOLOGICAL SYSTEMS

If the essential features of T4 DNA replication are shared by other organisms, a protein with properties similar to 32-protein should be found in all replicating systems. In theory, the detection of this type of protein is made difficult in uninfected bacteria by the fact that there are only a few replicating forks in each cell. Recently, we (and others) have found that the gene 5-protein of fd bacteriophage is of the 32-protein type (B.A. and L. Frey, ms in preparation; J. Oey and R. Knippers, personal communication): it binds tightly and cooperatively to single-stranded (but not to double-stranded) DNA; it forms a complex in which the DNA strand is unfolded and fully hyperchromic; and it rapidly denatures poly dAT at 25°. Like 32-protein, the gene 5-protein is synthesized in large amounts (20), and it is required for synthesis of the single-stranded, progeny DNA strands (21).

REFERENCES

1. A. Kornberg, I.R. Lehman and E.S. Simms. Fed. Proc. 15 (1956) 291.

2. D.W. Smith, H.E. Schaller and F.J. Bonhoeffer. Nature 226 (1970) 711.

3. B.M. Alberts, F.J. Amodio, M. Jenkins, E.D. Gutmann and F.L. Ferris. Cold Spring Harbor Symp. Quant. Biol. 33 (1968) 289.

4. B.M. Alberts. Fed. Proc. 29 (1970) 1154.

5. B.M. Alberts and L. Frey. Nature 227 (1970) 1313.

6. J. Marmur and P. Doty. J. Mol. Biol. 5 (1962) 109.

7. F.W. Studier. J. Mol. Biol. 41 (1969) 199.

8. J. Tomizawa, N. Anraku and Y. Iwama. J. Mol. Biol. 21 (1966) 247.

9. R.H. Epstein, A. Bolle, C.M. Steinberg, E. Kellenberger, E. Boy de la Tour, R. Chevalley, R.S. Edgar, M. Susman, G.H. Denhardt and A. Lielausis. Cold Spring Harbor Symp. Quant. Biol. 28 (1963) 375.

10. A. Kozinski and Z.Z. Felgenhauer, J. Virol. 1 (1967) 1193.

11. S. Riva, A. Cascino and E.P. Geiduschek. J. Mol. Biol. 54 (1970) 85.

12. H.R. Warner and M.D. Hobbs. Virology 33 (1967) 376.

13. H.V. Aposhian and A. Kornberg. J. Biol. Chem. 237 (1962) 519.

14. M. Goulian, Z.J. Lucas and A. Kornberg. J. Biol. Chem. 243 (1968) 627.

15. D.P. Snustad. Virology 35 (1968) 550.

16. R. Werner. Cold Spring Harbor Symp. Quant. Biol. 33 (1968) 501.

17. P. Amati. Science 168 (1970) 1226.

18. T. Okazaki and R. Okazaki. Proc. Nat. Acad. Sci., U.S. 64 (1969) 1242.

19. M. Nomura. Bacteriological Reviews 34 (1970) 228.

20. T.J. Henry and D. Pratt. Proc. Nat. Acad. Sci., U.S. 62 (1969) 800.

21. D.A. Marvin and B. Hohn. Bacteriological Reviews 33 (1969) 172.

The experimental work described in this article was supported by grants from the U.S. National Institutes of Health (GM-14927) and The American Cancer Society (E-510).

DISCUSSION

V.G. ALLFREY: Many of the proteins which associate with the proteins of mammalian chromosomes are phosphoproteins. Is either the 32-protein or the f_d protein phosphorylated or reversibly de-phosphorylated?

B.M. ALBERTS: A graduate student, Larry Moran, attempted at one time to label T_4 DNA-binding proteins with ^{32}P. There was no indication that 32-protein contains phosphorus in his experiments.