Minireview

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 263, No. 1, Issue of January 5, pp. 1-4, 1988
Printed in U.S.A.

DNA Replication

Arthur Kornberg

From the Department of Biochemistry, Stanford University School of Medicine, Stanford, California 94305

Whither DNA Replication

DNA replication is crucial to life: in the duplication of the genome, in its rearrangements, and in its repair when damaged. The maturity of research on replication is attested to by monographs (1, 2), reviews (3), and symposia (4). Yet, many interesting and accessible questions remain unanswered. What we know and would like to know will be discussed under: how a DNA chain is started and extended, and how chromosomes are initiated, terminated, and segregated.

How Is a DNA Chain Started?

Because DNA polymerases can only extend a preexisting chain, a new chain, almost always, is started by a short RNA transcript, known as a primer. There are two exceptions to RNA priming. (i) A 3′-hydroxyl (primer) end is created by a specific endonuclease, as in rolling-circle replication of some phages and plasmids, or by a folding back of the 3′-end of the linear, single-stranded DNA of parvoviruses. (ii) A virus-encoded protein serves as a primer by pairing a single deoxynucleotide with the end of a template strand, as in the replication of the linear chromosomes of adenoviruses and *Bacillus subtilis* phage ϕ29.

The specialized RNA polymerases that synthesize the mini-transcript serving as a primer are called *primases*. Their variety is fascinating. Some read sequences of a few bases; others instead synthesize a primer of an exact length by counting a multiple of bases (*e.g.* 10), but are indifferent to their sequence. Some accept deoxynucleotides (as well as ribonucleotides) for elongation; others do not. Some are tightly bound to a DNA polymerase (eukaryotic), others to a helicase (phage T4), and some contain the primase and helicase activities within a single polypeptide (phage T7). *Escherichia coli* primase associates with a complex containing dnaB, dnaC, dnaT (protein i), and three other proteins (about 20 polypeptides in all) to form a mobile, multifunctional primosome.

In no instance is it clear how a primase, by analogy with RNA polymerases, recognizes its "promoter" or terminates transcription. While primases, like all polymerases, extend chains in the 5′ → 3′ direction, they must, in the discontinuous synthesis of the lagging strand, maintain their position at a replication fork (see Fig. 2 below) by moving in the opposite direction on the template. This translocation must somehow be integrated with the actions of polymerase and helicase and for priming be responsive to signals in the template. For several viral, microbial, and animal primases, the stage is now set, the genetic script is in hand, and the enzyme actors are in full view. They await the interested biochemist to attend and direct their performance.

How Is a DNA Chain Extended?

The key features of DNA polymerase are fidelity and catalytic efficiency. *Fidelity* is achieved by a polymerase in two stages. Initially, a deoxynucleotide is added to the 3′-hydroxyl end of the primer DNA chain to pair with the base next in line on the template chain; the error rate is about 10^{-4} per base replicated. Next, a 3′ → 5′ exonuclease proofreads the newly added nucleotide and removes those mismatched with the template, thereby lowering the error rate to about 10^{-6}. This exonuclease activity is found within a domain of the polymerase polypeptide, as in DNA polymerase I of *E. coli* (pol I), or in a separate subunit, such as the ϵ subunit of DNA polymerase III (pol III) holoenzyme (5); ϵ becomes effective in proofreading only upon attachment to the α subunit, the polymerase activity (6). Eukaryotic polymerases, until recently, appeared to lack an associated proofreading 3′ → 5′ exonuclease. Examples have now been found which possess the activity in a cryptic state (7) or as part of a higher molecular weight form of the enzyme (8).

Catalytic efficiency depends on the processivity of the polymerase and the capacity to remain associated with the template strand and move without delay to the newly generated primer end. Because dissociation of the enzyme and its reassociation with the primer terminus consumes a minute or so, compared to the millisecond required to add a nucleotide, processivity has a huge effect on the overall rate. Pol I, with a processivity of 10 to 20 nucleotides, is thus suited to filling short gaps, whereas DNA polymerase III holoenzyme, responsible for most replication and equipped to clamp itself to the template, has virtually unlimited processivity. Thus, replication of a nucleotide chain which takes the holoenzyme only 10 s would take pol I several hours.

In the *molecular anatomy of polymerases*, the only three-dimensional structure probed by x-ray diffraction is the classic pol I, and only the amino-terminal fragment of the enzyme ("large fragment") with polymerase and proofreading functions has been crystallized. At a resolution of 2.75 Å, a cleft or sleeve is seen for accommodating the primer-template, with the presumed site for proofreading the growing chain end some 30 Å away (9); crystals of a polymerase-DNA complex indicate that single-stranded DNA binds near the growing end site. Slightly beyond the extreme simplicity of a single polypeptide is the polymerase encoded by coliphage T7, a virus streamlined for replication and transcription. This 80-kDa polypeptide adopts the 12-kDa host thioredoxin, thereby achieving (by still obscure means) rapid and processive replication (10).

A more elaborate polymerase, encoded by phage T4, resembles the pol I of its *E. coli* host (3, 11) but, by enlisting three auxiliary subunits, it attains processivity in chain extension and begins to approach the complexity of the host pol III holoenzyme. The latter, containing some 20 polypeptides, 10 of them distinctive from one another, is almost elaborate enough to be dubbed a "-some." Unlike RNA polymerases and ribosomes, the "replisome" complex comes apart readily. The effort and frustrations of isolating the friable pol III holoenzyme (with all its subunits in place) has been compensated by finally having most of the subunits in hand and the opportunity to examine them individually to see how they work and fit together. Several of the subunits are essential to processivity. Among them, the β subunit (37 kDa) is held loosely in the holoenzyme, with a K_D of 1 nM, and dissociates during the cycling from one template to another. The holoenzyme appears to be organized as an *asymmetric dimer*: a pair of core subassemblies, each with a potential for polymerase

action (Fig. 1), has an asymmetric distribution of auxiliary subunits which may endow each core with different properties, one suited to the continuous synthesis of one strand and the other to the discontinuous synthesis of the other strand (see Fig. 2).

The complexity of eukaryotic polymerases, with stringent requirements for fidelity, may match or exceed that of *E. coli*. Bedeviled by proteolytic degradation during isolation, it has become clear to investigators that, in isolating a eukaryotic polymerase, "bigger is better." Polymerase α in animal cells, acknowledged for a role in chromosome replication, is now seen as composed of four subunits: a polymerase (180 kDa) with a potent $3' \rightarrow 5'$ exonuclease (proofreading) activity which is masked in the presence of a 70-kDa polypeptide, and two polypeptides (50 and/or 60 kDa) which provide the primase activity (12). Polymerase δ, similar in size and organization to α but with a manifest exonuclease function, may well be a companion replicative enzyme. Of striking interest is a loosely associated 37-kDa auxiliary protein which, like the β subunit of pol III holoenzyme, endows polymerase δ with processivity (8). The 37-kDa protein has been discovered in two other guises (13): one as a prominent, cell-cycle-regulated protein, the proliferating cell nuclear antigen (PCNA, also called cyclin), and the other as one of three host components required for the replication of the circular tumor virus, SV40. These properties firmly associate polymerase δ with replication in the cell cycle, but no relationship to polymerase α has yet been demonstrated. If these polymerases prove to be truly separate, they may team with each other to create a replisomal complex as proposed for the asymmetric, dimeric pol III holoenzyme (Fig. 1). The existence of additional polymerases, β for DNA repair and γ for replication of mitochondrial DNA, is known, but how they operate in these processes is not.

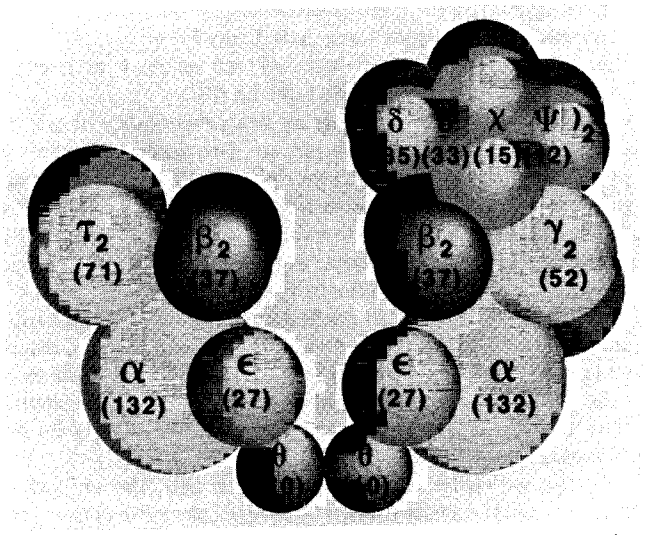

FIG. 1. **Hypothetical scheme for composition of polypeptides and their asymmetric, dimeric organization in DNA polymerase III holoenzyme of *E. coli*.**

The *dynamics of polymerase* action are largely unexplored. The kinetics and affinities of binding of polymerases to single strands, to duplexes, and to termini, and their movements on DNA are known in only a few instances. As an example, pol III holoenzyme, activated by ATP, forms a stable initiation complex with a primer-template; neither the affinities nor the "on" and "off" rates are known. How DNA-binding proteins move on DNA to recognize origins, primers, promoters, recombination and restriction sites, lesions, etc. is a fascinating question. When pol III holoenzyme reaches the end of available template because a primer strand, annealed to the template, lies in its path of replication, the polymerase can slide over the duplex stretch without expending energy and resume replication at the new primer terminus within a fraction of a second (14).

Advancing the replication fork (Fig. 2) depends on opening of the duplex by the action of helicases (such as rep protein

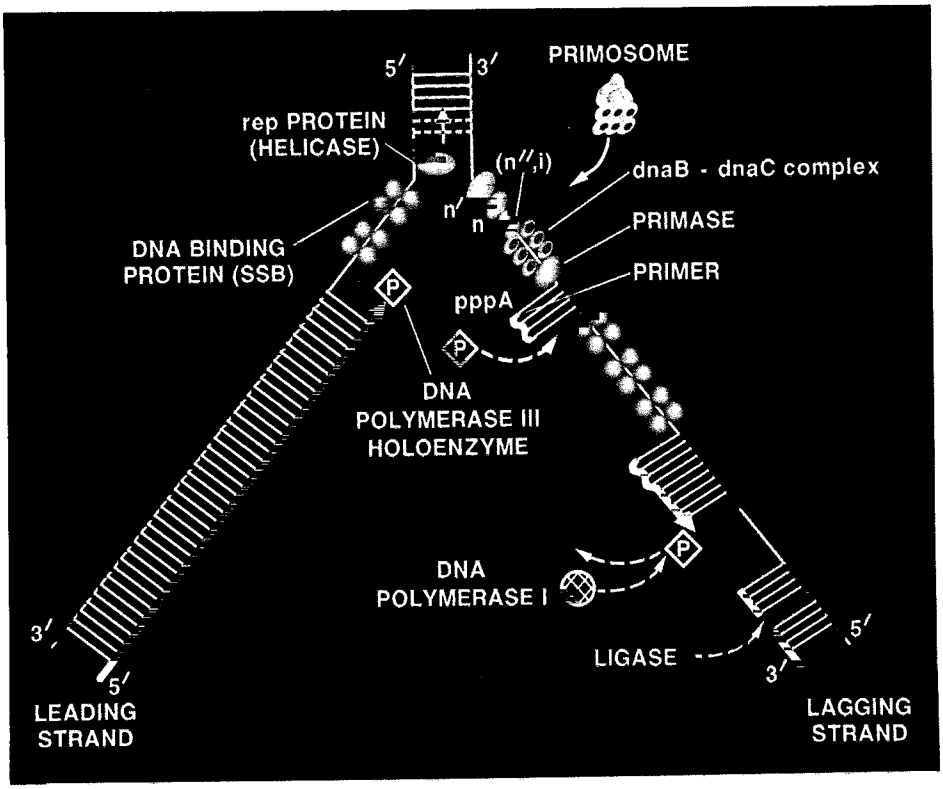

FIG. 2. **Scheme for enzymes operating at a replication fork of *E. coli* as proposed in Ref. 2, p. S122.**

of *E. coli* acting on one strand possibly in concert with the primosome acting on the other) and relief of the consequent positive supercoiling by the swivelling action of a topoisomerase (*e.g.* gyrase). Single-strand binding protein promptly covers the bared single strands, protecting them against nuclease action, preventing them from reannealing, and configuring them to serve as templates. Replication proceeds by continuous synthesis of the "leading" strand and discontinuous synthesis of the (other) "lagging" strand.

Essentially *concurrent replication* of both strands, rather than the jerky sequence of synthesis of one strand and then the other, might be achieved were priming of nascent fragments of the lagging strand integrated with continuous synthesis of the leading strand. Concurrent replication would also require a more complex and facile holoenzyme (a replisome) possessing primase and twin active sites for polymerization. Accumulating evidence for such a structure includes: (i) twin polymerase subassemblies in the pol III holoenzyme,[1] (ii) replication of both strands of a duplex by phage T4 polymerase with kinetics consistent with action by one polymerase molecule, and (iii) complexing of primase by some polymerases (11). In this hypothetical scheme for concurrent replication (Fig. 3), looping of the lagging strand template by 180° (perhaps halfway around the polymerase) endows it with the same orientation as the leading strand at the fork. A primer generated by primase is extended by polymerase as the lagging strand template is drawn past it. When synthesis reaches the 5′-end of the previous nascent fragment, the lagging strand template is released and unlooped. Helicase action and continuous synthesis of the leading strand periodically expose lengths of template for priming of nascent fragments.

The agenda for polymerase studies, beyond the items mentioned, includes: regulation of the synthesis and assembly of multisubunit polymerases; disposition of special subassemblies of these polymerases for replication, repair, and recombination; mechanisms of fidelity and processivity; modifications of polymerases for special needs (*e.g.* mutagenesis) (15); and unique polymerase forms for action on viral, mitochondrial, and other episomal elements.

How Are Chromosomes Initiated?

A major question exists: How might an increase in *E. coli* cell mass trigger the initiation of replication that commits the cell to start a new cycle? We need to know the biochemistry of the replication switch, which in *E. coli* regulates the cell cycle and in eukaryotes responds to signals that turn the embryonic cell to adult quiescence, or the quiescent cell to proliferation. Progress with *E. coli* has come largely from cloning the unique 245-base pair chromosomal origin (*oriC*) in plasmids and the isolation and analysis of the multiprotein system that replicates these plasmids (16). The proteins fall into three groups: (i) *initiation proteins* (*e.g.* dnaA, dnaB, dnaC, and HU proteins, and RNA polymerase) that recognize supercoiled *oriC*, alter its structural conformation, and lead to its further opening by dnaB helicase action; (ii) *specificity proteins* (*e.g.* topoisomerase I, ribonuclease H, and protein HU) that suppress potential origins elsewhere on the chromosome, and (iii) *replication proteins* (*e.g.* primase, polymerase, gyrase, helicase, and single-stranded binding protein) that prime and elongate chains on the opened plasmid and propel the two forks in bidirectional replication.

Initiation of the chromosome in *E. coli* depends primarily on dnaA protein. After binding its four recognition "boxes" (9-mers) in *oriC*, the protein can then perform a specific duplex opening reaction in an adjacent A-T-rich region (identified by three tandem 13-mers).[2] Having done so, dnaA protein, with the aid of dnaC protein, can admit and position dnaB protein in the open complex, which by helicase action can create a bubble of a few hundred residues, a locus for the priming of replication. Organization of *oriC* in *B. subtilis* and in each of many plasmids and phages is remarkably like that in *E. coli*, containing iterated sequences for binding of the initiator protein and neighboring A-T-rich repeats that can provide the site for opening the duplex. Initiation is profoundly influenced by the level of dnaA protein, by proteins that affect superhelical density (*e.g.* topoisomerase I, gyrase, and HU protein), by transcription (see below), and by other factors (*e.g.* ATP and cardiolipin).

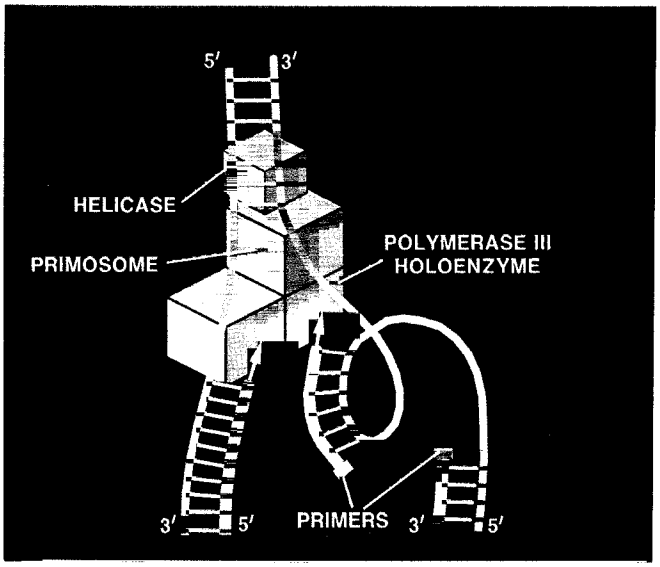

FIG. 3. **Hypothetical scheme for concurrent replication of leading and lagging strands (Fig. 2) by an asymmetric, dimeric polymerase (Fig. 1) associated with a primosome and a helicase in a replisome as proposed in Ref. 2, p. S125.**

Initiation of a eukaryotic genome, if a papovavirus can be used as an example, resembles that of *E. coli*. The T antigen of SV40, binding specific sequences in the origin, applies its helicase action (17, 18) to open the duplex preparatory to priming and replication by cellular enzymes (19). Factors influencing eukaryotic initiation may also be like those mentioned for *E. coli*, such as the histone binding of DNA, transcription, and the abundance of initiator proteins. Cell-cycle-dependent synthesis of key proteins and foci of transcription might determine when and where replication starts are made.

Transcriptional activation, in the initiation of chromosome replication, is likely as important as the RNA synthesis that primes the start of a DNA chain. In a reconstituted *E. coli* enzyme system, a transcript hybridized to a region near the origin can activate replication of a plasmid which may be inert for one of several reasons: reduced superhelicity, binding by HU protein, or a suboptimal temperature (20).[3] In eukaryotic systems, binding of a protein to a promoter sequence near the origin of replication appears to activate both transcription and initiation of adenoviral replication (21); similar factors may influence the replication of SV40 and of bovine papilloma virus as well. Whether these transcription factors directly activate replication by transcription has yet to be determined.

[1] H. Maki, S. Maki, and A. Kornberg, manuscript in preparation.
[2] D. Bramhill and A. Kornberg, manuscript in preparation.
[3] T. A. Baker and A. Kornberg, manuscript in preparation.

How Are Chromosomes Terminated and Segregated?

Completion of daughter chromosomes may require untangling catenated circles with topoisomerases (22); addition of telomeric sequences to linear chromosomes appears to involve a novel nucleotidyltransferase operating without a DNA template (23). However, it is uncertain whether terminator sequences exist in DNA and, if so, how proteins might respond to them. Virtually nothing is known about the molecular events that determine partition of prokaryotic chromosomes between daughter cells. Although membrane attachment is a plausible basis for segregation, the chemical evidence is scanty. In the mitotic apparatus of eukaryotic cells, spindles of microtubules and associated proteins arrange and divide the chromosomes in an elaborate sequence of reactions. For simplicity, plasmids are attractive, especially those with a low copy number which encode proteins that block cell division unless replication of the plasmid has been achieved (24). The world is waiting for an *in vitro* assay of chromosome partition so that the responsible proteins can be identified and isolated.

Frontiers in the Enzymology of Replication in Eukaryotes

Relative to the many opportunities for exploring the enzymology of replication in prokaryotes, similar pursuits of eukaryotic systems may seem daunting. The complexities of size and organization of eukaryotic genomes and the general lack of genetic analysis impose genuine difficulties. Yet the directions seem clear. Discovery and *functional* characterization of helicases, topoisomerases, and binding proteins, as well as polymerases and primases, will have a great impact. The basic mechanisms and enzymes of replication found in *E. coli* provide helpful guides, as they have for transcription and translation, but novel features are likely to be uncovered, much the same as in studies of gene expression. Technical difficulties, such as proteolytic degradation, inadequate genetic information, and limited source material are being circumvented with improved techniques (*e.g.* amplification by cloning and antibody affinity chromatography) and a focus on better sources (*e.g.* viruses, yeast, and *Drosophila*). The extraordinary scientific interest and practical importance of knowing the fine details of replication in various cells and tissues, in health and disease, promise great rewards for increased efforts in this area. The somatic mutations and genomic rearrangements responsible for antibody diversity will likely enlist the actions of nucleotidyltransferases, recombinases, and polymerases. In view of the general concern with controlling AIDS, it is remarkable that so little is known about reverse transcriptases and how they synthesize duplex DNA and prepare it for integration into a host chromosome. Not a single paper on reverse transcriptase appeared in this Journal in 1986.

REFERENCES[4]

1. Kornberg, A. (1980) *DNA Replication*, W. H. Freeman & Co., San Francisco
2. Kornberg, A. (1982) *Supplement to DNA Replication 1980*, W. H. Freeman & Co., San Francisco
3. Nossal, N. G. (1983) *Annu. Rev. Biochem.* **52,** 581–615
4. McMacken, R., and Kelly, T. J. (1987) *Replication and Recombination: UCLA Symposia on Molecular and Cellular Biology, New Series*, Vol. 47, Alan R. Liss, Inc., New York
5. Scheuermann, R. H., and Echols, H. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81,** 7747–7751
6. Maki, H., and Kornberg, A. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** in press
7. Cotterill, S. M., Reyland, M. E., Loeb, L. A., and Lehman, I. R. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** in press
8. Tan, C.-K., Castillo, C., So, A. G., and Downey, K. M. (1986) *J. Biol. Chem.* **261,** 12310–12316
9. Joyce, C. M., and Steitz, T. A. (1987) *Trends Biochem. Sci.*, in press
10. Tabor, S., Huber, H. E., and Richardson, C. C. (1987) *J. Biol. Chem.*, in press
11. Selick, H. E., Barry, J., Cha, T.-A., Munn, M., Nakanishi, M., Wong, M. L., and Alberts, B. M. (1987) *Replication and Recombination: UCLA Symposia on Molecular and Cellular Biology, New Series*, pp. 183–214, Alan R. Liss, Inc., New York
12. Kaguni, L. S., Rossignol, J.-M., Conaway, R. C., Banks, G. R., and Lehman, I. R. (1983) *J. Biol. Chem.* **258,** 9037–9039
13. Prelich, G., Kostura, M., Marshak, D. R., Mathews, M. B., and Stillman, B. (1987) *Nature* **326,** 471–475
14. O'Donnell, M. E., and Kornberg, A. (1985) *J. Biol. Chem.* **260,** 12875–12883
15. Lu, C., Scheuermann, R. H., and Echols, H. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 619–623
16. Funnell, B. E., Baker, T. A., and Kornberg, A. (1987) *J. Biol. Chem.*, in press
17. Stahl, H., Dröge, P., and Knippers, R. (1986) *EMBO J.* **5,** 1939–1944
18. Dean, F. B., Bullock, P., Murakami, Y., Wobbe, C. R., Weissbach, L., and Hurwitz, J. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 16–20
19. Li, J. J., and Kelly, T. J. (1985) *Mol. Cell. Biol.* **5,** 1238–1246
20. Ogawa, T., Baker, T. A., van der Ende, A., and Kornberg, A. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 3562–3566
21. Jones, K. A., Kadonaga, J. T., Rosenfeld, P. J., Kelly, T. J., and Tjian, R. (1987) *Cell* **48,** 79–89
22. Wasserman, S. A., and Cozzarelli, N. R. (1986) *Science* **232,** 951–960
23. Greider, C. W., and Blackburn, E. H. (1985) *Cell* **43,** 405–413
24. Ogura, T., and Hiraga, S. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 4784–4788

[4] I have set an arbitrary limit of recent references. When a reference is lacking, consult one of the reviews (1–4) or write me.