CHAPTER

3

REPLICATION

Chapter Outline

- Replication is the process in which individual strands in a DNA duplex are separated, before each strand is copied into a new complementary partner. This copying enables genetic information to be passed on to two daughter DNA molecules, and it is the process that underlies inheritance.
- DNA polymerase is the enzyme responsible for this process, and many DNA polymerases replicate different parts of a chromosome simultaneously in large replication factories.
- The mechanics of synthesis is complicated, and involves the cooperation of many other enzymes to separate parental strands, copy small segments of the template, stitch those copies together, and proofread them to see if any mistakes have been made.
- A critical step controlling the rate of replication is the first one—the initiation of synthesis. Not surprisingly, this step is tightly controlled. It involves attachment of an "origin" of replication within the body of a chromosome to a polymerase.
- Special problems are posed by replicating the very ends of a chromosome.

PRINCIPLES

All organisms must duplicate (replicate) their DNA before every cell division. This requires the coordinated action of many different enzymes, which must be fast enough to complete the complex task in a reasonable time, yet accurate enough to introduce few errors during the process. Base-pairs between the two strands of the double helix are broken at the **replication fork** to allow new base pairs to form between incoming *d*eoxy*n*ucleotide 5′-*t*riphos*p*hates (dNTPs) and bases in the template strand. **DNA polymerase** is the enzyme responsible for polymerizing these dNTPs into DNA. It can copy any sequence in a template by catalyzing the reaction

$$(\text{dNMP})_n + \text{dNTP} \leftrightarrow (\text{dNMP})_{n+1} + \text{PP}_i$$

so a polymer containing *n* subunits of *d*eoxy*n*ucleotide *m*ono*p*hosphate (dNMP) is elongated by one dNMP, as pyrophosphate is produced. The enzyme has two substrates—a primer base-paired to a template, and a dNTP. Nucleophilic attack by the terminal 3′-OH group of the primer on the α-phosphate of an incoming dNTP generates a 3′,5′-phosphodiester bond linking the two substrates (Fig. 3-1). Cleavage of pyrophosphate by pyrophosphatase makes the reaction essentially irreversible. The polymerase

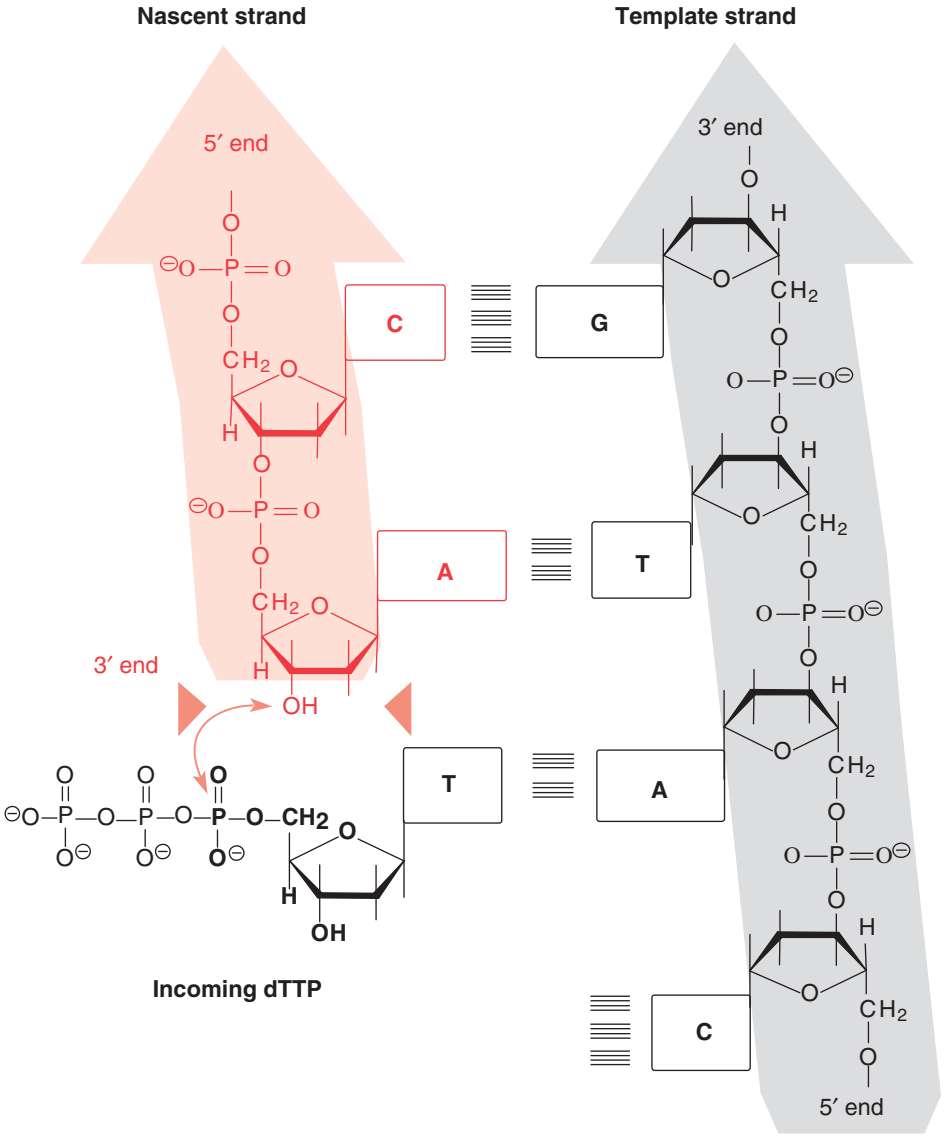


Figure 3-1.

DNA synthesis. Two *arrowheads* mark the polymerization site. An adenine (*A*) in the template strand pairs with an incoming dTTP. TMP is then linked to the 3′ end of the nascent strand, pyrophosphate is released, as nascent and template strands—which are base-paired together—are extruded upwards. The whole process is driven by the favorable free-energy change that accompanies pyrophosphate release. Now, the cycle can repeat; this time an incoming dGTP (not shown) will pair with the C in the template strand.

requires a primer and so cannot initiate synthesis of new chains. Nevertheless, it extends preexisting chains quickly and accurately; bacterial and mammalian enzymes polymerize ~500 and ~50 nucleotides/s, respectively (Box 3-1).

The basic principles involved in replicating eukaryotic DNA are:

- DNA synthesis is restricted to a specific phase of the cell cycle known as S phase. Following mitosis, cells progress through G1 phase, into S phase, and then into G2 phase. This cell cycle is discussed in Chapter 7.
- S-phase DNA synthesis occurs **semiconservatively** (Fig. 3-2). After replication is complete, each daughter DNA molecule contains one original strand and one newly-made strand. In the alternative (ie conservative replication), one daughter would end up with two new strands while the other got two old ones.
- Replication does not initiate at the ends of human DNA molecules, but at many internal sites spaced every ~50 kbp known as **origins**. DNA on both sides of an origin is replicated at the same time. The overall rate of replication is determined largely by the number of origins used, and the rate at which they initiate; the rate of elongation of different nascent DNA chains varies little.

BOX 3-1.
Bacterial DNA Polymerases

E. coli possesses three major DNA polymerases: I, II, and III. Polymerase I was the first to be purified from cells and it has been characterized in detail. Then, mutants were discovered that grew normally but lacked polymerase I; they were deficient in repairing damage in DNA. This finding prompted the search for the true replicative activity; eventually, additional polymerases were discovered, including polymerase III—the principal replicative activity. (Another polymerase is encoded by *UmuC*; this is part of a complex that can replicate a damaged template.)

There are only 10–40 molecules of polymerase III in each bacterium. After a ~7500-fold purification, it exists as a complex of ~20 different polypeptides with a mass of ~900 kD; this complex then dissociates on dilution into various subassemblies. The holoenzyme contains 10 different polypeptides that break down into the γ-complex and core. This core contains three subunits: α (the polymerase), ε (the 3′→5′ exonuclease), and θ.

The history of bacterial polymerases highlights two enduring problems faced by biochemists trying to purify eukaryotic enzymes:

- Isolation of a major activity seen *in vitro* does not necessarily mean that particular activity will be the major one *in vivo*; other polymerases involved in repairing damage can easily obscure the one involved in replicating cellular DNA, especially when broken templates are used during assay.
- Large molecular assemblies involved in chromosomal replication inevitably break up during fractionation, so the "polymerase" is arbitrarily defined by the particular procedure used during purification and assay.

(Baker and Wickner, 1992)

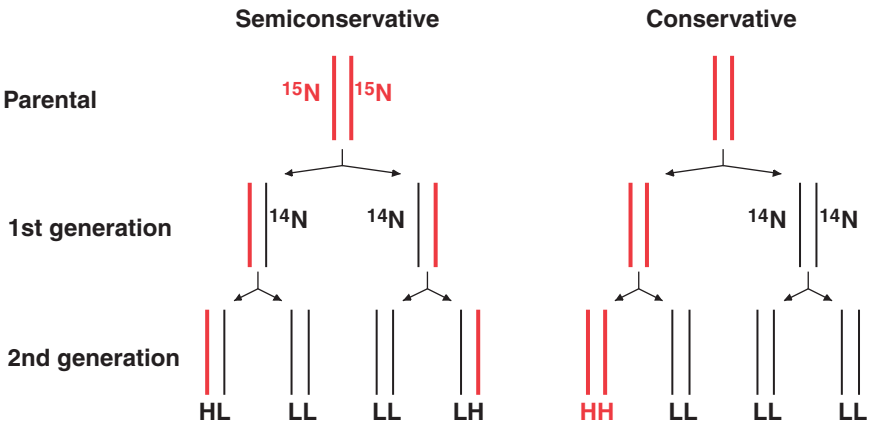


Figure 3-2.

Semiconservative and conservative modes of replication. The two modes of replication were originally distinguished by growing bacteria for several generations in a medium containing a "heavy" isotope of nitrogen (i.e., in $^{15}NH_4Cl$), and then switching the cells to normal ("light") medium (i.e., containing $^{14}NH_4Cl$). At different periods after switching, DNA was isolated and spun on a density gradient containing caesium chloride to resolve "heavy-heavy" DNA (HH; containing ^{15}N in both strands), "heavy-light" DNA (HL; ^{15}N in one strand, ^{14}N in the other), and "light-light" DNA (LL; ^{14}N in both strands). After growth for one generation in $^{14}NH_4Cl$, DNA was exclusively heavy-light; after two generations half was heavy-light, and half was light-light. This result was consistent with semiconservative replication.

- Many different parts of a chromosome are duplicated simultaneously in an enormous **replication factory** (diameter 100–1000 nm), where the appropriate factors are concentrated. Several different polypeptides can be purified as a unit known as a DNA polymerase, which can copy one template strand at a time. Sets of at least four such DNA polymerases form part of a polymerizing machine that duplicates the four parental strands around one origin. Many such polymerizing machines are housed in a factory.
- The duplication of a DNA segment involves several steps; each involves a complicated set of reactions (Fig. 3-3). First, the origin attaches to one machine in the factory (stage 1). The two strands in the origin separate, exposing bases to give a **replication bubble** that contains two **replication forks**, each associated with a pair of polymerizing sites (stage 2). After synthesis begins, the four parental duplexes slide through the four polymerizing sites, as two daughter duplexes are extruded (stages 3–5). Both chain growth and template movement are driven by the favorable free-energy change that accompanies release of pyrophosphate. Although no direct measurements have been made on the pushing/pulling power of a DNA polymerase, a bacterial RNA polymerase has enough power to push/pull double-helical templates of the necessary size (page ___).
- RNA synthesis is required for DNA synthesis. A DNA polymerase cannot start a new DNA chain from the 3′ OH of the first dNTP. However, it can create a phosphodiester bond between an incoming dNTP and the 3′-OH group of a preexisting RNA molecule. Such RNA **primers** are made in short pieces by a special RNA polymerase associated with a DNA polymerase called a **primase**.

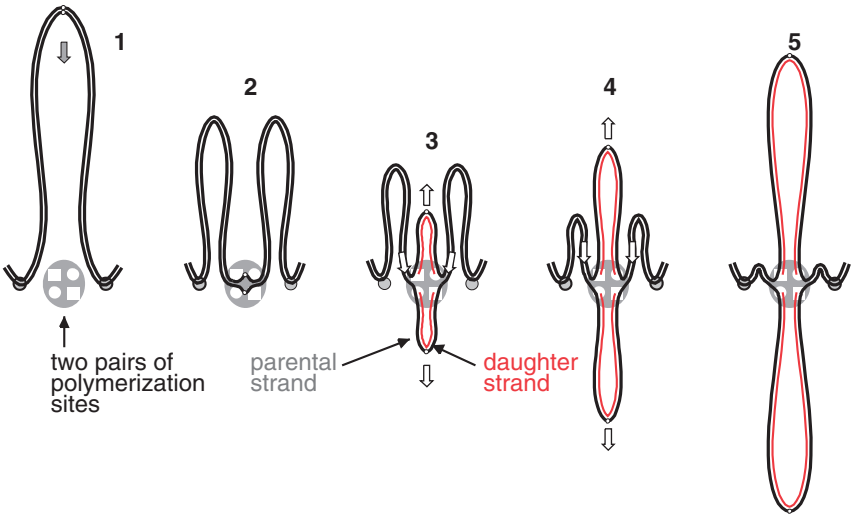


Figure 3-3.

The basic steps during replication. (1) Origin binding. An origin (*small circle*) within a chromatin loop binds (*arrow*) to a polymerizing machine containing two pairs of polymerizing sites on the surface of a factory. (2) Strands within the origin separate, exposing bases; synthesis is ready to begin. (3, 4) Elongation. Daughter strands are extruded (*arrows*), as parental duplexes slide (*arrows*) through polymerizing sites; during this process, parental loops shrink and daughter loops grow, until most DNA is duplicated. (The origin is shown here and in later figures to detach from the factory after initiation, but it may remain attached throughout the cycle.) (5) Most DNA has been replicated; special mechanisms are required to duplicate the last pieces on each side (page 143). [Modified from P. Hozák *et al.* (in *Eukaryotic DNA Replication,* ed J.J. Blow); copyright © 1996, by permission of Oxford University Press.]

- New DNA strands grow at the 3′ end—growth is said to be in the 5′-to-3′ direction—as template strands are pulled through polymerizing sites.
- DNA synthesis occurs **continuously** on one strand, and **discontinuously** on the other.

Tracking Versus Immobile DNA Polymerases

(Wessel *et al.*, 1992; Jackson and Cook, 1986)

The polymerization site and template must move relative to each other, to allow each new template base to occupy the site in turn. Until recently, it was tacitly assumed that a (small) DNA polymerase tracked along a larger template, and this is reflected by the models generally drawn in textbooks (Fig. 3-4, *A–C*). However, recent evidence is inconsistent with this assumption, and suggests that the polymerase remains still, while the template moves instead (Fig. 3-4, *D–F*). Note that both models require a motor activity, either to move the template, or to move the polymerase. In both models, energy is derived from the same source—hydrolysis of deoxynucleotide triphosphates.

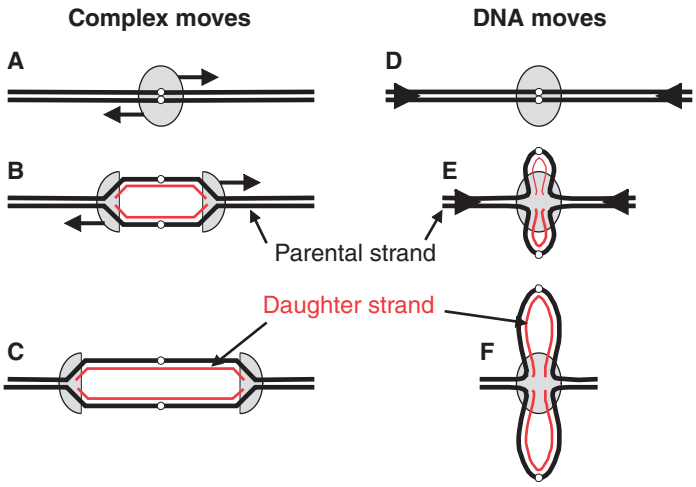


Figure 3-4.

Models for replication involving (*A–C*) tracking and (*D–F*) immobile polymerases. (*A-C*) The polymerizing complex (*oval*) binds to the origin (*small circles*), splits into two, and the two halves track along the template (*thick lines*). Nascent DNA (*thin red lines*) is made at the two pairs of polymerization sites as they move apart. (*D-F*) The template passes through a fixed complex as nascent DNA is extruded; all four polymerizing sites remain together. [Modified from Trends Cell Biol., Volume *4*, P. Hozák and P.R. Cook, pages 48–52, copyright © 1996, with permission from Elsevier Science.

What is the evidence for the old idea that a DNA polymerase tracks like a locomotive along the template (as in Fig. 3-4, *A–C*)? Although many results that will be described can be accommodated by models involving tracking polymerases, few experiments address this question directly. Consider, for example, results obtained by DNA fiber autoradiography (Fig. 3-17; Box 2-4). In such autoradiographs, a track of grains is seen that reflects the underlying incorporation by an individual polymerizing complex of [^{3}H]thymidine into nascent DNA. Longer incubations give longer grain tracks. Those that have been brought up on old textbooks that depict moving polymerases naturally interpret the lengthening track in terms of an underlying movement of the polymerase along the template. However, such linear grain tracks are only seen after destroying nuclear organization and spreading the now-naked DNA; they only reflect indirectly the structure found *in vivo*. Moreover, they tell us nothing about polymerase movement, though they may confirm our preconceptions.

Perhaps the best evidence for the movement of polymerases is even more indirect. It is based upon the following argument: modern biochemistry has proved very successful at dissecting the details of replication, and as we assume that this success is based upon the correctness of the conventional model, then—the argument runs—that model must be correct. This can be put in another way: if small soluble polymerases purified by biochemists work *in vitro*, there is little need to postulate the existence of larger immobilizing structures *in vivo*. Nevertheless, this argument is specious; whether polymerases work *in vitro* tells us nothing about whether or not they move *in vivo*. Moreover, we now know that we have misjudged the size of polymerases; they are contained in enormous factories that dwarf the template. And although small soluble isolates do work, they do so poorly.

The above argument for tracking by DNA polymerases has been compromised by results of a simple experiment using pure proteins. As we shall see, a helicase is required at each replication fork to separate the two intertwined strands into the single ones on which the polymerase can act; multimeric complexes of the *t*umor (T) antigen play this role during replication of the monkey virus, SV40. Pure T antigen was added to a linear template carrying the SV40 origin of replication, and the structure of the resulting complexes was analyzed by electron microscopy (Fig. 3-5). A dodecameric T antigen complex bound initially at the origin, and in the presence of ATP (and a single-strand binding protein used to stabilize unwound single strands), the expectation was that the dodecamer would split, and then the two half-complexes would each track like a locomotive away from the origin, unwinding the template to generate a replication bubble with a hexamer at each fork (Fig. 3-5*B*). However, both half-complexes (Fig. 3-5*B*) and intact dodecamers associated with the single-stranded loops (Fig. 3-5*D*) were seen—an inconclusive result. Nevertheless, this prompted the questions, did two (active) half-complexes aggregate to form a dodecamer, or did an active dodecamer break down into two half-complexes? Fortuitously, addition of a monoclonal antibody directed against the T antigen stimulated unwinding fivefold and increased the proportion of dodecamers, showing that they were the active forms. These results are consistent with DNA moving through the active dodecamer, followed by breakdown to give two inactive half complexes.

So in this simple system involving pure reagents, each fork is attached to, and so immobilized by, its partner. It seems that the two half-complexes do not move like two locomotives down a track away from each other; instead, they try to move toward each other but cannot. The result is that the two rails in the track (i.e., helical strands) are inevitably pulled past them, separated, and extruded into two loops. Since the two half-complexes dictate the geometry of the two replication forks, polymerases acting at those forks must adopt the same geometry—and immobility. As we shall see, direct imaging of polymerases in living bacteria also shows that the two forks are together.

Studies on nuclear matrices provided the first good evidence that nascent DNA, and so the associated polymerases, were attached to an underlying skeleton (Box 2-6); however, results obtained with these structures were never widely accepted because it was impossible to counter the criticism that the attachments seen had been generated artifactually during the isolation procedure. Now, more convincing evidence has been obtained

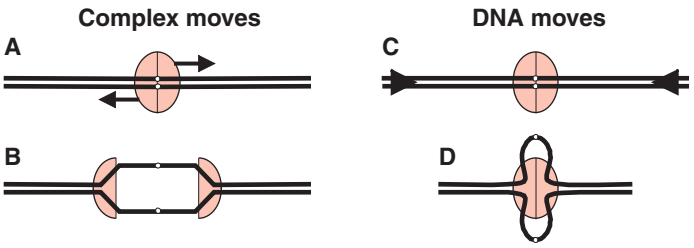


Figure 3-5.

The T antigen complex does not track along the DNA. (*A, B*) A dodecameric T antigen binds to the origin, splits, and the two hexamers track along the two strands, unwinding them. No DNA synthesis occurs as no polymerase is present. (*C, D*) Duplexes slide from each side through the dodecamer, as two single-stranded loops are extruded. Results seen by electron microscopy are consistent with this model.

using more physiological conditions, and the approach illustrated in Figure 3-6. Human cells were encapsulated in agarose microbeads to protect the fragile cells during subsequent manipulation, and permeabilized (Fig. 2-11). Models involving tracking or immobile polymerases were distinguished by cutting chromatin loops into small fragments, and then removing the fragments by electrophoresis (Box 1-5). If polymerases tracked around chromatin loops, they should be detached from the underlying structure with most chromatin. On the other hand, if they were attached, they should remain associated with the structure. It was found that polymerizing activity remained in nuclei, despite removal of most chromatin. Nascent DNA also resisted elution whether it was pulse-labeled *in vivo* using [^{3}H]thymidine, or *in vitro* using [^{32}P]TTP. These results are simply explained if polymerases are attached, directly or indirectly, to a large structure that cannot be eluted.

Attachment of polymerases to a large structure in the nucleus (i.e., a factory) has made both the purification of relevant enzyme activities and the development of *in vitro* assays for polymerization difficult. For example, there is still no efficient *in vitro* system derived from mammalian cells that initiates DNA synthesis on intact chromatin templates. Much of what we know is derived from the use of frog egg extracts (page 36) or small viral templates supplemented with SV40 T antigen; both represent special cases. And most biochemists prefer to work with soluble activities that can be fractionated easily on columns. Therefore, they usually detach polymerizing activities as a first step in their procedure, but these detached activities work inefficiently. As a result, there is controversy about which detached activities (i.e., DNA polymerases α, δ, and ϵ) are responsible for DNA synthesis on which of the parental strands. These controversies will

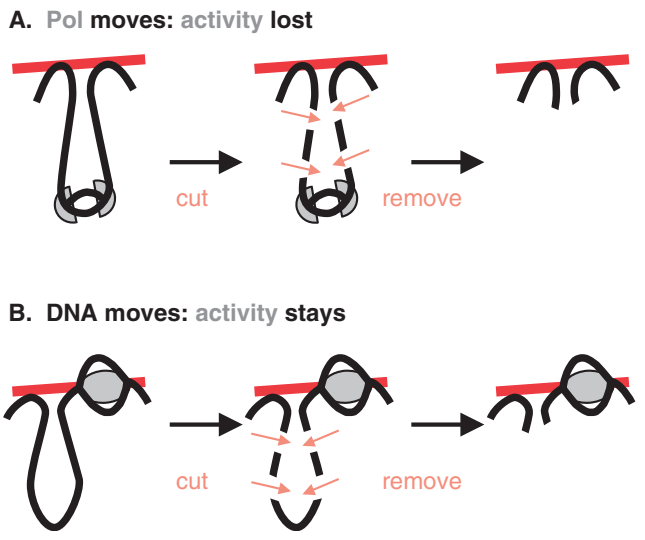


Figure 3-6.

Are active DNA polymerases attached to large structures in the nucleus? (*A*) A chromatin loop is shown attached to the nucleoskeleton, and polymerases (*semicircles*) track away from each other as they replicate out in the loop. After cutting chromatin into ~10 kbp pieces with a restriction enzyme (*arrows*), chromatin should electroelute with associated polymerases and be lost. (*B*) DNA is replicated as it moves through polymerases fixed in a factory (*oval*) attached to a nucleoskeleton. Despite cutting (*arrows*) and electroelution to remove 75% chromatin, all polymerizing activity should remain. This is the result found.

probably not be resolved until efficient systems that replicate native chromatin templates are developed, and this will require less-destructive fractionation. Therefore, it should be borne in mind that the models presented here and elsewhere for the mechanism of replication are based on extrapolations from the inefficient activities seen *in vitro*; they will have to be modified, or perhaps altered in some major aspects, as new evidence comes to light.

REPLICATION FACTORIES

(Nakamura *et al.*, 1986; Hozák *et al.*, 1993; Lemon and Grossman, 1998; Cook, 1999)

Seeing is believing; the most convincing evidence for immobilized polymerases is provided by pictures of replication sites in whole cells. However, visualizing the actual sites of synthesis is difficult. Although the relevant polymerases can be immunolabeled, many may be inactive and stored far from synthetic sites, so it is better to detect sites where tagged DNA precursors (e.g., [^{3}H]thymidine) are incorporated into nascent DNA. This, however, brings associated problems, and these problems must be viewed against the rapid rate of synthesis that occurs *in vivo*.

Consider, for example detecting synthetic sites by autoradiography after incubating living cells in [^{3}H]thymidine. The radiolabel must first enter the cell and be converted to [^{3}H]dTTP. The nuclear [^{3}H]dTTP concentration is initially low compared with that of the endogenous pool of dTTP, so little [^{3}H] is incorporated into DNA. Then, as more [^{3}H]dTTP becomes available, more [^{3}H] is incorporated, but it takes several minutes before sufficient is incorporated to be detected. Therefore, during a typical labeling period of 10 min, ~15,000 nucleotides might be polymerized into DNA, and—if uncompacted—this would stretch across a nucleus with a diameter of ~5 μm (Fig. 1-2; Table 1-1)! So even if nascent DNA can be detected, it does not necessarily lie at its site of synthesis. Autoradiographic detection further compounds the problem (Box 2-4); pathlengths of the β-particles emitted by ^{3}H can be many tens, even hundreds, of nanometers long, so that the resulting silver grains will not necessarily lie over the nascent DNA (Fig. 1-2).

Factories were first imaged in rat embryo fibroblasts using a higher-resolution technique. Cells were grown to confluency so that they stopped replicating, and replated at a low density to induce them to reenter S-phase semisynchronously. A thymidine analogue—bromo-deoxyuridine (Br-dU)—was added so that it could be incorporated into DNA, and then labeled using fluorescently tagged antibodies. This approach provides better spatial resolution than autoradiography (Fig. 1-2). Label first appeared in small, discrete, foci scattered throughout nuclei. Subsequently these enlarged, implying that DNA was extruded from them once it had been made.

Problems associated with the rapid rate of replication *in vivo* can be sidestepped by permeabilizing cells and washing away endogenous pools; then elongation rates can be reduced by lowering the triphosphate concentrations. Figure 3-7 illustrates results obtained when permeabilized HeLa cells are incubated with the thymidine analogue, biotin-dUTP, before sites containing incorporated biotin were immunolabeled; the fluorescent foci (i.e., factories) have a characteristic size and distribution that change during S-phase. Calculations show that between 40 and many thousands of forks are active in the smallest and largest factories respectively. (This calculation requires knowledge of the number of factories, rate of fork progression, spacing between forks, genome size, and length of S-phase.)

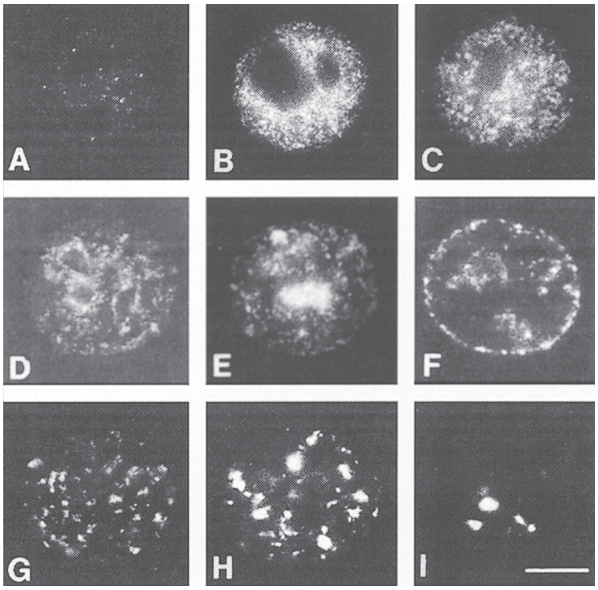


Figure 3-7.

Fluorescence micrographs of replication factories at different stages of S-phase. Synchronized HeLa cells were encapsulated in agarose, permeabilized, incubated with biotin-dUTP, and incorporation sites indirectly immunolabeled with fluorescently tagged antibodies. Fluorescence marks replication sites, which change in number and distribution as cells progress from (*A*) early to (*I*) late S-phase. Bar: 5 μm. [Reprinted, with permission, from P. Hozák *et al.* (J. Cell Sci. *107*, 2191–2202); copyright © 1994 Company of Biologists Ltd.]

All factories with sizes below the resolution limit of the light microscope appear ~200 nm in diameter (Box 1-2). Higher resolution can be achieved by electron microscopy. This is only useful if samples are allowed to incorporate label for short periods (so incorporated analogs cannot move far from synthetic sites), and if immunochemical methods of sufficient sensitivity are available (so those analogs can be detected). Such methods have now been developed. In the experiment illustrated in Figure 3-8, permeabilized HeLa cells were incubated briefly with biotin-dUTP so nascent DNA chains were extended by only ~500 nucleotides; then sites containing incorporated biotin were immunolabeled with 5-nm gold particles. Visibility of the underlying structures was improved by removing ~90% chromatin (by cutting chromatin with an endonuclease and then removing it by electrophoresis, as in Fig. 3-6). Residual clumps of chromatin and several electron-dense bodies—the factories—can be seen attached to a nucleoskeleton. Most immunolabeling gold particles marking nascent DNA are associated with the factories. As the incubation time *in vitro* was increased, longer pieces of DNA were made and gold particles were found progressively further away from factories; this implied that nascent DNA was extruded from factories as templates passed through. Immunolabeling also shows that factories contain DNA polymerases. [Even using this high-resolution technique, gold particles still mark synthetic sites imprecisely (Fig. 1-2). First, they are connected to incorporated biotin by an antibody bridge of ~20 nm. Second, ~about 500 nucleotides (contour length ~170 nm when fully stretched) was polymerized during incubation, so some particles mark nascent DNA that has moved away from synthetic sites.]

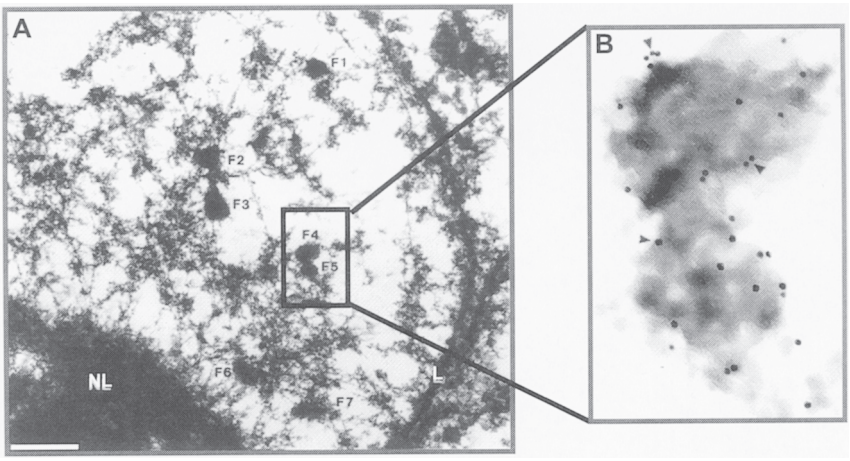


Figure 3-8.

Replication factories imaged in the electron microscope. Encapsulated cells in early S phase were permeabilized, incubated with biotin-dUTP, treated with nucleases, ~90% chromatin removed, and sites of biotin incorporation immunolabeled with 5-nm gold particles; then, samples were stained, embedded in a wax, sections (400 nm) cut, the wax removed, and the sections imaged. (*A*) Seven replication factories (F1-7). *NL*: nucleolus. *L*: nuclear lamina. 72% of the 180 gold particles in the nuclear region (not visible at this magnification) were in factories, showing that factories were sites of DNA synthesis. Bar: 0.5 μm. (*B*) Higher-power view of F4 and F5; three *arrowheads* indicate some of the 30 gold particles. [Reprinted, with permission, from P. Hozák *et al.* (Cell *73*, 361–373); copyright © 1993 Cell Press.]

Bacterial polymerases are also fixed in factories. The enzymes in living *B. subtilis* were visualized using a construct in which the catalytic subunit of PolC was fused with the green fluorescent protein (Box 1-3). If the two forks moved independently (as in Fig. 3-4, *A–C*), we would expect to see two fluorescent spots somewhere in the area occupied by DNA. However, one discrete spot was generally seen in the middle of the cell (as in Fig. 3-4, *D–F*). As it was unlikely that the method was sensitive enough to detect only four polymerases (one on each arm of the two forks), many of the ~40 polymerase molecules in a cell were probably concentrated in one factory. This prompts the question: why are so many polymerases found in the factory? Two related answers can be suggested. First, a high local concentration is probably required for efficient initiation, and these high numbers may simply persist during elongation. Second, those high numbers may also facilitate template re-engagement following repair. Although DNA is a stable molecule, it nevertheless degrades spontaneously in its "normal" surroundings (Fig. 5-1). Therefore, an engaged polymerase may inevitably encounter damage as it replicates a long template, and so have to disengage to allow access of the repair machinery. Then, a local concentration of polymerases will facilitate re-engagement, and so replication of the whole genome.

The discovery that replication takes place within factories in eukaryotes poses many questions: Is a factory made from identical subunits? Are specific components concentrated within specific regions of each factory? What directs those components to a factory, or to particular sites within a factory? Are different factories assembled at different times of S-phase? What triggers factory assembly?

THE MECHANICS OF SYNTHESIS AT THE FORK

Two properties of DNA polymerase ensure that the stereochemistry of replication is not straightforward: the polymerase can neither initiate synthesis of a new chain, nor can it extend a growing chain from the 5′ end. This means that special mechanisms are required to initiate chains, and to elongate one strand. These special mechanisms, and others involved during the elongation of nascent DNA chains, will now be discussed.

Separating Parental Strands
(West, 1996)

The double helix is often drawn as a ladder, so we might imagine that the two strands can be separated simply by breaking the rungs (i.e., the base pairs). But this view ignores the fact that each strand is entwined (i.e., interlocked) about its partner, and must be untwined during replication. This **unwinding problem** (Fig. 3-9) can be solved in two ways:

- The entire chromosome could rotate about its axis, once for every 10 bp separated. Pulling apart (intertwined) strands in a piece of string provides a simple analogy; as the strands are separated from one end, the whole piece of string rotates about its axis. However, the ends of the chromosome are not free to rotate; in bacteria there are no ends (the genome is circular), and in eukaryotes each end of a chromatin loop is tied down. If the end of our piece of string is not free to rotate, we can pull apart the strands but we can never completely separate them as they remain interlocked in a higher-order tangle.
- If rotation about ends is not permitted, the two strands can only be untwined by cutting one or other of the strands (or both), passing one (or both) strands through the break, and then resealing the break. **Topoisomerases** are the enzymes that achieve this remarkable reaction (Box 3-2), and they are absolutely required during replication to solve the unwinding problem by providing the necessary (transient) swivel.

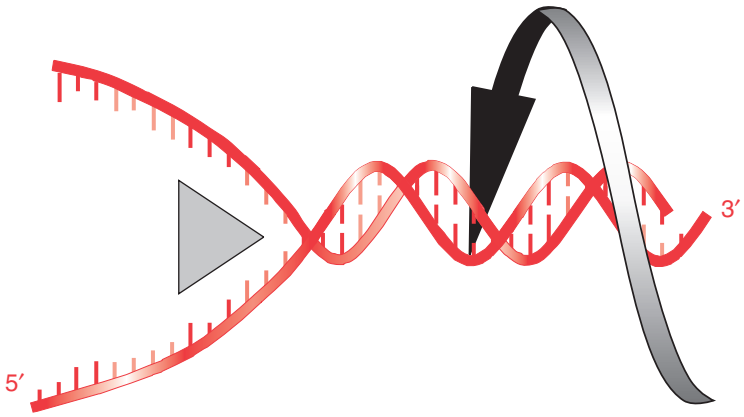


Figure 3-9.
The unwinding problem. Bacteria replicate DNA at ~500 nucleotides/s, so the parental helix must unwind and rotate at ~50 revolutions/s to separate strands as it passes into the replication complex. Strand separation is achieved by a helicase (*arrowhead*), and unwinding is facilitated by a topoisomerase (not shown) that allows the parental duplex to spin as it advances. The helicase is represented here and in some other figures as an *arrowhead*, but it probably forms a ring around the template, much like the circular clamp to be discussed later.

BOX 3-2.
Topoisomerases and Anticancer Drugs

Topoisomerases are reversible nucleases; they cut and then mend DNA strands, usually after changing its higher-order folding. [The term *topoisomerase* is derived from the Greek (*topos* = place, *isos* = equal).] There are two main types:

- Type I enzymes transiently cleave a single DNA strand in the duplex to allow one (or both) cut ends to rotate around the intact strand, using a phosphodiester bond as a swivel. Torsional tension in the helix drives this rotation so that these topoisomerases release any tension that builds when the helix is unwound during replication or transcription. They are monomeric, and the reaction involves a transesterification to a tyrosine.
- Type II enzymes transiently break both strands of a helix. All belong to a single family; the prokaryotic enzymes (DNA gyrase and topoisomerase IV) are A_2B_2 tetramers, while the amino- and carboxy-terminal parts of each of the two subunits of the eukaryotic topoisomerase II are homologous to the bacterial B and A subunits respectively. The crystal structure of the yeast enzyme reveals a heart-shaped dimeric protein with a large central hole; it leads to a model in which two sets of articulated jaws clamp two duplexes and pass one through the other. Cleavage is a transesterification between a pair of tyrosyl residues (one in each half of the dimeric enzyme) and a pair of DNA phosphodiester bonds four base pairs apart. During the process, phenolic oxygens of tyrosines become (temporarily) covalently linked to phosphoryl groups at the 5′ ends of the transiently broken DNA, leaving a pair of hydroxyl groups at the recessed 3′ ends. The intervening ends then separate, and the pair of tyrosine-linked 5′ ends move away from each other in an ATP-dependent process, opening a "gate" in the double helix for the transport of another duplex. After the second duplex has passed through, the gate closes, and a transesterification between the 3′ hydroxyl groups and the phosphotyrosyl linkages restores the continuity of the DNA strands and breaks the covalent bonds between enzyme and DNA. This mechanism requires a dramatic change in conformation to open the gate by ~4 nm to allow the second duplex to pass through; it is formally equivalent to the action of other molecular motors. Therefore, type II enzymes can separate two interlocked DNA circles or loops and are essential for the segregation of newly replicated chromosome pairs. Their usefulness can be appreciated by anybody who has tried to disentangle an entangled kite string without using scissors. Some type II enzymes also use ATP to introduce supercoils into DNA.

As topoisomerases are required during replication, and as tumor cells rapidly replicate their DNA, these enzymes are natural targets for anticancer drugs. The drugs include m-AMSA, etoposide, and teniposide. However, cisplatin—which is the only really successful anticancer drug to be discovered in the last 20 years—prevents unwinding, but not by inhibiting a topoisomerase. The platinum uses two of its ligand-binding sites to cross-link the two DNA strands, so that they cannot separate before replication. Its therapeutic activity was discovered in a

roundabout way by Barnett Rosenberg, who was studying how bacteria grew in an electric field. When platinum electrodes were used, the bacteria grew long and thin, like spaghetti; the platinum combined with ammonia in the buffer to create a complex that inhibited cell division. Although this compound had been known for more than 100 years, it had not been tested as an anticancer agent; Rosenberg did so, and found that it worked. It inhibited the growth of the rapidly dividing cancer cells more than others.

(Wang, 1991; Berger *et al*., 1996)

DNA helicase pries apart (**melts**) the two strands of the double helix to provide the single-stranded region that can be used by a polymerase (Fig. 3-10). Because strands can only be separated by heating pure DNA in a salt solution to ~90°C, it is not surprising that helicases need ATP for the process. However, it is surprising that some can unwind more than 1000 bp each second. Usually, a particular helicase will unwind in either the 5′→3′ or the 3′→5′ direction, but not both; for example, T antigen is a 3′→5′ helicase. Helicases are assayed by their ability to melt (in an ATP-dependent manner) a short radiolabeled oligonucleotide base-paired to a longer strand; the released oligonucleotide is easily resolved from the partially double-stranded complex by gel electrophoresis (Box 1-5). Like another DNA-binding protein known as *p*roliferating *c*ell *n*uclear *a*ntigen (PCNA; Fig. 3-11), the bacterial enzyme encoded by the *DnaB* gene forms a hexameric ring with a central hole that is just large enough to accommodate a DNA duplex. Although they were originally conceived to be DNA-unwinding enzymes, their role is more as a molecular motor or pump that drives a DNA duplex through the central hole and so into the replication complex. More than 12 helicases acting on DNA or RNA have been characterized in *E. coli*, with all sharing seven distinct and conserved motifs. Similar motifs are found in >1000 different human proteins, and mutations in these motifs in two proteins give rise to Bloom's (Box 5-2) and Werner's syndromes (characterized by premature aging), respectively.

Once the two strands have been separated, they are prevented from base-pairing again by ***s*ingle-*s*trand DNA-*b*inding (SSB) proteins**, or helix-destabilizing proteins. The gene 32 protein (gp32) of phage T4 is the prototypic example of this class of protein; it was isolated by its ability to bind to single-stranded (but not double-stranded) DNA. It is absolutely required for replication, and the X-ray crystal structure reveals an unstructured DNA strand in a cleft in the molecule. Each monomer binds to ~10 nucleotides, and many monomers bind cooperatively with an affinity for single-stranded DNA that can be as much as 10^5 times greater than for duplex DNA. Because infected cells contain ~10^4 gp32 molecules and ~60 replication forks, there is sufficient to cover ~850 bp on each of the parental strands at each fork. Even if fewer molecules are involved, the single-stranded regions made by the helicases are still likely to be extensive.

RNA Primers

DNA polymerases cannot initiate synthesis of new chains; they can only extend pre-existing chains. However, RNA polymerases can initiate new chains by joining two nucleotide triphosphates together. The special RNA polymerase that acts in DNA synthesis in this way is known as a **primase**. It is found in a complex with DNA polymerase

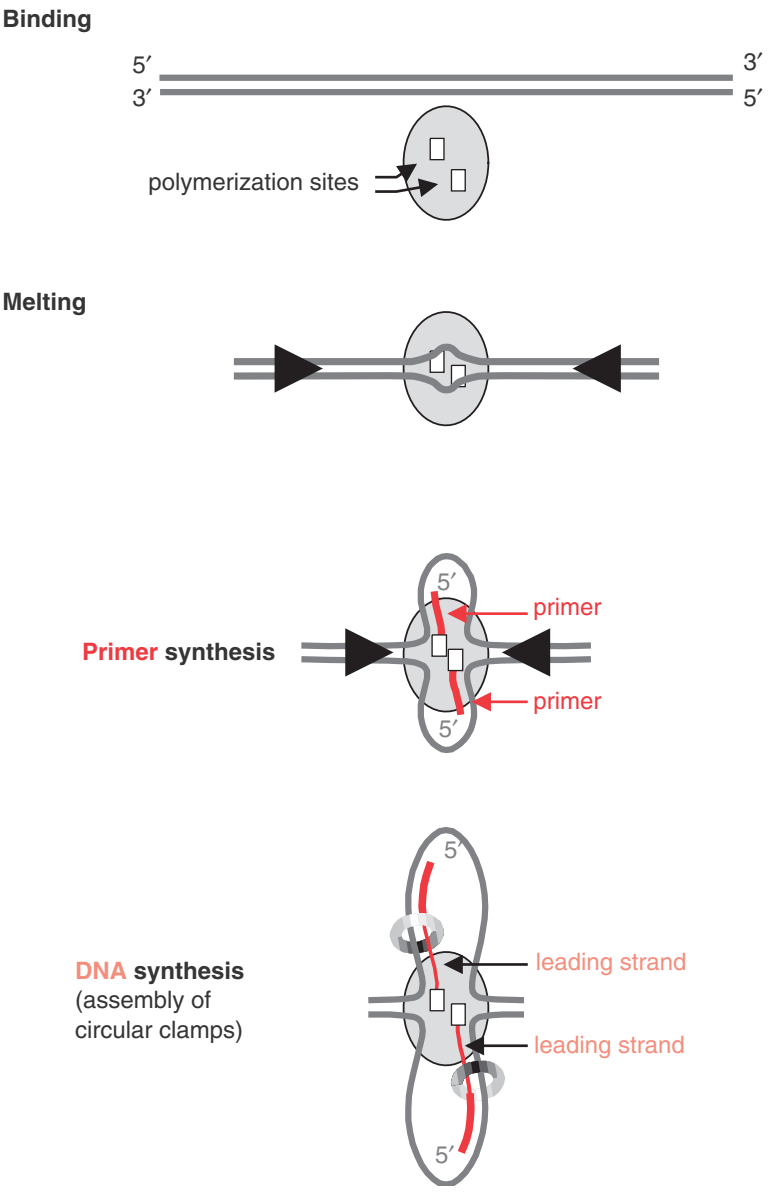


Figure 3-10.

Initial steps during the synthesis of one of the two daughter strands (i.e., the leading strand). Here, synthesis of the other—lagging—strand is omitted for clarity; therefore, only two of the four polymerization sites are shown. After binding to the polymerizing complex, the parental duplex is reeled in from both sides (*arrowheads*) and is melted. Synthesis (5′→3′) of RNA primers and then daughter DNA strands occurs as parental strands pass through the two polymerization sites in the complex; dissociation of templates is prevented by the assembly of circular clamps.

α. Primases make short chains or **RNA primers** of 10–20 nucleotides; as these chains contain a 3′ OH, a DNA polymerase can take over and catalyze addition of a deoxynucleotide on to this 3′ OH (Fig. 3-10). Primase activity can be detected by incorporation of radiolabeled UTP into growing RNA-DNA chains, and the presence of labeled ribonucleotides in DNA can be confirmed by removing the label with an RNase (but not a DNase). Because RNA primers are covalently linked through a phosphodiester bond to the growing DNA chain, they must be removed subsequently and replaced with DNA. As we shall see, this involves nucleolytic degradation of RNA to generate a gap, DNA synthesis by a DNA polymerase to fill the gap, followed by ligation of the ends to give a continuous strand of DNA.

The Asymmetric Fork

(Ogawa and Okazaki, 1980; Stukenberg *et al.*, 1991; Krishna *et al.*, 1994; Wyman and Botchan, 1995; Hozák *et al.*, 1996; Waga and Stillman, 1998)

During synthesis, two segments of duplex flow from each side into one complex containing four polymerizing sites. Copying one of the two template strands in each duplex is straightforward (Fig. 3-10). For the sake of clarity, this synthesis—which is known as **leading-strand synthesis**—will first be discussed alone, although synthesis of the lagging strand occurs concurrently. Once an RNA primer has been made by the primase, DNA polymerase α in the other half of the primase-polymerase complex probably copies a few nucleotides before the main leading-strand polymerase—polymerase δ—takes over. This enzyme then copies most of the loop until the terminus has been reeled in. As the two parental strands in each of the two segments are antiparallel, only one can be copied in this way (i.e., into an RNA primer that grows 5′→3′, with subsequent addition of DNA to the 3′ end). This results in two new DNA chains that have opposite polarities to the template strands.

A remarkable property of the bacterial polymerizing complex is its ability to polymerize ~2×10^6 nucleotides into a leading strand without releasing the template; as the template moves past, it must catch hold of each base in turn without ever missing. The polymerizing subunit of DNA polymerase III—the α-subunit—achieves this by firmly attaching itself to a circular ring clamped around the template. The circular clamp is formed by a dimer of the β subunit. X-ray crystallography shows that this ring fits snugly around a duplex, so that the surface of the strongly positive hole abuts the negatively charged backbone of DNA (Fig. 3-11, *right*). Moreover, once the two monomers fuse into a ring around a circular template, they become irreversibly bound to it. So in much the same way that getting a bead off a circular necklace without breaking the string is impossible, the dimeric β-subunit is essentially irreversibly bound to the circular template. This can be demonstrated by assembling pure [^{3}H]labeled β-rings on to circular (nicked) template in a reaction that requires a "clamp-loader" and ATP; the complex produced is stable enough to pass intact down a gel-filtration column. However, cutting the circle with a restriction enzyme allows the [^{3}H]ring to slip off the end of the now-linear molecule, so that the two can be separated on the column. This ring-like structure is carried over to the mammalian analogue formed from three subunits of PCNA (Fig. 3-11, *left*).

How are the two template strands that are not illustrated in Figure 3-10 copied, when they are antiparallel to those discussed above? One simple possibility considered initially was that some unknown polymerase extended chains from the 5′ end so that

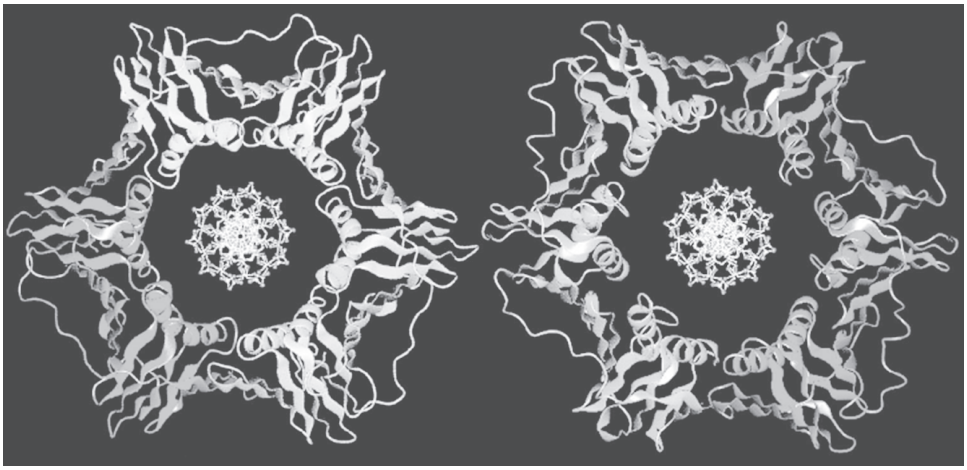


Figure 3-11.

Computer-generated structures of DNA bound to a trimer of yeast PCNA (*left*), and a dimer of the β subunit of bacterial DNA polymerase III (*right*). Polypeptide backbones are represented by ribbons (α helices as spirals and β sheets as flat ribbons); individual polypeptides within each ring are distinguished by different colors. A model of 10 bp of B-form DNA has been placed in the geometric center of each structure; it is viewed end-on, and individual bonds are shown as sticks. PCNA seems to be a molecular adaptor; it carries binding sites for other components of the replication and repair machineries (e.g., RFC, DNA ligase 1, FEN1, XPG). [Reprinted, with permission, from T.S. Krishna *et al.* (Cell *79*, 1233-1243); copyright © 1994 Cell Press.] Figure also appears in Color Figure section.

they grew 3′→5′. Such an unknown enzyme and the known polymerase could then act in concert to copy one parental duplex into two new strands, one growing 3′→5′ and the other 5′→3′. However, no polymerase capable of 3′→5′ extension has been discovered.

There is a logical solution to this problem—the two template strands must move locally in opposite directions past the two adjacent polymerizing sites, so that the two new chains can both grow 5′→3′. But how can two template strands move locally in opposite directions when they move into the complex base-paired together? This process can be achieved after melting the two template strands by passing the leading strand "forwards" through one polymerization site, and the lagging strand "backwards" through the other (Fig. 3-12). This inevitably means that the lagging strand is made in short pieces that are then stitched together. Before discussing such movements in detail, the evidence for the synthesis of the lagging strand in small pieces will be discussed.

As many millions of base pairs can be polymerized into a leading strand, the discovery that the other strand was synthesized in short pieces that were then stitched together came as a surprise. The original discovery was made using T4 bacteriophage; infected bacteria were cooled to 8°C to reduce the replication rate, incubated for 5–50 s in [^{3}H]thymidine, DNA purified, denatured, and nascent chains sized by sedimentation in sucrose gradients. It was expected that label would be incorporated into growing viral chains that ranged in size from the recently initiated to the full-length, and the labeled chains would grow with time. Initially, half the labeled chains had the expected size, but the other half had a length of only 1000–2000 nucleotides. As the incubation time increased, the proportion of these short chains declined. These short chains—which are called **Okazaki fragments** after their discoverer—represent a pool of lagging-strand

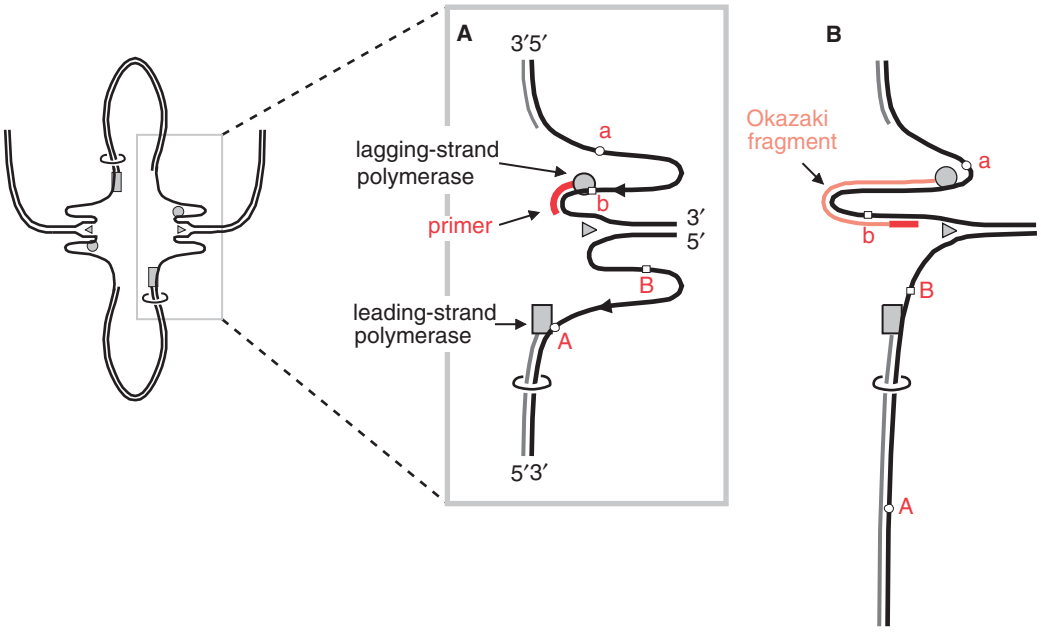


Figure 3-12.

The leading strand moves "forwards" (with duplex flow) through one polymerization site, while the lagging strand moves "backwards" (i.e., against the flow) through the other. Circular clamps associated with lagging-strand polymerases are not shown for clarity. *Left*: One polymerizing complex containing four polymerizing sites is shown. Each of the four single-stranded regions shown in Figure 3-3, stage 3, has been extended into a loop; how this transition occurs will be discussed later. *Panel A*: A blow-up of one side of the complex. Base pair A•a was reeled in past the helicase (*triangle*) before B•b. Base A continues to move with the flow past the leading-strand polymerase and so will be copied before base B. Although base a entered before base b, the flow of this template strand is now reversed. It has attached to the lagging-strand polymerase and moved past it as a primer was extruded; base b will now be copied into DNA. *Panel B*: All the single-stranded DNA in the lower loop in panel A has slid past the leading-strand polymerase so that base A has been copied; base B will be copied soon. Single-stranded DNA in the upper loop in panel A slid backwards against the flow, was copied, and the resulting primer/Okazaki fragment is now on the other side of the polymerase. Base b has been copied before base a.

chains synthesized discontinuously in the 5′→3′ direction. They are then stitched together by **DNA ligase**. Although the number of chains in this pool remains constant during ongoing replication as new ones are made and others are removed by ligation, the relative amount of label in it declines as label accumulates in the ligated pieces. Such continuous synthesis of one strand and discontinuous synthesis of short pieces that are then stitched together inevitably means that the replication fork is asymmetric. Similar results were soon found in mutant bacteria that are unable to stitch the pieces together, and in mammalian cells where the Okazaki fragments are shorter (i.e., 100–200 nucleotides).

Just as an RNA primer is required to initiate synthesis of each leading strand, another RNA primer is also required for each Okazaki fragment. In higher eukaryotes, each primer-Okazaki fragment is probably made successively by a primase, polymerase α, and perhaps polymerase δ. Therefore, both leading and lagging strands contain RNA

at their 5′ ends. This must be removed (by an RNase H activity that can degrade the RNA strand of a DNA/RNA helix), the gaps filled in (by a DNA polymerase), and ends ligated (by a DNA ligase). Figure 3-13 illustrates a possible model for the movements that take place during lagging-strand synthesis.

Figure 3-14 illustrates how template movements might be choreographed to ensure that global template movements in one direction can be reconciled with local movements of the two template strands with—and against—duplex flow. The energy for the movement of the DNA strands is provided from two sources: ATP (used by helicases), and the dNTPs (used at the various polymerizing sites). We can now see how the inability of DNA polymerases either to initiate synthesis of new chains, or to extend growing chains from the 5′ ends, means that daughter DNA strands are synthesized in pieces (Fig. 3-15).

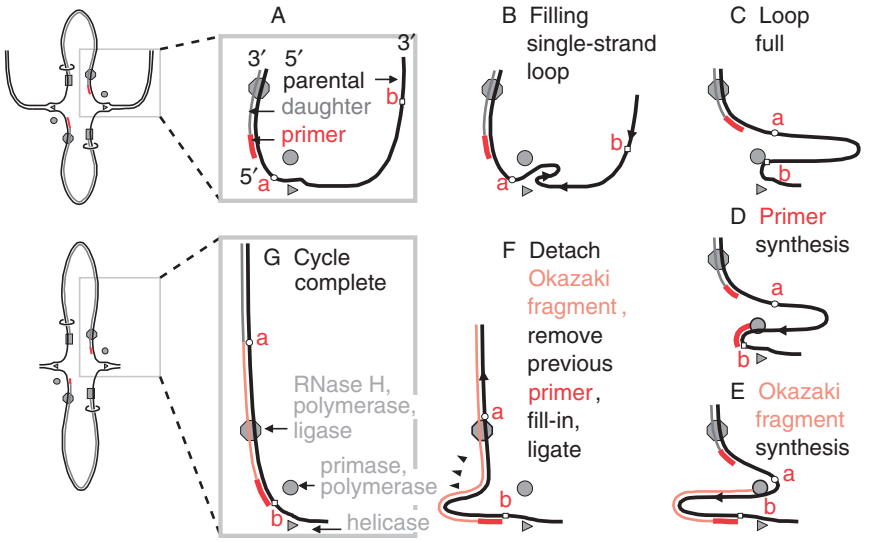


Figure 3-13.

A model for the movements of the lagging strand. We will follow how bases between a and b are copied. *Arrowheads* indicate movement. Circular clamps associated with lagging-strand polymerases are not shown for clarity. *Top left*: One polymerizing complex containing four polymerizing sites is shown (see Fig. 3-3, stage 3). (*A*) A blow-up of part of the complex. Base a has just been melted as it slid past the helicase; base b is still unmelted. A primer/Okazaki fragment made a moment before remains within the complex. (*B*) The parental strand is extruded by the helicase into a loop. (*C*) Filling the loop is now complete; base b engages the primase/polymerase site. (*D*) A primer is extruded as the template strand flows "backwards" past the primase/polymerase; the 5′ end of the primer becomes anchored to the complex. (*E*) Once the primer is complete, DNA synthesis begins; the Okazaki fragment will be complete when base a is copied. (*F*) Now the primer rearranges. The 3′ end of the completed primer/Okazaki fragment detaches from the primase/polymerase, and the 5′ end detaches from its anchorage point. Simultaneously, the new primer/Okazaki fragment and the one made previously slide through a processing site where the "older" primer is removed by RNase H, the gap between the two Okazaki fragments is filled in by a polymerase, and a ligase joins the two. (*G*) The cycle is now complete. *Bottom left*: A view of the now enlarged chromatin loops. The cycle is ready to repeat.

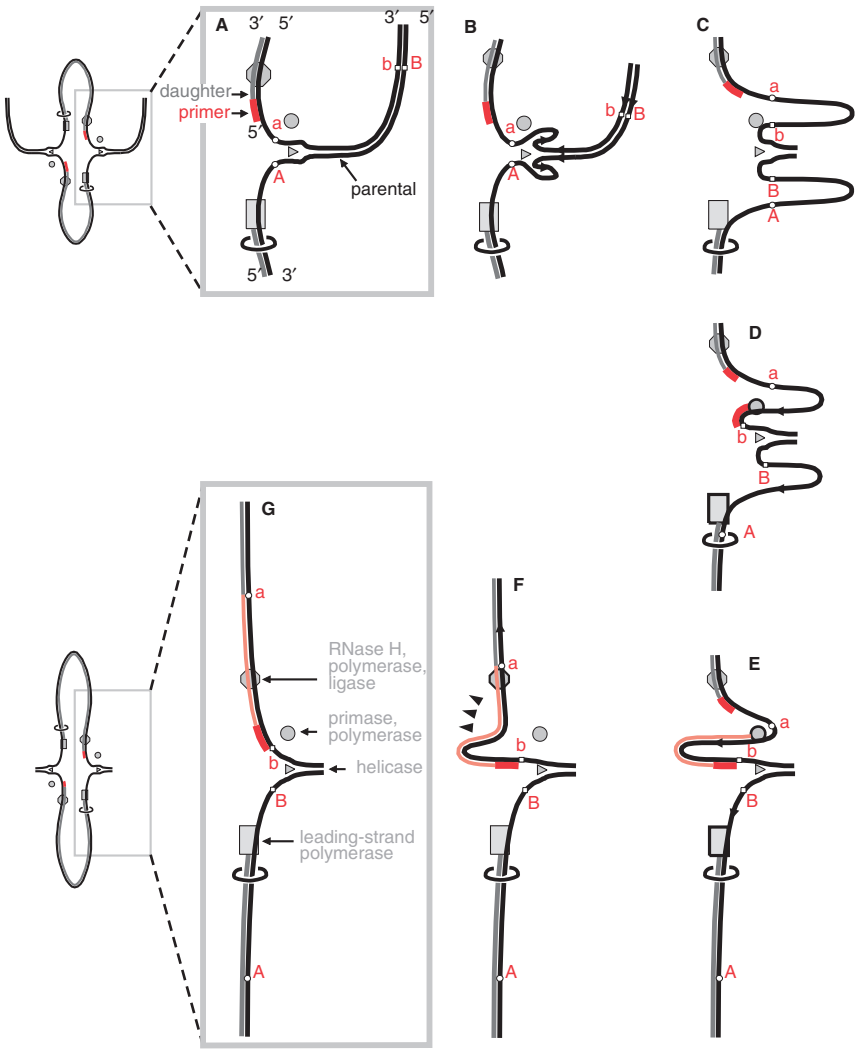


Figure 3-14.

A model for the coordinated synthesis of leading and lagging strands. *Top left*: A tetrameric polymerizing complex (like the one illustrated in Fig. 3-3, stage 3), is shown; *panel A* illustrates a blow-up of one replication fork. We will follow how a segment of DNA lying between A•a and B•b is copied. *Arrowheads* indicate movement. Circular clamps associated with lagging-strand polymerases are not shown for clarity. Continuous (leading-strand) synthesis is straightforward (*bottom half of each panel*). (*A–C*) The parental duplex is melted as it flows into the complex to create a single-stranded loop. (*D, E*) Base A is reeled in through the leading-strand polymerase/PCNA complex as a nascent strand is extruded. Lagging-strand synthesis is more complicated (*top of each panel*). (*A*) Base a has just been melted as it slid past the helicase; base b is still unmelted. A primer/Okazaki fragment made a moment before remains within the complex. (*B, C*) The parental strand is extruded by the helicase into a loop, and base b engages the primase/polymerase site in the lagging-strand polymerase. (*D*) A primer is extruded as the template strand flows "backwards" past the primase/polymerase; the 5′ end of the primer becomes anchored to the complex. A complex rearrangement and processing now occur. Both the 3′ and 5′ end of the completed primer/Okazaki fragment detach, and the new primer/Okazaki fragment and the one made previously slide through a processing site where the "older" primer is removed by RNase H, the gap between the two Okazaki fragments is filled in by a polymerase, and a ligase joins the two. (*E*) Once the primer is complete, DNA synthesis begins; the Okazaki fragment will be complete when base a is copied. (*F*) The cycle is now complete. *Bottom left*: A view of the now-enlarged loops. The cycle is ready to repeat.

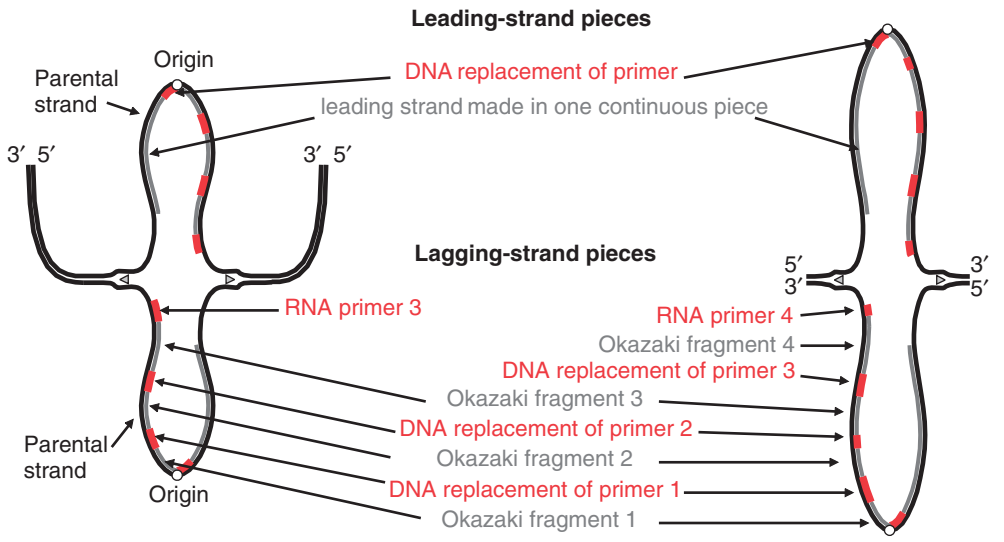


Figure 3-15.
Both leading and lagging strands are made in many different pieces ligated together to give complete daughter strands. Duplicated DNA resulting from the activity of one polymerizing complex (with four polymerizing sites) is shown, with the history of some different segments in the symmetrical structure indicated.

Only some different enzymic activities required during DNA synthesis have been discussed here. They include:

- topoisomerases,
- DNA helicases that melt the double helix to provide the single-stranded region that can be used by polymerases,
- single-strand DNA-binding proteins, or helix-destabilizing proteins, that prevent bases from pairing again,
- proteins that help assembly of the circular clamp,
- circular clamps (e.g., PCNA) and clamp-loaders,
- primases,
- leading- and lagging-strand polymerases,
- RNase H activities that remove RNA primers,
- polymerases that fill in the resulting gaps, and
- ligases that stitch together Okazaki fragments.

Polymerization occurs at four major sites in each polymerizing machine, plus two additional sites where gaps between Okazaki fragments are filled in.

Proofreading
(Echols and Goodman, 1991)

Mistakes happen; an A does not invariably pair with a T, nor a G with a C. As a result, bases mispair at a frequency of ~1 in 10^4, perhaps with lethal consequences. For example, a rare tautomeric form of C can pair with A, leading to its (mis)incorporation. Not

surprisingly, an efficient proofreading enzyme exists that clips off mispaired bases and reduces the error rate to ~1 in 10^9. This enzyme is often found as an extra catalytic domain or subunit in the polymerase; its exonuclease works from the 3′ end (i.e., in 3′→5′ direction; Fig. 3-16).

The overwhelming need for accuracy perhaps probably explains why no polymerase able to replicate in the 3′→5′ direction evolved (page 128); such an enzyme would be unable to incorporate a hydrolytic proofreading activity of this type. This follows because the growing 5′ end would then carry the activating triphosphate, and any proofreading activity would remove a mispaired base to leave a bare 5′ end; because this end lacks an activating triphosphate, chain growth would inevitably cease.

The main replicative enzyme—polymerase δ—possesses this intrinsic proofreading activity, but polymerase α does not. Polymerase α forms a complex with the primase that initiates synthesis at the origin, and at each Okazaki fragment where it makes a short RNA primer attached to DNA. Therefore, a considerable amount of replicative synthesis is potentially error-prone. However, fidelity is improved by the combined action of RNase H1 and FEN1. The former digests away the RNA primer, whereas the latter removes the last ribonucleotide and edits a faulty 5′-end. If polymerase α misreads the template and generates a mismatch close to the 5′ end of the Okazaki fragment, the decreased stability of the product allows FEN1 to excise the short 5′ terminal sequence up to the mismatch. Completion of synthesis of the preceding Okazaki fragment by polymerase δ will then fill in the gaps left by primer removal and mismatch repair. This strategy achieves high-fidelity synthesis without the need for a 3′ exonuclease activity.

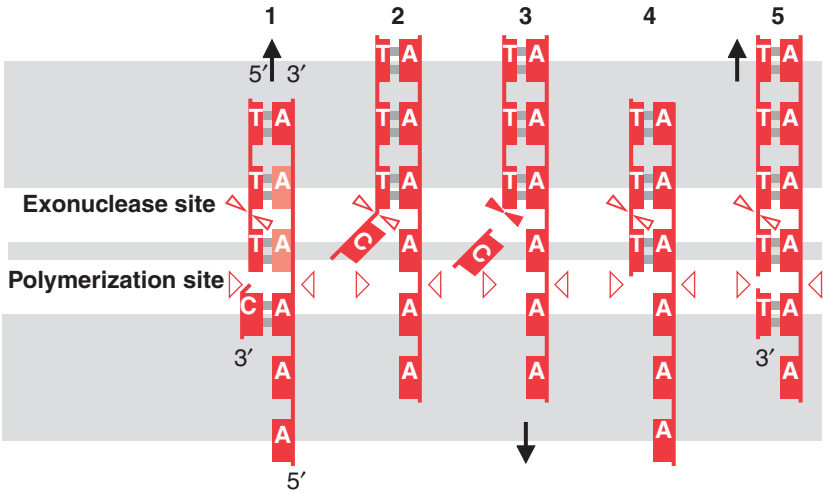


Figure 3-16.

Proofreading during the replication of a strand containing a run of As. (1) An incoming dCTP mispairs with an A in the template strand as it slides upwards through the polymerization site. (2) The exonuclease site senses the misincorporated and unpaired dCMP. (3) The dCMP has been excised; now the nascent and template strands both slide backwards. (4) The 3′ end of the nascent chain has reached the polymerization site. (5) A correctly paired dTTP is about to be added, as template and nascent strands slide upwards. Error rates can be determined from the amount of [^{32}P]dCTP (mis)incorporated during the replication of a dA•dT template.

Replicating Chromatin
(Krude, 1999)

There has been much speculation about what happens to preexisting nucleosomes during replication. Some possibilities are:

- An RNA polymerase has enough pulling power to strip DNA off a nucleosome as it reels in the template (page ___), so a DNA polymerase might also release the naked template that it is just about to replicate.
- Nucleosomes might remain bound, perhaps splitting into half nucleosomes to increase access to the replication machinery (Fig. 6-8).
- A chaperone might "hand over" a nucleosome from a parental to a daughter duplex.
- Preexisting nucleosomes on the parental duplex could segregate together (i.e., "conservatively") to only one of the daughters, while new ones were deposited on the other. Alternatively, "old" nucleosomes could segregate randomly (i.e., "semiconservatively") between the two daughters.

Preliminary results with SV40 minichromosomes—from infected cells or assembled *in vitro*—provide some pointers as to what is happening. For example, new chromatin is assembled very quickly; only ~225 and ~285 nucleotides on the leading and lagging strands respectively are sensitive to nucleases (and so free of nucleosomes; page ___), and full nucleosomes containing H1 are detected within 450–650 bp from the fork. Moreover, when minichromosomes are replicated (in the absence of additional histones) using a crude cell extract as an enzyme source, electron microscopy shows that old nucleosomes are transferred semiconservatively to both daughter duplexes. In addition, nucleosomes are not found on protein-free competitor templates, so they seem to be transferred without dissociation.

About half the weight of a chromatin template is histone, so duplication of the template requires enormous numbers of histone molecules. Each mammalian cell possesses ~20 sets of the genes for the five types of histones and—unlike most proteins synthesized throughout interphase—histones are synthesized mainly during S phase. This increased protein synthesis results from ~30-fold increase in mRNA levels; this is due to a three- to five-fold increase in transcription, and a five-fold reduction in mRNA degradation. Once histones are made, they are probably assembled first into $(H3\text{-}H4)_2$ tetramers; then H2A-H2B dimers are added, before the octamer binds to the nascent DNA and the nucleosome is assembled (page 78). A *c*hromatin-*a*ssembly *f*actor (CAF-1) composed of three subunits (i.e., p150, p60, and p48) facilitates such assembly; it is found in nuclei throughout interphase and becomes concentrated in factories during S phase. Other assembly factors (e.g., *n*ucleosome *a*ssembly *p*rotein; NAP-1) may be involved, as disrupting the genes encoding the CAF-1 subunits still gives viable yeast cells.

THE INITIATION OF SYNTHESIS
(Huberman and Riggs, 1968; DePamphilis, 1999)

One critical step that controls the replication rate is the initial one where the template binds to the polymerization site. In bacteria, the same short sequence in the chromosome called an **origin of replication** is the first to bind (Box 3-3). This sequence is the

BOX 3-3.
The Origin of Replication of *E. coli*

The genome of *E. coli* (~3.5 Mbp) contains a single origin—*oriC*. This region is defined as an origin because:

- it is essential (deletions are lethal),
- it is the first region to be replicated after returning *dnaA* ts mutants (which are blocked at initiation but can elongate normally) to the permissive temperature, and
- autonomous replication can be restored to plasmids lacking their region responsible for autonomous replication by inserting *oriC*.

The minimal region required for origin function is 245 bp. Five regions of 15–20 bp within it are highly conserved in enteric bacteria, and most mutations in these regions are not tolerated. Largely random sequences of fixed length lie between these regions. *OriC* contains motifs that—in combination—are essential for function, including:

- four 9-mers containing the specific recognition sequence [5′-TTAT(C/A)CA(C/A)] for the initiator protein, dnaA,
- three 13-mers that melt easily,
- 11 potential sites (i.e., GATC) of methylation by the Dam methylase,
- two back-to-back promoters that may be involved in the initiation of replication.

A cell-free system has been established that initiates synthesis in superhelical circular DNA containing *oriC*. Synthesis depends on various proteins including RNA polymerase, DNA gyrase (a type II topoisomerase; Box 3-2) and dnaA protein. DnaA is a central player in the initiation of chromosomal duplication as mutants cannot initiate but can elongate normally, and dnaA binds cooperatively to the 9-bp dnaA boxes in *oriC* (promoting template melting and loading of dnaB helicase).

(Sugimoto *et al*., 1979)

final target for several control circuits that sense whether conditions are suitable for rapid growth. If conditions are favorable, binding is followed by the progressive replication of the whole genome; the two parental duplexes slide into each side of the polymerization complex and the four polymerization sites each make a leading or lagging strand at ~500 nucleotides/s. As a result, circular genomes of ~4 Mbp can be replicated in ~40 min.

A typical human chromosome contains ~100 Mbp and it would take one DNA polymerase ~600 h to copy one strand, polymerizing ~50 nucleotides/s. As the S phase of many human cells lasts only ~8 h, it is clear that a chromosome could not be replicated

from one origin. This was proved using an elegant method—**DNA fiber autoradiography**—devised in the 1960s. It depends on the enormous length of DNA molecules. Human cells were grown for 5 or 10 min in [^{3}H]thymidine; then, a few were placed on a microscope slide and lysed with a strong detergent. The slide was held at an angle so that as individual strands of DNA were released from each cell they were stretched over the surface of the slide. (Fully stretched DNA of a "typical" chromosome is ~5 cm long.) After coating the slide with photographic emulsion, the pattern of labeled DNA was determined by autoradiography (Box 2-4). Tracks of silver grains that overlie the nascent DNA were clearly seen by light microscopy though individual DNA molecules were not visible (Fig. 3-17). The grain tracks were discontinuous, often appearing in pairs of equal length. The characteristic length and spacing of the tracks suggested that:

- replication was bidirectional,
- replication took place from origins spread every ~50 μm along the chromosome (i.e., the center-to-center distance between pairs of tracks),
- replication forks moved at ~50 nucleotides/s (pulse length increases track length), and
- adjacent origins tended to initiate simultaneously (all tracks in one line have roughly the same length).

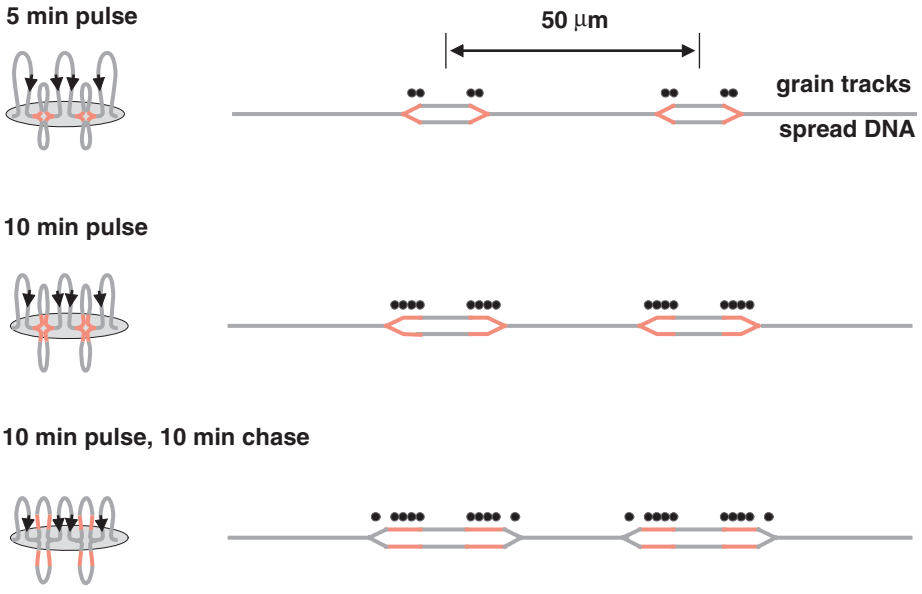


Figure 3-17.
DNA fiber autoradiography. *Left*: Two polymerizing sites in one factory are shown. Looped duplex DNA (*gray lines*) is reeled in from each side (*arrows*) into the sites as nascent DNA (*pink lines*) is extruded from each site into loops. *Right*: If cells are incubated with [^{3}H]thymidine for 5 or 10 min, radiolabel is incorporated into DNA and can be detected (after spreading DNA on a slide) as grain tracks in an autoradiograph. If the 10-min incubation is followed by a 10-min chase without [^{3}H]thymidine, radiolabel becomes diluted and the grain intensity tails off at the ends of the tracks.

These conclusions were confirmed using a 10-min pulse of [^{3}H]thymidine followed by a 10-min chase in unlabeled medium. During the chase, incorporation of label into nascent DNA tails off, and so the grain density at the ends of each pair of tracks flanking an origin diminishes.

A "typical" human chromosome of 100 Mbp contains ~2000 origins distributed every ~50 kbp. The region of DNA between one origin and the next would take ~1 h to replicate, so the whole chromosome can easily be replicated in ~8 h. As the distance between origins roughly equals the contour length of a chromatin loop (page 85), there is probably one origin per loop.

Because the origin plays such an important role in controlling bacterial replication, equivalent sequences have been sought in higher eukaryotes. But despite the discovery of origins in the monkey virus, SV40, and the yeast *S. cerevisiae*, it seems that replication in mammals begins in broad zones covering several kbp, rather than at discrete origins (Box 3-4).

BOX 3-4.
Two Methods for Mapping Origins

Various methods are now available for mapping origins. One involves two-dimensional gel electrophoresis (Box 1-5). DNA is cut with a restriction enzyme; nonreplicating DNA is cut into linear fragments, but replicating DNA is cut into various nonlinear forms (e.g., Y-shaped forms) that have odd electrophoretic mobilities (Fig. 3-18*A*). Fragments are first separated by virtue of mass, and then according to shape (using a field at right angles to the first). Linear fragments of different sizes lie along a diagonal (shown as a thick line in the bottom panel of Fig. 3-18*A*), but Y-shaped forms (resulting from a replication fork) and fragments containing replication bubbles (from the origin) move aberrantly and are offset from the diagonal. Finally, specific DNA sequences are localized after Southern blotting and hybridization with a labeled probe (Box 1-5). Then, maps of the origin and its surroundings can be built using different probes along the chromosome. Unfortunately, and because so few DNA molecules are replicating at any one time, this method is insensitive; however, it has been used to confirm that yeast ARSs are indeed origins.

Another approach uses the *p*olymerase *c*hain *r*eaction (PCR; Fig. 3-18*B*), and the exquisite sensitivity of this technique allows the method to be applied to higher eukaryotes.

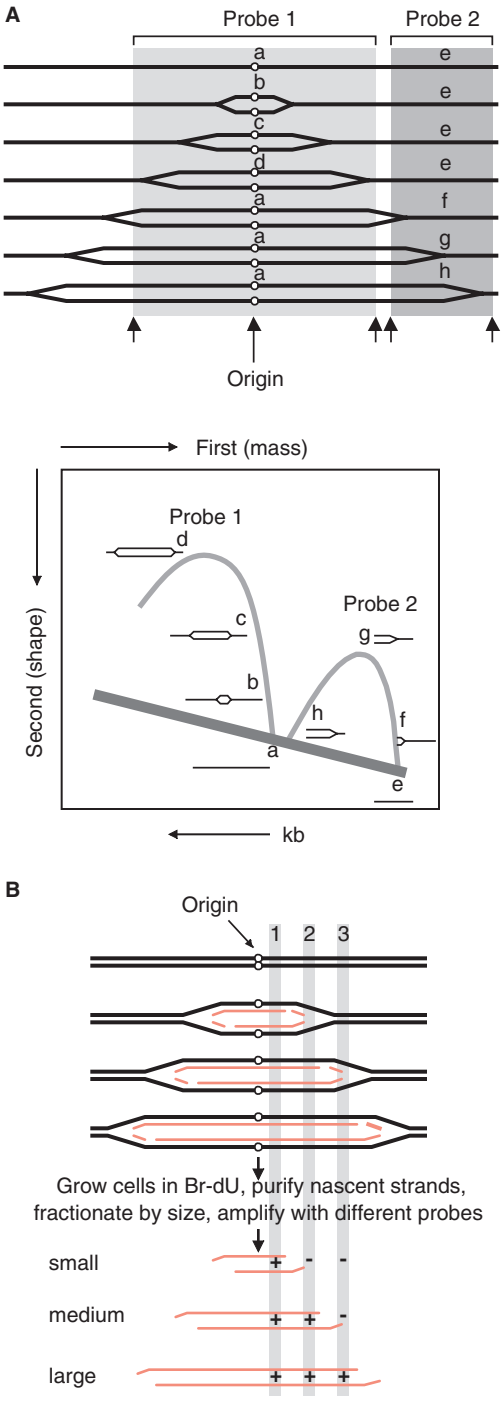
(Brewer and Fangman, 1987; Vassilev and Johnson, 1989)

Simple Origins of SV40 Virus and Yeast
(Borowiec *et al.*, 1990; Bell and Stillman, 1992; Marahrens and Stillman, 1996)

The genome of SV40 virus is only ~5 kbp long; it encodes three viral coat proteins, plus the large and small *t*umor (T) antigens. The simplicity of the genome means that the virus parasitizes its host, with the multifunctional T antigen playing an essential role in subverting normal cellular controls. (T antigen probably combines functions of the bac-

Figure 3-18.

Mapping origins (*A*) Using two-dimensional gels. *Top*: Seven stages during the replication of one region of a chromosome are shown after deproteinization and disruption of chromatin loops. Each line represents a duplex, and all stages are present in an unsynchronized population. Regions around the origin are duplicated before those at the ends. After cutting with a restriction enzyme (*arrows*), fragments (labelled *a–h*) are resolved by two-dimensional electrophoresis, blotted, hybridized with probes complementary to regions shown (Box 1-5), and an autoradiograph prepared (Box 2-4). *Bottom*: A diagram illustrating the position of different fragments in the gel. Linear fragments of different sizes lie along the *thicker line*; most do not hybridize with either probe. Some replication intermediates that hybridize with probes 1 and 2 lie on one of the *two curves*, and the positions of the specific intermediates are shown. Repeating the experiment with other probes hybridizing with different regions along the chromosome allows precise mapping. (*B*) Using PCR. Four different stages in the replication of one region of a chromosome are shown (after deproteinization and linearization); parental and nascent strands are represented by *thick* and *thin pink lines*, respectively. Cells are grown in Br-dU so that the nascent chains can be purified by virtue of their different density on gradients containing caesium chloride; then the chains are fractionated according to length, and the different fractions amplified by PCR using three different pairs of primers that flank three different regions (i.e., regions 1, 2, and 3). Primer pairs hybridizing to region 1—which is nearest the origin—will amplify sequences in all size classes of nascent DNA (shown by +), whereas those from further away will only work with longer chains.



terial dnaA protein with those of a helicase.) The origin was first defined genetically—mutants were replication deficient. Deletion mapping showed it to be ~450 bp, with an essential 64 bp core. It contains:

- four copies of the pentamer GAGGC organized as an inverted repeat—the prime binding site for the dodecameric T antigen,
- a region of 17 AT base pairs likely to be the site where the duplex first melts,
- an imperfect 15 bp palindrome,
- signals for the initiation of transcription of "early" and "late" viral genes.

The definition of the first origin in a eukaryotic chromosome depended on the powerful selection methods available in yeast (see also page 99). Yeast mutants lacking the *His* gene are unable to form colonies on plates lacking histidine. Even on transformation with a bacterial plasmid carrying the yeast *His* gene, few colonies result; this is because the bacterial plasmid is unable to replicate along with the yeast chromosomes and is soon diluted out. However, if random pieces of yeast DNA are inserted into the plasmid, a few will contain an origin and so can replicate in yeast cells. Cells carrying such a plasmid will grow into a colony because they contain both the *His* gene and an origin that allows the plasmid to replicate. DNA sequences identified in this way are called ***a*utonomously *r*eplicating *s*equences** (ARSs), and two-dimensional gel electrophoresis (Box 1-5) confirms that replication initiates within them (Box 3-4).

Molecular dissection of several yeast ARSs uncovered an essential 11-nucleotide "core consensus sequence" (the A element), plus three additional elements (B1, B2, and B3) that enhance ARS function. A multiprotein complex—the ***o*rigin *r*ecognition *c*omplex** (ORC)—is probably the eukaryotic equivalent of the bacterial DnaA protein; it binds to the A and B1 elements and protects them from attack by DNase 1. B3 is a binding site for the DNA-binding protein ABF1 that can be substituted with the binding sites of other transcription factors. The six polypeptides in ORC are encoded by *ORC1–6*. Temperature-sensitive mutants of the *ORC2* and *ORC5* genes are defective in replication at the restrictive temperature, with fewer replication bubbles in *ARS1* being detectable by two-dimensional gel electrophoresis (Box 1-5; Fig. 3-18). Genomic footprinting (Box 1-6) using DNase I shows that ORC is bound to *ARS1* throughout S, G2 and M phase. Late in mitosis, the footprint is enlarged to define a prereplication complex (page ___).

Despite many attempts, no ARS analogs have yet been discovered that enable plasmids to replicate in mammalian cells. However, several ORC homologs have been characterized in higher eukaryotes, and they clearly play a role in initiation. For example, ORC1 or ORC2 can be removed from *Xenopus* egg extracts using antibodies, and then the extract is unable to support the initiation of replication (page 36).

Complex Mammalian Origins

(Harland and Laskey, 1980; Craig and Bickmore, 1993; Jackson and Pombo, 1998; Dimitrova and Gilbert, 1999)

Despite the important role that origins play during initiation in bacteria and yeast, replication in higher eukaryotes can clearly occur without specific origins. When DNA of bacteriophage λ—which lacks any sequences required for replication in eukaryotes—is

added to an extract made from frogs' eggs, it is assembled into a pseudo-nucleus, replication factories form, and DNA is replicated semiconservatively (see also page 36). Here, some phage sequence must serve as an origin.

Are specific sequences used as origins in most mammalian cells? It seems there is considerable flexibility over the choice of origin; choice is not as tightly controlled as in SV40 virus, but nor is it as loosely controlled as in *Xenopus* eggs. For example, an origin has been mapped (Box 3-4) in the ~135 kbp of the human β globin locus; replication tends to initiate in ~2 kbp between the δ and β globin genes. If this region is deleted—for example, in patients suffering from β-thalassemia—replication initiates outside the locus.

In mammals, it takes ~1 h to replicate DNA between adjacent origins. As S phase is ~8 h, this means that most origins are unlikely to be active at any one time. Are they activated randomly, or do they fire in a specific sequence? A specific order has been shown by measuring (using Southern blotting; Box 1-5) the copy number of several genes as synchronized cells progress around the cell cycle (Box 7-1); in a fibroblast, "housekeeping" genes expressed in all cells (e.g., nonmuscle actin) tend to be duplicated before genes active only in other specialized cells (e.g., globin).

Specificity can also be shown using Br-dU. (An analogous approach is shown in Figure 3-7.) Cells at different stages of S phase are incubated for 15 min with Br-dU, the cells harvested at the subsequent mitosis, and chromosomes prepared; then regions containing incorporated bromine can be visualized by its altered staining properties, or after immunolabeling with an anti-BrDNA antibody. Results show that different regions of each chromosome are replicated at a particular time in S phase. The incorporation of [^{3}H]thymidine followed by autoradiography of chromosomes at the subsequent metaphase gives similar results, but with poorer resolution (Fig. 1-2; Box 2-4).

This general approach has been extended to show that the structure organizing a chromatin "cloud" (page 92) persists from one cycle to the next, and that regions that fire early in one cycle fire early in a later one. HeLa cells were synchronized at the beginning of S phase just after early-replicating origins had fired (using successive thymidine, nocodazole, and aphidicolin blocks; Box 7-1). Then, cells were pulse-labeled for 20 min with Br-dU, before the label was removed and the cells allowed to grow and divide for 4 d. Now, cells were resynchronized and pulse-labeled as before, but this time using I-dU. Again, cells were washed and regrown, but only until they reached the next G1 phase. Finally, any Br-DNA and I-DNA were immunolabeled with different colors. Confocal microscopy revealed that I-DNA was concentrated in small foci reminiscent of those found by direct labeling at the beginning of S-phase (e.g., Fig. 3-7*A*); early-replicating chromatin loops must remain together in their "clouds" from the last S-phase to the last G1 phase. Br-DNA was found in fewer foci clustered in a several chromosome "territories" (page 94); these early-replicating loops also remained together in their clouds through successive divisions, but fewer foci were seen because some were lost when Br-DNA strands segregated semiconservatively during those divisions (Fig. 3-2). Significantly, foci containing Br-DNA also contained I-DNA, showing that clouds (and origins) that fired early in one cycle did so again several cycles later. This conclusion was confirmed at higher resolution using spread DNA fibers (see also Fig. 3-17). Cells were labeled with Br-dU between 5 and 20 min from the beginning of one S-phase, and with I-dU between 10 and 30 min in the next. In doubly labeled fibers, >95% (unlabeled) origins were flanked by inner regions containing Br-DNA and outer regions containing I-DNA. At the level of resolution obtained (i.e., to within a few kbp), most origins are used during successive cycles.

The order of replication of different chromatin clouds is largely determined by structure; centromeric heterochromatin and the inactive X chromosome in a Barr body are both replicated late during S phase (page ___). Moreover, most R bands (which contain many "housekeeping" genes) replicate early, whereas G bands (which contain few genes) replicate late (page 102). This order is defined early in G1 phase. When nuclei isolated from Chinese hamster cells late in G1 phase are incubated with biotin-dUTP in extracts of *Xenopus* eggs, biotin is incorporated into clouds in the correct temporal sequence (see Fig. 3-7, page 36). This order is lost with nuclei isolated earlier in G1 phase; now clouds that replicated late during the previous cell cycle (and so can be labeled with IdU) sometimes replicate early in the extract, and *vice versa*. Some event occurring 1–2 h after mitosis establishes what the temporal sequence will be.

Role of Transcription During Initiation

(Hassan *et al.*, 1994)

A primase synthesizes the short piece of RNA used as a primer for the synthesis of both leading and lagging strands; therefore, RNA synthesis is required for replication. However, there is an additional and confusing relationship between transcription and origin firing. First, RNA polymerase is required *in vitro* to replicate bacterial templates containing *oriC* under certain conditions. Moreover, eukaryotic origins are also always closely associated with transcription units (e.g., the SV40 origin), and they contain binding sites for transcription factors (e.g., the ABF1 site in yeast *ARS1*). However, it is also clear that yeast origins generally lie between transcription units, and that incorporation of an origin into a transcription unit prevents it from firing. Second, some mammalian cell-cycle mutants that arrest late during G1 can be rescued by microinjection of RNA polymerase II, or by transfection with a cDNA encoding a polymerase subunit or a transcription factor. Third, transcriptionally active regions replicate early during S phase (see above). Fourth, double-labeling experiments show that replication begins close to transcription factories at the beginning of S phase (Fig. 3-19). Subsequently, the small replication factories grow and apparently move about the nucleus (Fig. 3-7).

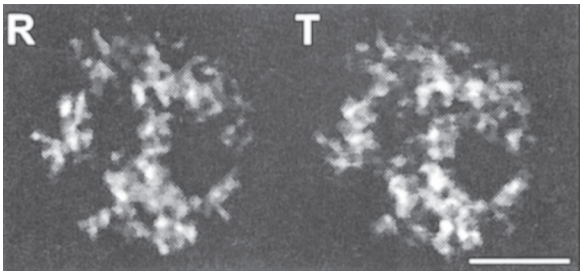


Figure 3-19.

Sites of replication and transcription lie close together at the beginning of S phase. HeLa cells that had just entered S phase were permeabilized, and allowed to make DNA and RNA in the presence of biotin-dUTP and Br-UTP; subsequently sites containing biotin-DNA (i.e., replication sites; R) and Br-RNA (i.e., transcription sites; T) were immunolabeled with red and green fluorochromes, respectively. The two views (collected using a confocal microscope and red or green filters) of one nucleus show that replication sites lie close to transcription sites. Bar: 5 μm. [Reprinted, with permission, from A.B. Hassan *et al.* (J. Cell Sci. *107*, 425–434); copyright © 1994 Company of Biologists Ltd.]

These observations suggest that transcription factories nucleate the formation of replication factories. Two extreme models can be envisaged. In one (Fig. 3-20, *left*), small replication factories (1–6) quickly assemble at the beginning of S phase around transcription factories and immediately become active. As one factory replicates many loops, not all origins need fire simultaneously. By mid S phase, some small replication factories (e.g., 1) would become redundant; after disassembly, their components would be incorporated into medium-sized factories that arise by growth and fusion of smaller factories (e.g., 2, 3). By late S phase, disassembly of most factories and growth/fusion of a few would generate large factories (i.e., 4, 5, 6) which replicate most heterochromatin; they will always contain nascent DNA as they grow from smaller active factories. Factories might apparently "move" along the chromosome as now-redundant components at one (inactive) end become soluble to be incorporated into the other (active) end. In a second model, factories are built *ab initio* at new sites on the chromatin fiber at different stages during S phase (Fig. 3-20, *right*). For example, if some factor was in short supply, there might only be sufficient to build the small factories A1, 2, and 3; then, once they had finished replicating the surrounding loops, their components might be re-used to build factory B, and then factory C.

In both models, transcription plays critical roles in determining when a DNA sequence is replicated. Transcription factories nucleate the formation of the needed replication complexes, and they determine the organization of the chromatin fiber into clouds (page 92) and so the proximity of a particular DNA sequence to those replication complexes. Then, which DNA sequence is used as an origin will depend on its affinity for the polymerizing machinery; an easily melting region close to an active transcription unit (e.g., the origin of SV40 virus) will be chosen more often than a distant inactive one. But where the pressure to replicate the genome becomes high enough (e.g., at the end of S phase in a mammalian cell), any sequence will be pressed into service. Under certain circumstances, transcription may even play no role. After fertilization, a *Xenopus* egg divides rapidly 12 times to give ~4000 cells; then, at the midblastula transition, the rate of division slows. Very little RNA synthesis—measured by incorporation of [^{3}H]uridine (by whole cells) or [^{32}P]UTP (after microinjection)—occurs before the midblastula transition, implying that DNA can be duplicated without transcription. Here, the pressure to replicate is paramount.

REPLICATING ENDS

(Hozák *et al.*, 1996; Nugent and Lundblad, 1998)

Special problems are posed by the replication of DNA that lies between two units of replication. Once most of a parental loop has slid through one or other of the polymerizing machines on each side, some parental duplex remains unreplicated (e.g., Fig. 3-3, stage 5). A possible solution to this problem is illustrated in Figure 3-21.

Replicating the very end of a eukaryotic chromosome—the telomere (page 101)—also requires specialized machinery (Fig. 3-22). DNA polymerases replicate 5′→3′, cannot initiate a DNA chain *de novo*, and use an RNA primer that is then removed. As a result, they can copy one strand to the very end, but not the other as a gap will inevitably be left where the primer was removed. Because this gap cannot be filled in by a conventional polymerase, the DNA end will be incompletely replicated so that after a series of cell divisions chromosomes will become progressively shorter. Such shortening can be shown directly in yeast by deleting the telomeric sequence from one end of a dispensable chromosome.

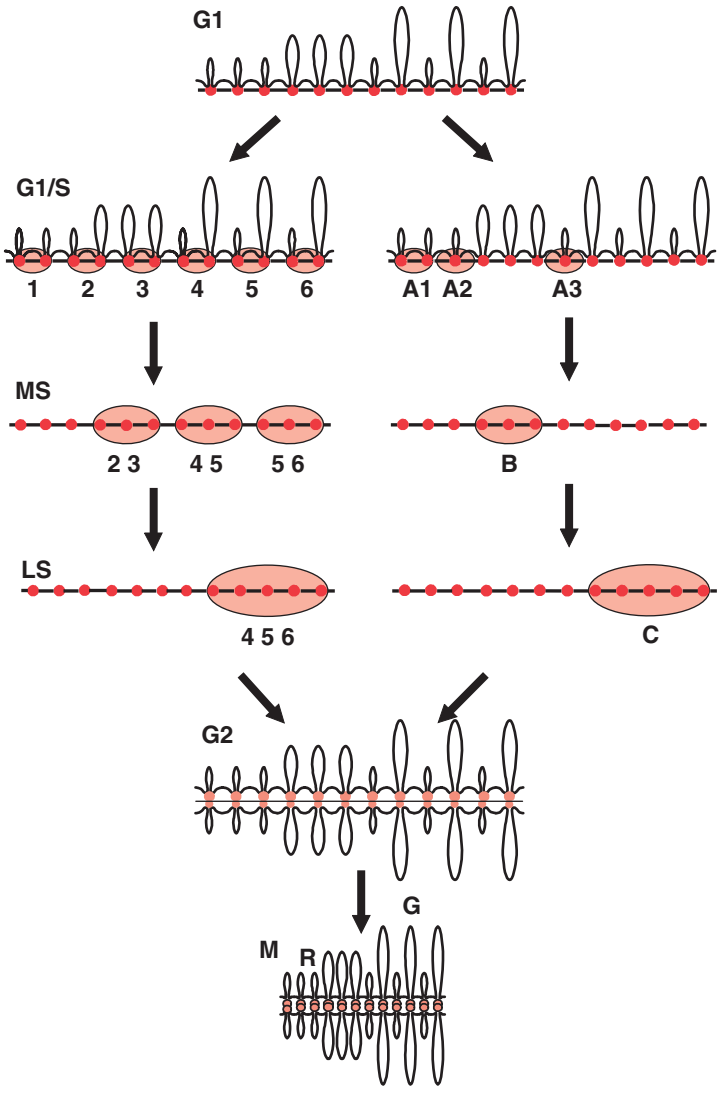


Figure 3-20.

Transcription factories may nucleate the assembly of replication factories. The G1 duplex is shown looped by attachment to transcription factories (*circles*) attached to a nucleoskeleton (*horizontal line*). Small and large loops represent eu- and hetero-chromatin. For the sake of clarity, only one of the ~16 loops attached to each transcription factory during G1, G1/S and G2—and none of those during mid and late S-phase (MS and LS)—are shown. Replication factories (ovals) are assembled around transcription factories and then most DNA is synthesized as templates slide through the factories, although some "tidying-up" replication occurs outside factories to give the duplicated G2 fiber. The left- and right-hand pathways illustrate two extreme models of factory evolution; large factories might grow from small factories (*left*) or be built *ab initio* at new sites on the chromatin fiber at different stages during S phase (*right*). On entry into mitosis (*M*), the skeleton disassembles (without changing the contour length of loops) and residual transcription factories (plus associated eu- and hetero-chromatin) collapse on to the chromosome axis to generate R and G bands. [Reprinted from P. Hozák *et al.* (in *Eukaryotic DNA Replication*, ed J.J. Blow); copyright © 1996, by permission of Oxford University Press.]

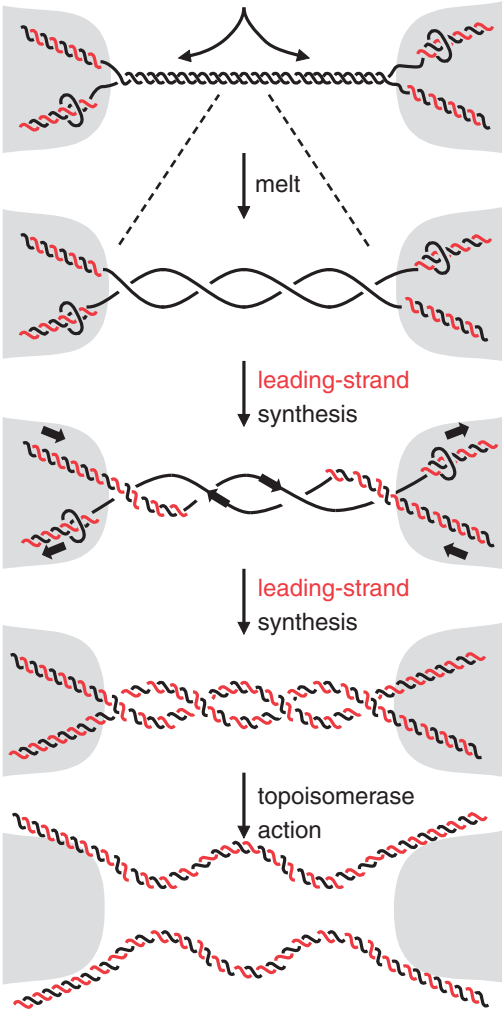


Figure 3-21.

A model for the replication of the last piece of parental DNA between two origins. A loop of DNA between two replication machines has been almost completely pulled through (*arrows*) the flanking polymerization sites (*shaded areas*); ~120 bp between the two forks remain unreplicated. For convenience, we will follow what happens to only ~20 bp. First, parental strands are melted; daughter chromosomes with 10–30 interlocks like this are found late during the replication of SV40 virus. Leading-strand synthesis then occurs on both sides (circular clamps are indicated), as the melted parental strands are pulled in opposite directions (*arrows*) through the two complexes. After complete replication, and ligation of ends (not shown), two interlocked daughters result. The interlocks are then resolved by topoisomerases (Box 3-2). An analogous model involving lagging-strand polymerases is equally possible.

Various mechanisms have evolved to solve this problem of replicating the very ends of DNA molecules. The human adenovirus uses protein-nucleotide priming; a protein bound at the end of the genome primes the synthesis of the first nucleotide. Another solution is to have a sequence at the end that allows the formation of a hairpin, concatamer, or circle (e.g., in vaccinia, T7, and lambdoid viruses); rearrangement of the

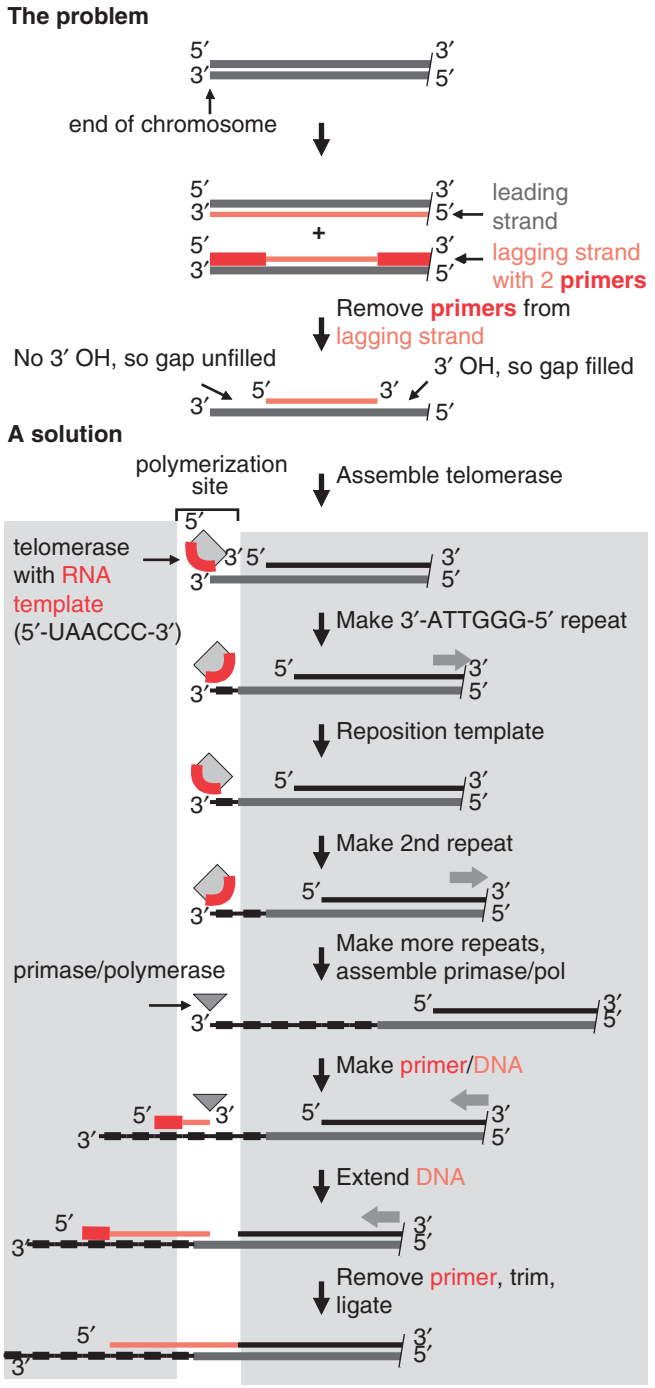


Figure 3-22.

Replicating a chromosome end. The problem: The left-hand end of a chromosome is shown (for simplicity, the end is shown without a 3′ overhang; page 101). The leading strand can be extended to the very end, but removal of the primer at the 5′ end of the lagging strand leaves a gap that cannot be filled, as no 3′OH is available. A solution: it is assumed that the chromosome initially has a 3′ overhang; this might result partly from removal of the last primer (as in the panel immediately above), and partly from a preexisting 3′ overhang (page 101). It is also assumed that all polymerizing sites are fixed (*large arrows* show template movements), the template in telomerase repositions before the addition of each telomeric repeat, and that conventional lagging-strand synthesis by a primase/polymerase is involved after telomerase action. Here, the daughter chromosome is longer than its parent.

hairpin, etc. then eliminates the need for replication of the "last" nucleotide. Alternatively, the ends of a linear molecule can be maintained by recombination (e.g., in T4 bacteriophage). In many organisms, telomere replication involves a specialized reverse transcriptase called **telomerase**.

Telomerase activity was first detected in *Tetrahymena* extracts at a stage when telomeres were forming (page 101). Telomeres contain tandem DNA repeats, with one G-rich strand and one C-rich strand; the G-strand has the repeat $(5'\text{-GGGGTT})_n$ running from the center of the molecule to the end. In man, the G-strand repeat is $(5'\text{-GGGTTA})_n$, again running from center to end. The repeat length varies from cell to cell, as restriction fragments bearing chromosomal ends (detected by Southern blotting using a telomeric probe; Box 1-5) have a range of sizes. Telomerase was discovered by its ability to extend the G-rich strand without a DNA template. Unusually for an enzyme, it is part protein and part RNA. The protein part has homology with reverse transcriptases, while the RNA part contains 8–30 nucleotides containing 1.2–1.9 copies of the C-strand repeat. For example, human telomerase contains the RNA sequence 5′-C**UAACCC**UAAC; the bases shown in bold are copied into G-strand DNA to give $(5'\text{-GGGTTA})_n$. Therefore, mutating the gene encoding this RNA alters the sequences that become incorporated into telomeric DNA. As a result, telomerase avoids the problem of end replication by using RNA, not DNA, to template the synthesis of telomeric DNA. It is unusual in that information within the enzyme itself determines the sequence of the product.

The details of how telomerase acts are obscure; the model in Figure 3-22 illustrates one possible mechanism. Here, telomerase binds through its RNA template to the parental strand, catalyzing the synthesis of new telomeric subunits that extend the parental strand. As this process is not tightly controlled, different numbers of subunits are added to different chromosomes, so that chromosomal length varies from cell to cell, even in a clonal population. The resulting length variations may play important roles in aging and cancer (Box 3-5). Once telomerase has made its copies, the gap at the 5′ end of the lagging strand is filled in by the conventional machinery (i.e., primase/polymerase, RNase H, and a ligase).

BOX 3-5.
Telomerase in Aging and Cancer

Telomerase deficiency may play an important role in aging and cancer. Human germ-line cells express telomerase and have telomeric repeats that are 15–20 kbp long. During development, telomerase activity declines, and telomeres of most somatic cells become shorter. This shortening can also be seen when diploid cells are explanted *in vitro*; after an initial period of growth, the cells senesce and stop growing. A minority of such cells become cancerous, escaping sensescence and becoming immortal. Senescence of the majority and renaissance of the few correlates with the loss and gain of telomerase. Then, it is attractive to suppose that a lack of telomerase in normal cells acts as a tumor-suppressor system. (Immortalization may also be accompanied by activation of a recombination-based system that maintains telomere length.)

The correlation between telomerase expression and a failure to age is strengthened by the following experiment. When the catalytic subunit of human telomerase (hTRT) is expressed in telomerase-negative normal cells (retinal pigment epithelial cells, foreskin fibroblasts), telomeres become longer, and clones

exceed their normal life-span by >20 doublings. However, "knocking out" both genes encoding *m*ouse *t*elomerase *R*NA (*mTR*) does not cause the resulting mice to age prematurely, even though their chromosomes lack detectable repeats at their ends. These mice are initially cancer-free, and breed successfully for at least six generations. But during generations 4–6, telomerase deficiency leads to fewer cells in rapidly dividing tissues (e.g., bone marrow, spleen, testis). This suggests that telomeres are partially maintained by a telomerase-free mechanism (e.g., through nonreciprocal recombination), but some telomerase activity is eventually required.

(Zakian, 1997; Greider, 1998)

Summary

- All organisms must replicate their DNA accurately before every cell division.
- This requires the coordinated action of many different enzymes that help DNA polymerases in doing the basic task. Base pairs in the double helix are broken at the replication fork. Then the polymerase facilitates the pairing of two adjacent bases in a template strand with their complementary dNTPs, and the formation of a phosphodiester bond linking the two; the stepwise addition of additional dNTPs after each has paired in turn with its complementary base leads to growth of a new DNA chain in the 5′-to-3′ direction.
- Synthesis is semiconservative; each daughter DNA molecule contains one original strand and one newly made strand.
- In higher eukaryotes, replication occurs only during S phase and begins at loosely defined origins spaced every ~50 kbp along the chromosome.
- Many different parts of one chromosome are duplicated simultaneously after origins attach to enormous replication factories (diameter 100-1000 nm) containing the required components. Each factory contains many polymerizing machines; each machine contains (at least) four DNA polymerases, and each polymerase is responsible for copying one of the four parental strands around an origin. Synthesis occurs as two parental duplexes are reeled into the complex by two helicases. The resulting four single strands then slide through the four polymerizing sites, as daughter duplexes are extruded into loops. Movement of parental and daughter strands is driven by the favorable free-energy change that accompanies release of pyrophosphate occurring at the helicase and polymerization sites.
- A magnetic tape is often used as an analogy for the DNA in a chromosome; the information in both is arrayed linearly, and it can be copied (replicated) accurately by the appropriate machine (i.e., a dubbing cassette player or a polymerase). The analogy can be extended to include the movements of the tape/DNA that take place during copying: the tape is reeled past the fixed head, and the DNA is copied as it is pulled through the polymerizing site. And just as very long tapes can be copied quickly by dividing them into segments and then copying all the segments simultaneously, chromosomes are divided into replicons, with different groups of replicons being copied in different factories.

- Note that the models presented here all involve fixed polymerases. They are based on recent studies that show that the active polymerases are localized within replication factories. However, older models presented in most textbooks involve individual polymerases that are free to track along the template. The analogy here would be one of the tape head tracking along the tape, which seems mechanically implausible.

REFERENCES

Baker, T.J. and Wickner, S.H. (1992). Genetics and enzymology of DNA replication in *Escherichia coli*. Annu. Rev. Genet. *26*, 447–477.

Bell, S.P. and Stillman, B. (1992). ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature *357*, 128–134.

Berger, J.M., Gamblin, S.J., Harrison, S.C. and Wang, J.C. (1996). Structure and mechanism of DNA topoisomerase II. Nature *379*, 225–232.

Borowiec, J.A., Dean, F.B., Bullock, P.A. and Hurwitz, J. (1990). Binding and unwinding—how T antigen engages the SV40 origin of DNA replication. Cell *60*, 181–184.

Brewer, B.J. and Fangman, W.L. (1987). The localization of replication origins on ARS plasmids in *S. cerevisiae*. Cell *51*, 463–471.

Cook, P.R. (1999). The organization of replication and transcription. Science *284*, 1790–1795.

Craig, J.M. and Bickmore, W.A. (1993). Chromosome bands—flavours to savour. BioEssays *15*, 349–354.

DePamphilis, M.L. (1999). Replication origins in metazoan chromosomes: fact or fiction? BioEssays *21*, 5–16.

Dimitrova, D.S. and Gilbert, D.M. (1999). The spatial position and replication timing of chromosomal domains are both established in early G1 phase. Mol. Cell *4*, 983–993.

Echols, H. and Goodman, M.F. (1991). Fidelity mechanisms in DNA replication. Ann. Rev. Biochem. *60*, 477–511.

Greider, C.W. (1998). Telomerase activity, cell proliferation, and cancer. Proc. Natl. Acad. Sci. USA *95*, 90–92.

Harland, R.M. and Laskey, R.A. (1980). Regulated replication of DNA microinjected into eggs of *Xenopus laevis*. Cell *21*, 761–771.

Hassan, A.B., Errington, R.J., White, N.S., Jackson, D.A. and Cook. P.R. (1994). Replication and transcription sites are colocalized in human cells. J. Cell Sci. *107*, 425–434.

Hozák, P., Hassan, A.B., Jackson, D.A. and Cook, P.R. (1993). Visualization of replication factories attached to a nucleoskeleton. Cell *73*, 361–373.

Hozák, P., Jackson, D.A. and Cook. P.R. (1996). The role of nuclear structure in DNA replication. In *Eukaryotic DNA Replication: Frontiers in Molecular Biology* (Ed. J.J. Blow), pp 124–142. Oxford University Press, Oxford.

Huberman, J.A. and Riggs, A.D. (1968). On the mechanism of DNA replication in mammalian chromosomes. J. Mol. Biol. *75*, 327–341.

Jackson, D.A. and Cook, P.R. (1986). Replication occurs at a nucleoskeleton. EMBO J. *5*, 1403–1410.

Jackson, D.A. and Pombo, A. (1998). Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to efficient activation and propogation of S phase in human cells. J. Cell Biol. *140*, 1285–1295.

Krishna, T.S., Kong, X.P., Gary, S., Burgers, P.M. and Kuriyan, J. (1994). Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. Cell *79*, 1233–1243.

Krude, T. (1999). Chromatin assembly during DNA replication in somatic cells. Eur. J. Biochem. *263*, 1–5.

Lemon, K.P. and Grossman, A.D. (1998). Localization of bacterial DNA polymerase: evidence for a factory model of replication. Science *282*, 1516–1519.

Marahrens, Y. and Stillman, B. (1996). The initiation of DNA replication in the yeast *Saccharomyces cerevisiae*. In *Eukaryotic DNA Replication* (Ed. Blow, J.J.), pp 66–95. Oxford University Press, Oxford.

Nakamura, H., Morita, T. and Sato, C. (1986). Structural organisation of replicon domains during DNA synthetic phase in the mammalian nucleus. Exp. Cell Res. *165*, 291–297.

Nugent, C.I. and Lundblad, V. (1998). The telomerase reverse transcriptase: components and regulation. Genes Dev. *12*, 1073–1085.

Ogawa, T. and Okazaki, T. (1980). Discontinuous DNA replication. Ann. Rev. Biochem. *49*, 421–457.

Stukenberg, P.T., Studwell-Vaughan, P.S. and O'Donnell, M. (1991). Mechanism of the sliding beta-clamp of DNA polymerase III holoenzyme. J. Biol. Chem. *266*, 11328–11334.

Sugimoto, K., Oka, A., Sugisaki, H., Takanam, M., Nishimura, A., Yasuda, Y. and Hirota, Y. (1979). Nucleotide sequence of *Escherichia coli* K-12 replication origin. Proc. Natl. Acad. Sci. USA *76*, 575–579.

Vassilev, L. and Johnson, E.M. (1989). Mapping initiation sites of DNA replication *in vivo* using polymerase chain amplification of nascent strand segments. Nucleic Acids Res. *17*, 7693–7705.

Waga, S. and Stillman, B. (1998). The DNA replication fork in eukaryotic cells. Ann. Rev. Biochem. *67*, 721–751.

Wang, J.C. (1991). DNA topoisomerases: why so many? J. Biol. Chem. *266*, 6659–6662.

Wessel, R., Schweizer, J. and Stahl, H. (1992). Simian virus 40 T-antigen DNA helicase is a hexamer which forms a binary complex during bidirectional unwinding from the viral origin of DNA replication. J. Virol. *66*, 804–815.

West, S.C. (1996). DNA helicases: new breeds of translocating motors and molecular pumps. Cell *86*, 177–180.

Wyman, C. and Botchan, M. (1995). A familiar ring to DNA polymerase processivity. Curr. Biol. *5*, 334–337.

Zakian, V.A. (1997). Life and cancer without telomerase. Cell *91*, 1–3.