Many of the most basic biochemical mechanisms of DNA replication have been conserved from prokaryotes to eukaryotes, but the evolution of eukaryotic cells resulted in many changes in the logic of the cell cycle and in the mechanisms that regulate DNA replication. In a prokaryotic cell, such as *Escherichia coli*, DNA replication is generally initiated at a single essential locus in the circular chromosome (replicator). The timing and specificity of initiation are determined by the binding of a protein complex (initiator) to specific sequences in the replicator (Jacob et al. 1964). In a rapidly growing *E. coli* cell, initiation events occur at intervals significantly shorter than the time required to complete the synthesis of daughter chromosomes. Under these circumstances, DNA synthesis is essentially continuous throughout the cell cycle, and cells contain several partially completed chromosomes. In contrast, in eukaryotic cells, DNA replication occupies only part of the cell cycle and alternates with mitosis. Moreover, each chromosome is duplicated only once in each cell cycle, producing two complete daughter chromosomes. To ensure the timely completion of the replication of the large genomes of eukaryotes, initiation of DNA synthesis occurs at multiple chromosomal sites (Huberman and Riggs 1966). The number of such origins of DNA replication ranges from a few hundred in a yeast cell to tens of thousands in a human cell. Given the complexity of genome duplication in eukaryotic cells, it is not surprising that novel regulatory mechanisms have evolved.
to coordinate DNA replication with other events in the cell cycle and to control initiation of DNA synthesis in both space and time (Kelly and Brown 2000; Bell and Dutta 2002; Machida et al. 2005). The machinery of DNA chain elongation is also considerably more elaborate than in prokaryotic cells, for reasons that are not completely clear but may be related in part to the more complex nucleoprotein structure of eukaryotic chromosomes.

WHERE TO START—REPLICATORS AND ORIGINS

Many laboratories have studied the nature of origins of replication in eukaryotic cells, but there are still numerous unanswered questions, especially in metazoans (Chapter 2). One important issue is the extent to which specific nucleotide sequence elements analogous to prokaryotic replicators determine the chromosomal sites where DNA replication is initiated. Sites of initiation are clearly distributed nonrandomly over the chromosomes, so origins of replication must ultimately be determined to some extent by local nucleotide sequence (Raghuraman et al. 2001; MacAlpine and Bell 2005). However, it appears that the importance of conserved replicator sequences differs significantly among eukaryotic species. At one end of the spectrum are the budding yeast Saccharomyces cerevisiae and its close relatives. Like prokaryotes, budding yeast have discrete replicators containing highly conserved sequence elements (Newlon 1996). Such replicators are referred to as autonomously replicating sequences (ARSs) because they can support the autonomous replication of extrachromosomal plasmids. A 17-bp ARS consensus sequence (ACS or A element) and a second sequence element B1 are present in every budding yeast origin and are required for binding of the initiator protein origin recognition complex (ORC) (Bell and Stillman 1992; Marahrens and Stillman 1992; Theis and Newlon 1997). Other sequences (B2 and B3 elements) distributed over a DNA segment of about 100 bp also contribute to replicator function. Importantly, it is possible to accurately predict the sequences in the budding yeast genome that function as origins of DNA replication from nucleotide sequence alone (Breier et al. 2004). This has not been possible for other eukaryotes, probably because replicators in most eukaryotes have considerably less stringent sequence requirements.

It is possible to identify sequences that function as replicators (ARS elements) in the fission yeast Schizosaccharomyces pombe. Unlike S. cerevisiae, however, the precise sequence of fission yeast replicators does not appear to be conserved, and no defined sequence elements comparable to the ACS have been identified (Dubey et al. 1994; Clyne and Kelly 1995). The required sequence elements of fission yeast replicators are dis-
tributed over very large DNA segments (500–1000 bp) and are rich in A and T residues. Indeed, the best predictor of the ability of a segment of DNA to function as an ARS is its length and AT content (Segurado et al. 2003; Dai et al. 2005). Moreover, sequence elements that contribute to ARS activity in fission yeast can be replaced by other sequences as long as they are AT-rich (Okuno et al. 1997). These unusual properties are largely explained by the finding that one of the subunits of the S. pombe ORC contains multiple copies of a DNA-binding motif called the AT hook (Chuang and Kelly 1999). The AT hook motif binds short AT tracts a few nucleotides in length. Thus, it appears that fission yeast replicators consist of multiple copies of relatively simple sequence elements (AT tracts), rather than a single, highly conserved sequence. To a first approximation, it is the number of such elements in a given segment of DNA that determines its probability of binding ORC and functioning as an origin of DNA replication. This more stochastic picture of replicator organization and activity is significantly different from that of prokaryotes and budding yeast (Dai et al. 2005).

The nature of replicators in metazoans is less well defined than in the yeasts (Chapter 2). Many careful studies have demonstrated that potential origins of DNA replication are very abundant in mammalian chromosomes. For example, analysis of the amplified DHFR locus in Chinese hamster cells indicated that there are at least 40–50 potential sites of initiation of DNA replication scattered through a 55-kb intergenic region (Vaughn et al. 1990). Most of these sites are utilized in a small fraction of cell cycles, but there is a wide variation in the probability of utilization, with some sites firing as often as once in five cell cycles. Genetic analysis has shown that no single origin is essential for initiation of DNA replication within the DHFR intergene (Mesner et al. 2003). Broad zones containing many relatively inefficient origins of replication have been observed in several other intergenes, suggesting that this is the typical pattern in metazoans (Dijkwel et al. 2000; Norio et al. 2005). Several other origins, notably the human lamin B2 and β-globin origins, appear to be more circumscribed and to have a relatively high efficiency of utilization (Aladjem et al. 1998; Abdurashidova et al. 2000). These origins represent one end of what is clearly a wide spectrum of origin efficiencies. Although sequence elements that contribute to origin efficiency have been identified in a few cases by genetic studies, no consensus replicator sequence common to metazoan origins has been identified (Aladjem et al. 1998; Altman and Fanning 2004; Paixao et al. 2004). The available evidence, taken as a whole, suggests that metazoan replicators, like those of S. pombe, consist of sequence elements that occur at frequent and nearly random intervals along chromosomal DNA, and only a small fraction of
such replicators are utilized in each cell cycle. This hypothesis is consistent with the observation that purified human or Drosophila ORC binds AT-rich DNA but does not have a strong preference for any particular DNA sequence (Vashee et al. 2003; Remus et al. 2004).

Although the evidence is largely circumstantial at this point, it seems likely that origin utilization is strongly influenced by local chromatin structure and other DNA-binding proteins. For example, experiments in budding yeast have shown that nucleosome positioning can affect the efficiency of initiation of DNA replication at specific replicators (Bell and Dutta 2002). In both yeast and Drosophila, DNA-binding proteins such as Abf1 and the Myb-protein complex bind to origins and can affect the efficiency of origin utilization (Li et al. 1998; Beall et al. 2002). In mammalian cells, an ATP-dependent chromatin-remodeling complex has been shown to be required for efficient replication of heterochromatin (Collins et al. 2002). There are also a number of studies linking the state of histone acetylation with initiation efficiency (Iizuka and Stillman 1999; Aggarwal and Calvi 2004). One unanswered question is whether there is active regulation of chromatin structure in the vicinity of active origins or whether initiation of DNA replication simply occurs passively within chromatin domains that, because of their state of acetylation or favorable distribution of nucleosomes, are permissive for assembly of initiation complexes. Likely related to the effects of chromatin on initiation of DNA replication is the long-noted correlation between transcriptional activity of a given chromosomai region and the time of its replication during S phase. This correlation has recently been studied in greater detail by the application of microarray methodology (MacAlpine et al. 2004). Chromosomal regions that are rich in active genes and have higher than average GC content tend to replicate early in S phase, whereas gene-poor regions and heterochromatin replicate late. Moreover, changes in the distribution of early-replicating origins can occur during the course of development and differentiation, concomitant with developmentally regulated changes in chromatin structure and transcriptional activity (Norie et al. 2005). The data suggest that transcription factors, perhaps acting indirectly through effects on chromatin structure, can facilitate initiation of DNA replication. However, analysis of replication timing in Drosophila cells suggested that the correlation occurs over large chromosomal domains of the order of 100 kb rather than at the level of individual genes, indicating that more complex, long-range interactions may be involved as well (MacAlpine et al. 2004). Whereas the relationship between transcription and replication is clearly discernible for a significant fraction of the genome, the majority of chromosomal segments in human cells do not appear to exhibit a fixed time of replication, but rather replicate asyn-
chronously throughout S phase (Jeon et al. 2005). This finding is consistent with the hypothesis that most metazoan origins are inefficient and that the pattern of origin utilization varies significantly from one cell cycle to the next.

THE FIRST STEP—LOADING THE HELICASE

When Watson and Crick initially advanced the structure of the double helix, it was immediately apparent that if the two strands could be separated, the base-pairing rules would dictate the sequence of two daughter duplexes. Thus, in both prokaryotes and eukaryotes, the key first step in DNA replication is loading of a helicase onto the DNA to unwind the double helix (Chapter 3). This process is carefully controlled to ensure that the genomic DNA is not unwound at an inappropriate time or place. Although the biochemical details of helicase loading in eukaryotes are not yet well understood, the identity of most, if not all, of the key players and the approximate order of events are now reasonably clear. The first step is the binding of the ORC to chromosomal DNA. The discovery of ORC was a major advance that made it possible to explore the early events which occur at origins of DNA replication. ORC was initially detected as a protein that bound specifically to budding yeast replicators. It was found to contain six subunits (Orc1-6) that are conserved in all eukaryotes. After binding DNA, ORC recruits two additional initiation factors, Cdc6 and Cdt1. The resulting complex acts as a machine to load the hetero-hexameric MCM2-7 complex, which almost certainly functions as the replicative helicase during eukaryotic DNA replication. The resulting assembly of proteins at origins of DNA replication is called the pre-replication complex (pre-RC) (Diffley et al. 1994) and is a major target of the regulatory mechanisms that control eukaryotic DNA replication (Fig. 1).

The loading of the MCM2-7 complex is a unidirectional process driven by the binding and hydrolysis of ATP. Ten of the polypeptide chains involved in pre-RC assembly belong to the AAA+ family of ATPases (Orc1, Orc4, Orc5, Cdc6, MCM2-7). Recent studies have suggested that Orc2 and Orc3 may be distantly related members of the AAA+ family as well (Speck et al. 2005). Analysis of mutant proteins has given some insight into the roles of ATP binding and hydrolysis in pre-RC assembly (Bell and Dutta 2002). ATP binding by Orc1 stimulates its binding to DNA. Once bound to DNA, the ATPase activity of ORC is much reduced relative to that of free ORC. The formation of ORC-Cdc6-Cdt1 complexes can occur in the presence of nonhydrolyzable ATP analogs, but the subsequent loading of Mcm2-7 is absolutely dependent on ATP hydrolysis (Gillespie et al. 2001; Randell et al. 2006). In vitro studies in the bud-
 ding yeast system suggest that ATP hydrolysis by Cdc6 plays a particularly important role in pre-RC formation. Both binding and hydrolysis of ATP by Cdc6 are dependent on interaction with ORC in association with origin DNA, and hydrolysis of ATP by Cdc6 is absolutely required for MCM2-7 loading. ATP hydrolysis by ORC may drive a later step in initiation. Thus, although much remains to be learned, it is already clear that the binding and hydrolysis of ATP play a central role in ensuring that the steps in pre-RC assembly occur in the proper order.

Interestingly, evidence from in vitro studies in both budding yeast and *Xenopus* indicates that multiple MCM2-7 complexes can be loaded by a single ORC-Cdc6-Cdt1 complex, at least when loading is uncoupled from downstream events (Edwards et al. 2002; Bowers et al. 2004). The ATPase activity of ORC appears to be required for the reiterative loading

Copyright 2006 Cold Spring Harbor Laboratory Press.
of MCM2-7, but not for the loading of the first MCM2-7. The biological meaning of this phenomenon is not yet clear, but it may serve in some way to increase the efficiency of initiation, since the events that occur after pre-RC formation appear to depend on MCM2-7, but not on ORC or Cdc6 (see below).

The evidence that MCM2-7 acts as the replicative helicase is now quite convincing (Takahashi et al. 2005; Chapter 4). However, so far only subcomplexes, consisting of the MCM4/5/6 subunits, have been shown to have helicase activity in vitro. Direct demonstration of helicase activity of the complete MCM2-7 complex will likely require purification or reconstitution of more complete complexes containing additional factors required for helicase activity (see below). Studies of eukaryotic MCM2-7 or its archaeal counterparts by electron microscopy and X-ray diffraction suggest that it forms a hexameric or double-hexameric ring with a central cavity (Fletcher et al. 2003; Yabuta et al. 2003). It seems likely that MCM2-7 hexamers encircle the DNA, but it is not clear whether one or both DNA strands pass through the central cavity. Elucidation of the biochemical mechanism of helicase action is a critically important unsolved problem.

MAKING A FORK—HELICASE ACTIVATION AND THE TRANSITION TO DNA CHAIN ELONGATION

Once the MCM2-7 helicase has been loaded on the DNA during G1, the functions of ORC, Cdc6, and Cdt1 appear to be dispensable, and, at least in some systems, Cdc6 and Cdt1 appear to dissociate from DNA (Hua and Newport 1998; Gillespie et al. 2001). The subsequent activation of helicase activity at the G1/S transition leads to the local unwinding of the duplex and the initiation of DNA synthesis (Chapter 4). Helicase activation requires the activity of two protein kinases and the association of MCM2-7 with numerous additional replication proteins, most of which were originally identified by genetic or biochemical studies in budding yeast. Studies of initiation of DNA replication in the *Xenopus* egg extract system have provided additional insights into the order of events after pre-RC assembly.

Helicase activation and local DNA unwinding require at least a dozen additional polypeptide chains in addition to DNA-bound MCM2-7: Mcm10, Dpb11, Sld2, GINS (a complex of Sld5, Psf1, Psf2, and Psf3), Cdc45, Sld3, DNA pol-ε (in yeast), and RP-A. These proteins assemble into a complex that has been called the pre-initiation complex to distinguish it from the pre-RC (Fig. 1). We use this term to include all of the
proteins that assemble with MCM2-7 prior to the synthesis of the first primer for DNA synthesis. Although the interactions of these proteins with each other and with MCM2-7 have been studied extensively, their biochemical functions are largely unknown (Chapter 4). Moreover, few of them contain conserved motifs that might provide clues about function. Whereas the cast of players appears to be largely the same in the unicellular budding yeast and the vertebrate *Xenopus laevis*, the available data show some variation in their order of assembly and their interactions with other factors in the pre-initiation complex. It is possible that this variation is more apparent than real, reflecting small differences in the relative affinities of the proteins for each other, but further biochemical work will clearly be required to understand in detail how these proteins assemble. Several of the proteins required for helicase activation, including Mcm10, GINS, and Cdc45, appear to remain associated with MCM2-7 and travel with active replication forks, suggesting that they may be required continuously for helicase activity or may contribute in other ways to the polymerization of growing DNA chains (Aparicio et al. 1997; Pacek et al. 2006; see below).

Helicase activation and initiation of DNA replication at the G1/S transition absolutely require the activities of two hetero-dimeric protein kinases: Cdc7-Dbf4 (DDK) and CDK-cyclin. The catalytic subunits of both kinases (Cdc7 and CDK) are expressed at relatively constant levels during the cell cycle, but the regulatory subunits (Dbf4 and cyclin) oscillate in abundance, increasing significantly at the onset of S phase (Chapter 4). Two critical issues for understanding how eukaryotic DNA replication is initiated are the identity of the key targets that are phosphorylated by these kinases and the biochemical mechanisms by which phosphorylation activates pre-initiation complexes. There is circumstantial evidence suggesting that one target of Cdc7-Dbf4 is the MCM2-7 complex itself. Several subunits of the MCM2-7 are substrates of purified Cdc7-Dbf4 in vitro, and the enzyme is required for the phosphorylation of at least one subunit (Mcm2) at the G1/S transition in vivo. Additionally, a mutation in the gene encoding Mcm5 of budding yeast (*bob-1*) bypasses the requirement for Cdc7-Dbf4 in initiation of DNA replication, although subsequent DNA replication is very inefficient (Hardy et al. 1997). It has been speculated that the *bob1* mutation might induce a conformational change in MCM2-7 that mimics the effects of phosphorylation by Cdc7-Dbf4. However, to date there has been no direct demonstration that the MCM2-7 complex is a critical target of the enzyme. In the case of CDK-cyclin, recent studies in budding yeast have demonstrated that Sld2 is a key substrate in vivo (Masumoto et al. 2002; Noguchi et al. 2002). Sld2 has several CDK phosphorylation sites whose elimination prevents assembly of pre-initiation complexes and initiation of DNA replication. The evidence
indicates that phosphorylation of Sld2 by the S-phase CDK-cyclins of yeast, Cdc28-Clb5 or Cdc28-Clb6, is required for its interaction with another protein, Dbp11. Genetic analysis has shown that association of the dimeric Sld2-Dbp11 complex with MCM2-7-DNA complexes is interdependent with the association of several other proteins required for helicase activation and initiation, including GINS, DNA pol-ε, and Cdc45 (Takayama et al. 2003). Thus, in yeast, CDK-cyclin phosphorylation controls a critical protein–protein interaction that is essential for the assembly of the pre-initiation complex as a whole. The critical substrate(s) of CDK-cyclin activity in metazoans has not yet been identified. Although metazoans contain a protein, RecQL4, that is required for DNA unwinding and exhibits homology with Sld2 at its amino terminus, it is not yet known whether this protein interacts with metazoan Dpb11 or is a target of CDK-cyclin activity (Sangrithi et al. 2005).

The association of the DNA polymerases and other components of the machinery involved in the priming and elongation of DNA chains with pre-initiation complexes is a comparatively late event, probably occurring after local unwinding of the duplex. One known exception is the association of DNA pol-ε in budding yeast, which appears to be essential for assembly of the pre-initiation complex (Morrison et al. 1990). The essential function can be fulfilled by the carboxy-terminal domain of DNA pol-ε, which lacks polymerase activity, suggesting that the role of the protein in the assembly of pre-initiation complexes is architectural. In *Xenopus* and perhaps other metazoans, DNA pol-ε does not appear to play a role in the assembly of pre-initiation complexes, because helicase activation and unwinding can occur in its absence (Mimura et al. 2000). DNA pol-α is absolutely required for initiation because it is the only source of primase activity (see below). Its association with DNA occurs only after helicase activation and DNA unwinding. Various studies have implicated MCM2-7, Mcm10, and Cdc45 in recruiting DNA pol-α to replication forks (Mimura and Takisawa 1998; Aparicio et al. 1999; Ricke and Bielinsky 2004). On the basis of studies in the simian virus 40 (SV40) cell-free replication system, stable association of DNA pol-δ with replication forks probably requires synthesis of a primer and the loading of PCNA by the replication factor C (RF-C) complex (see below).

**CONTROLLING INITIATION OF REPLICATION**

DNA replication in eukaryotic cells is largely regulated at the level of initiation, and most of the known regulatory mechanisms act on pre-RC assembly or activation. As described above, the onset of S phase appears to be triggered by the activation of S-phase CDK and DDK protein kinases,
which in turn depends on accumulation of the corresponding regulatory subunits and other mechanisms. How kinase activation is coupled to cell growth is a major issue that is not yet completely understood. It is also important to keep in mind that the existence of origins which reproducibly fire early or late in S phase means that factors other than kinase activation, perhaps including local chromatin structure, can influence the time of initiation at individual origins. The nature of these factors is also poorly understood.

An equally important regulatory question is how DNA replication is limited to a single round each cell cycle. The fundamental problem is to ensure that the hundreds or thousands of initiation events that occur during S phase are confined exclusively to the parental DNA and never occur on the newly synthesized daughter DNA. The general logic of the mechanisms that accomplish this feat can be described rather simply: The assembly of pre-RCs, but not their activation, can occur in one part of the cell cycle (late M through G1), whereas activation of pre-RCs, but not their assembly, can occur in another part of the cell cycle (S to M). Surprisingly, given the central importance of this regulatory mechanism, its specific molecular implementation appears to be somewhat different in yeasts and metazoans (Chapter 3).

In unicellular eukaryotes, oscillation of CDK activity during the cell cycle largely determines the cellular states that are permissive or nonpermissive for pre-RC assembly (Chapter 15). From late M through G1, CDK activity is low and pre-RC assembly is efficient. The rise in CDK activity at the G1/S transition triggers initiation but also inhibits further pre-RC assembly. The importance of CDK in preventing pre-RC assembly has been demonstrated by experiments in which genetic manipulations were employed to reduce CDK activity after the onset of S phase. Under these conditions, “re-replication” occurs, resulting in abnormal increases in cellular DNA content because of inappropriate initiation on daughter DNA molecules (Hayles et al. 1994; Dahmann et al. 1995). A number of studies have demonstrated that CDK acts on multiple targets to prevent pre-RC assembly (Gopalakrishnan et al. 2001; Nguyen et al. 2001; Yanow et al. 2001). In both budding and fission yeast, CDK phosphorylation of Cdc6 results in its SCF-dependent ubiquitinylation and degradation by the proteasome (Drury et al. 1997; Jallepalli et al. 1997). There is also evidence in both yeasts that CDK-dependent phosphorylation of ORC subunits contributes to inhibition of pre-RC assembly by an unknown mechanism that does not appear to involve regulated degradation (Nguyen et al. 2001). In budding yeast, CDK phosphorylation of MCM2-7 causes it to be exported from the nucleus, but this does not appear to be an important regulatory mechanism in most other eukaryotes. Finally, in fission...
yeast, the regulation of Cdt1 contributes significantly to prevention of pre-RC assembly after the onset of S phase. Cdt1 is degraded in S phase with kinetics similar to those of Cdc6. However, it is not known whether Cdt1 degradation depends on phosphorylation by CDK. Deregulated expression of Cdt1 and Cdc6 in fission yeast is sufficient to induce re-replication (Gopalakrishnan et al. 2001; Yanow et al. 2001). The Cdt1 homolog in budding yeast is not periodically degraded but, like MCM2-7, is exported from the nucleus after the G1/S transition (Tanaka and Diffley 2002).

CDK-dependent inhibition of pre-RC assembly is also important in metazoans (Chapter 16). In mammalian cells, Cdt1 is phosphorylated by CDK at the G1/S transition and then degraded in a SCF (Skp2)-dependent manner (Liu et al. 2004; Nishitani et al. 2004; Sugimoto et al. 2004). Human Orc1 is degraded at the onset of S phase, but it is not yet clear whether this is dependent on CDK. Interestingly, metazoan Cdt1 is also targeted for inactivation by two CDK-independent mechanisms that have no clear precedent in unicellular eukaryotes. First, in Xenopus egg extracts, degradation of Cdt1 has been shown to be dependent on DNA replication. Recent work has demonstrated that this replication-dependent degradation requires a specific interaction between Cdt1 and the replication processivity factor, proliferating cell nuclear antigen (PCNA) (Takeda et al. 2005; Arias and Walter 2006). Mutation of the PCNA interaction site stabilizes Cdt1 and results in re-replication. Second, Cdt1 activity in metazoan cells is regulated by the inhibitor geminin, which accumulates from the onset of S phase until M phase, when it is degraded by the APC (Machida et al. 2005). Binding of geminin to Cdt1 in the pre-RC blocks loading of the MCM2-7 complex. The regulation of Cdt1 activity by geminin appears to be an important regulatory mechanism, because depletion of geminin in human and Drosophila cells induces re-replication.

**COMMITMENT TO ENTER S PHASE**

Following the activation of CDK-cyclin and DDK, one of the first processes to be induced is the means to produce the deoxyribonucleoside triphosphates (dNTPs), the precursors for DNA synthesis. The rate-limiting step is the synthesis of deoxyribonucleoside diphosphates (dNDPs) from ribonucleoside diphosphates (NDPs) by an enzyme called ribonucleotide reductase (RNR) (Jordan and Reichard 1998; Eklund et al. 2001). This enzyme reduces a hydroxyl group on the ribose sugar of each of the four NDPs to produce four deoxyribonucleotides, yielding products dCDP, dUDP (that is then converted into dTDP), dGDP, and finally, dADP. The diphosphates are then phosphorylated to produce dNTPs. dATP is unique because it feeds back to inhibit the RNR when sufficient
dNTPs are made (Chabes and Thelander 2003). RNR is activated after commitment to cell division and thus has very low activity in G1 when the pre-RC are being assembled. Cells go to considerable lengths to ensure that RNR has low activity in the G1 phase of the cell cycle, sufficient for mitochondrial DNA synthesis. The \textit{RNR1} gene is transcriptionally regulated, the Rnr1 protein is unstable, and any enzyme activity that does exist is subject to allosteric feedback inhibition (Elledge et al. 1992; Chabes and Thelander 2003). Once in S phase, the dNTPs produced are made in limiting amounts. At any given time in S phase, there are enough dNTPs to replicate a small fraction of the genome. Thus, inhibition of RNR activity with hydroxyurea will retard DNA replication fork progression.

Studies in the budding yeast show that cells have another cycle called the oxidative-reductive cycle which may represent an intrinsic cell clock that controls transcription and when DNA synthesis occurs (Klevecz et al. 2004). Cells synchronize into two main phases, a respiration phase, when mitochondria are actively producing energy via the oxidative cycle, and a reductive phase. DNA synthesis occurs in the reductive phase. The relatively low levels of DNA-damaging reactive oxygen species in the reductive phase may provide a better environment to replicate DNA. RNR may play an active role in restricting DNA replication to the reductive phase because RNR utilizes a radical tyrosyl residue and reduced cysteines in its catalysis. A reducing cellular environment would favor regeneration of active RNR after each catalytic reaction and hence promote DNA synthesis. How regulation of the oxidative-reductive cycle ties into the known cell-cycle regulation by CDK-cyclin and DDK remains to be determined, but it is most likely a result of coupling transcription to the oxidative-reductive cycle.

**THE CORE DNA REPLICATION FORK MACHINERY**

Commitment to cell division activates the regulators of DNA replication that are discussed in Chapters 15 and 16. The DNA replication machinery is very complex, perhaps involving 100 proteins at a single DNA replication fork. Many of these proteins form the core of the DNA synthesis machinery, whereas others play accessory roles. Proteins in the latter class ensure accurate synthesis of DNA or monitor and coordinate the progression of all active forks at any given time in S phase. Other proteins couple DNA replication to inheritance of chromatin structures, such as assembly of nucleosomes and epigenetic states of gene expression.

Biochemical studies on the replication of SV40 DNA in extracts from human cells resulted in the discovery of the core DNA replication fork machinery in eukaryotes (Waga and Stillman 1998; Kelly and Brown 2000).
2000). This was made possible by the development of a cell-free extract that could replicate SV40 origin containing DNA (Li and Kelly 1984). The only virus-encoded protein necessary for replication in cells and in vitro is SV40 T antigen, a protein that has at least four functions in DNA replication: recognition of specific DNA sequences at the SV40 origin of DNA replication (ori), primasome loading, DNA helicase activity, and recruitment of topoisomerase 1 (Borowiec et al. 1990; Waga and Stillman 1998; Eichman and Fanning 2004). Primasome loading involves the recruitment of the DNA primase enzyme that in eukaryotes is a part of the DNA polymerase α enzyme (Hubscher et al. 2002). Primase synthesizes the necessary RNA primers that are utilized by DNA polymerases to begin DNA synthesis and, as such, marks the first step in actual DNA synthesis. T antigen recruits the four-subunit polymerase α/primase complex to the origin along with another cellular replication protein RP-A (replication protein A). RP-A is the eukaryotic single-stranded DNA (ssDNA)-binding protein that contains multiple OB fold domains within its three subunits, domains that are similar to the OB fold in the bacterial ssDNA-binding protein SSB (Wold and Kelly 1988). Interestingly, T antigen, topoisomerase 1, and RP-A by themselves can recognize the SV40 origin, cause local unwinding, and then activate the helicase activity of T antigen to unwind most of the circular DNA template in the absence of DNA synthesis (Borowiec et al. 1990). Thus, the DNA helicase and RP-A provide the principal driving force to produce the ssDNA templates for both leading- and lagging-strand synthesis.

Because DNA polymerases cannot spontaneously incorporate dNTPs into a growing chain without a 3’OH at the end of a primer that is base-paired to a template DNA, the short RNA primer produced by polymerase α/primase provides the solution for initiation of the leading strand at each origin of DNA replication and for the initiation of DNA synthesis at each Okazaki fragment on the lagging strand (Fig. 2) (Hubscher et al. 2002). The short RNA is immediately extended by the DNA polymerase α activity of the complex to synthesize a short initiator DNA (iDNA) of about 30 bases. Because polymerase α does not have a proofreading exonuclease, an enzyme that can correct mistakes inserted by the DNA polymerase catalytic activity, it would be potentially mutagenic to include the iDNA in the final product of DNA replication. Thus, the iDNA only serves as a primer for more extensive DNA synthesis by another DNA polymerase called DNA polymerase δ, which has such a proofreading active site in addition to its DNA polymerase active site (Lee et al. 1989; Tsurimoto et al. 1990). The iDNA is later removed during maturation of Okazaki fragments, reducing any possibility of mutations.

DNA polymerase switching occurs at the start of leading-strand syn-
thesis and during initiation of every Okazaki fragment on the lagging strand (Fig. 2) (Waga and Stillman 1994) and is perhaps a unique feature of eukaryotic DNA replication because the bacterial replicative polymerase utilizes the RNA primer directly and synthesizes long DNA strands (Johnson and O’Donnell 2005). The question arises as to why DNA polymerase δ does not directly utilize the RNA primer made by primase and why in eukaryotes primase is linked to a DNA polymerase. Two explanations spring to mind. From an evolutionary perspective, bacteria may have evolved to be more efficient by eliminating a polymerase that does not have proofreading activity. Alternatively, DNA polymerase α/primase may be required to monitor simultaneous DNA synthesis from multiple origins or be a target of DNA damage-induced arrest of DNA replication fork progression in S phase. There is some evidence implicating the production of the RNA–DNA primers in S-phase checkpoint signaling, a regulatory process that prevents DNA replication until DNA damage has been repaired (Pellicioli et al. 1999; Michael et al. 2000).

Polymerase δ is a four-subunit enzyme that cannot function efficiently as a DNA polymerase on its own and thus needs accessory factors to help it load onto the primer and to make it a processive polymerase capable of synthesizing long DNA strands without dissociating from the template. The accessory factor that allows the polymerase to move along the template DNA without dissociating is called PCNA (Maga and Hubscher 2003; Bowman et al. 2005). PCNA is a trimer of a 36-kD protein which forms a doughnut-shaped structure that enables double-stranded DNA (dsDNA) to pass through the hole in the middle (Fig. 2). Because the PCNA trimer encircles dsDNA and binds the polymerase that is actively copying the template strand, it acts as a sliding molecular tether to hold DNA polymerase δ onto the DNA. Even if the polymerase were to stop incorporating nucleotides into the nascent DNA strand, PCNA holds the polymerase in close proximity to the 3’-hydroxyl at the end of the nascent DNA strand so that the local concentration of the primer-template DNA remains high. Thus, the polymerase can continue without dissociating. For these reasons, the PCNA molecule is commonly called a DNA polymerase clamp; bacterial, bacteriophage T4, and archaeal versions of polymerase clamps are known (Chapter 14; the mechanism of DNA polymerase clamps is also discussed in Chapters 5 and 21).

Because the PCNA molecule is ring-shaped, there must be a mechanism to load the ring onto the double-stranded iDNA that acts as a primer for DNA polymerase δ. A specialized clamp loader protein called RF-C performs this function in eukaryotes, and homologs of clamp loaders exist in bacteria, bacteriophage T4, and archaeal species (Waga and Stillman 1998; Johnson and O’Donnell 2005; see Chapter 14). The clamp
Figure 2. DNA replication forks. Scheme for how the core of the DNA replication machinery operates at DNA replication forks. A DNA helicase unwinds the DNA in a bidirectional manner, and RP-A binds to the ssDNA to create template strands for leading- and lagging-strand DNA synthesis. Via interactions with the helicase and RP-A, polymerase α/primase is loaded onto the template and an RNA–DNA hybrid primer (red) is synthesized. Primer DNA is then recognized by RF-C in an ATP- and RPA-dependent manner. RF-C then loads the DNA polymerase clamp PCNA onto the duplex region of the DNA. This process expels the polymerase α/primase, and PCNA and RF-C recruit another DNA polymerase, such as DNA polymerase δ. Polymerase α/primase is recycled for priming of additional Okazaki fragments on the lagging strand. This scheme is based on studies of SV40 DNA replication, where the helicase is T antigen. Many other proteins, including another DNA polymerase (pol ε) and some pre-IC proteins (see Fig. 1) participate at cellular DNA replication forks; however, how they work is not yet clear. MCM hexamers are part of the DNA helicase at such forks. Not shown is the process of removal of the RNA–DNA primer at the 5’ end of Okazaki fragments and ligation of the fragments for a continuous lagging strand.
loader RF-C is a five-subunit enzyme; all subunits are members of the AAA+ family of proteins that have a three-dimensional structure common to ATPases (Bowman et al. 2005). Each AAA+ subunit embraces a neighboring subunit, and hydrolysis of ATP by one subunit is stimulated by an adjacent subunit. RF-C recognizes the primer-template DNA structure in an ATP-dependent manner that also requires RP-A to bind to the single-stranded template DNA (Waga and Stillman 1998). ATP binding promotes structural changes in the RF-C subunits that ultimately cause an α-helix in one of the subunits to thrust itself into the PCNA molecule and break an interface between two of the PCNA subunits. This force creates a break in the clamp ring through which dsDNA can pass (Kazmirski et al. 2005). When the clamp loader is in its active state, the bound PCNA clamp has a structure like an open lock-washer. Upon passage of the dsDNA through the open clamp, DNA-activated ATP hydrolysis within the RF-C subunits causes the RF-C to release from the DNA and to allow the polymerase to engage the clamp. In some ways, this is reminiscent of origin DNA sequences activating the ATPase activity of the AAA+ proteins in the ORC-Cdc6. Details of the mechanism of clamp loading and the function of other related clamp loaders involved in response to DNA damage and in sister chromatid cohesion are discussed in more detail in Chapter 21.

Because the clamp does not touch the dsDNA that it surrounds at the back of the moving DNA polymerase, it slides along much like a hydrofoil planes across water. In addition to DNA polymerase, PCNA acts as a sliding molecular landing pad for many proteins that participate in the duplication of DNA and chromatin (Maga and Hubscher 2003). The outer surfaces of the clamp bind to DNA polymerase, but because the clamp has threefold symmetry provided by its three identical subunits, the clamp can bind other proteins. Indeed, a plethora of PCNA-binding proteins have been discovered. They include the cell-cycle regulator p21CIP1 CDK inhibitor, cytosine DNA methyltransferase involved in inheritance of methyl groups on DNA, chromatin assembly factor 1 (CAF-1) that forms new nucleosomes at the DNA replication fork, enzymes such as DNA ligase and the endonuclease FEN1 that mature the Okazaki fragment, DNA repair proteins such as 8-oxoguanine-DNA glycosylase and DNA-N-glycosylases, as well as other DNA polymerases that deal with damage in the template DNA during DNA synthesis (Maga and Hubscher 2003; see Chapters 5, 6, 20, and 21).

The mechanism of DNA polymerase switching from polymerase α to polymerase δ first observed during replication of SV40 DNA is used to load other DNA polymerases that can handle abnormal bases in the template. For example, the trans-lesion DNA polymerases η, 1, and κ bind...
to PCNA and can synthesize through damaged bases in the template DNA and then switch back to the replicative polymerase (Haracska et al. 2002, 2005; Plosky and Woodgate 2004; Lehmann 2005; Prakash et al. 2005). It is possible that the PCNA molecule binds to two DNA polymerases at the same time, allowing efficient switching back and forth. A discussion of the role of various DNA polymerases in eukaryotic DNA replication and repair is presented in Chapter 20.

Following the polymerase switch during Okazaki fragment synthesis at the DNA replication fork, polymerase δ tethered to its PCNA clamp extends the nascent DNA chain until the polymerase runs into a previously synthesized Okazaki fragment (penultimate fragment) in the lagging-strand template (Fig. 2). Two mechanisms are now recognized to remove the RNA primer and dDNA from the 5´ end of the penultimate fragment (see Chapter 5). In one model, RNase H and FEN1 couple to remove the RNA primer and, together with DNA ligase, seal the two fragments of DNA (Rumbaugh et al. 1997). An alternative mechanism in cells involves the Dna2 helicase-endonuclease that cooperates with FEN1 to create a FLAP structure to remove the RNA-dDNA primer, allowing ligation of the two Okazaki fragments by DNA ligase (Bae et al. 2001; Kao et al. 2004). Both mechanisms involve recruitment of protein such as FEN1 and DNA ligase to the site of Okazaki fragment maturation by the PCNA clamp that remains on the newly synthesized strand after the DNA polymerase is displaced. RF-C, the clamp loading ATPase, can also act to unload PCNA from dsDNA so that the PCNA can recycle for the synthesis of other Okazaki fragments. Details of these reactions are discussed in Chapter 5.

CELLULAR DNA REPLICATION FACTORS REQUIRED FOR COPYING CHROMOSOME DNA

The SV40 system has been extremely valuable in figuring out the basic mechanism of DNA replication in eukaryotes, but there were indications that this system did not reflect all of the events at the replication forks copying cellular DNA. One can think of the SV40 replication fork as a stripped-down version of the cellular replication apparatus. This is partially due to the multiple functions contributed by SV40 T antigen that need to be provided in cells and also because virus DNA replication does not have to deal with the complexities of cellular DNA replication, such as initiation from multiple origins, handling errors in the template DNA, dealing with stalled replication forks, and restricting DNA replication to one round per cell cycle. The virus has the luxury of discarding defective genomes, whereas replication of cell chromosomal DNA cannot afford such inaccuracies.
Most of the evidence for other cellular DNA replication factors comes from the yeast *S. cerevisiae*. One of the first enzymes observed to be required for cellular DNA replication was DNA polymerase ε, a four-subunit enzyme (Morrison et al. 1990). This polymerase is activated by PCNA, but under some circumstances the polymerase catalytic activity is not essential for chromosome replication, although DNA replication is clearly not normal (Ohya et al. 2002). There has been much speculation about the role of DNA polymerase ε, but a definitive role has not been deduced. It is commonly drawn in texts that DNA polymerase ε functions on the leading strand, whereas polymerase δ works on the lagging strand. Genetic studies in yeast suggest that the two DNA polymerases work on opposite strands (Karthikeyan et al. 2000; Pavlov et al. 2001). It is equally possible that polymerase ε is required for initiation of DNA replication at origins and recovery from stalled DNA replication forks, a view supported by genetic studies that link newly discovered initiation factors such as Dbp11, GINS, Sld3, and Mcm10 to this DNA polymerase (Kamimura et al. 1998; Takayama et al. 2003). Thus, the role for DNA polymerase ε remains to be determined.

Another obvious group of proteins that is essential for cellular DNA replication but is dispensable for SV40 DNA replication are the six MCM proteins that form the core of the eukaryotic replicative DNA helicase. Recent evidence demonstrates that MCM proteins are not only required for assembly of the pre-RC (discussed above), but also for the active DNA helicase at the DNA replication fork in cells (Labib et al. 2000; Byun et al. 2005; Ying and Gautier 2005). The six MCM proteins form a hexameric structure that is remarkably similar to SV40 T antigen and the papillomavirus E1 helicase (Forsburg 2004), but in eukaryotes, the MCM hexamer lacks helicase activity, perhaps requiring other DNA replication proteins. This is in contrast to the archaeal MCM double hexamer that works efficiently as a DNA helicase (see Chapter 14). It is most likely that the MCM accessory proteins play a role in coordinating replication from many origins at a single time in S phase. For example, some of these pre-initiation complex proteins are known to associate with DNA replication forks throughout S phase. What these proteins do and how they interact with the DNA polymerases is one of the major remaining questions in understanding the enzymology of eukaryotic DNA replication.

ENSURING GENOMIC INTEGRITY DURING DNA REPLICATION

As cells pass through S phase and the DNA is copied, a number of mechanisms exist which ensure that DNA replication occurs uniformly across the genome and with minimal mistakes. The above-mentioned DNA-
polymerase-associated exonuclease activities contribute to the accuracy of copying the template strands by removing incorrectly inserted nucleotides. However, the DNA replication fork also has to cope with damaged template DNAs such as bulky adducts attached to the bases, thymidine dimers induced by UV, natural genome barriers for replication fork progression, and DNA strand breaks (see Chapters 17–21). This type of damage control takes a number of different forms.

First, there are damaged bases in the template DNA that the replicative DNA polymerases cannot copy. The DNA replication fork stalls and does not resume until the damage is repaired. Alternatively, an altered base can be copied either by an error-prone mechanism that causes mutations or by an error-free mechanism, or the lesion bypassed by recombination-based mechanisms (Lambert and Carr 2005; Lehmann 2005; Prakash et al. 2005; Cahill et al. 2006). The mechanism depends on the type of DNA polymerase recruited to the lesion and whether or not the damaged base is excised (see Chapters 19 and 21).

If the DNA replication fork stalls due to a lesion on one strand, then the coordinated replication of leading and lagging strands is compromised and needs to be reestablished. Proteins found at the DNA replication fork, such as Mrc1, Cms1, Tof1, Dbp11, and DNA polymerase ε, are involved in signaling to a cellular signal transduction pathway that the fork is compromised (Alcasabas et al. 2001; Zegerman and Diffley 2003). The signal transduction pathway involves the Rad53 kinase, Mec1 kinase and its partner Ddc2 which signal to the regulators of DNA replication, DDK and CDK, that further initiation should not occur until the damage is corrected (Bartek et al. 2004; Lukas et al. 2004). This is often called the S-phase checkpoint pathway. Other proteins such as the Sgs1 helicase-Top3 complex are required for recovery of the stalled DNA replication fork so that DNA synthesis can resume, at the same time preventing potentially dangerous recombination intermediates (Fabre et al. 2002; Ira et al. 2003; Ui et al. 2005; detailed discussion of these pathways is presented in Chapters 17 and 18).

dsDNA breaks provide an extra challenge because DNA replication cannot occur when a template is missing. A dsDNA break in G1 lacks sister chromatids for recombination-based repair that is possible after the DNA has replicated, and thus, it is imperative that entry into S phase be delayed until repair of such damage occurs. Two major pathways have been recognized, nonhomologous end-joining and homologous DNA recombination, both of which are active in mammalian cells (Pastink et al. 2001). Inherited mutations in genes involved in these processes are discussed in Chapters 19, 22, and 28. We still do not know, however, how the dsDNA break-repair pathways are coordinated with DNA replication.
Repair often involves resection of long stretches of ssDNA to create a template for Rad51-mediated strand invasion, followed by replication of the exposed template by the same enzymes that are involved in duplication of chromosomal DNA, particularly RP-A, polymerase δ, PCNA, and RF-C (Wang et al. 2004). Such DNA synthesis is linked to checkpoint signaling to prevent cell-cycle progression, perhaps by RP-A coating the ssDNA. In many situations, damaged DNA causes repair that exposes ssDNA, which is then bound by RP-A. The half-life for RP-A-binding to such structures is much longer than the time RP-A spends associated with ssDNA at the DNA replication fork. Thus, one model for checkpoint signaling is the longer time for RP-A on ssDNA that enables recruitment of checkpoint proteins such as RF-C-like checkpoint clamps, DNA repair proteins, and the checkpoint signaling machinery.

CHROMATIN ASSEMBLY AT THE DNA REPLICATION FORK

Each time the DNA is duplicated, the associated protein complexes must be copied, including the histones that form the core of nucleosomes. Histones H2A and H2B form dimers and are loaded onto the chromatin after the core H3/H4 tetramer is formed, and the loading of these dimers is facilitated by chaperone proteins such as NAP1 and nucleoplasmin that bind H2A/H2B dimers (Laskey et al. 1993; Adams and Kamakaka 1999). The available evidence suggests that H2A/H2B dimers associated with the unreplicated DNA are displaced from the chromatin as the DNA replication fork passes and both new and old H2A/H2B histone dimers are reassembled onto the chromatin shortly after passage of the DNA replication fork. In the SV40 system, mature nucleosomes are assembled within 300 nucleotides of the newly incorporated DNA at the fork (Herman et al. 1981; Gruss et al. 1990).

The key and rate-limiting step in nucleosome assembly occurs at the DNA replication fork and is mediated by H3/H4 chaperones CAF-1 and ASF1 (Verreault 2000; Krude and Keller 2001; see Chapter 6). Both bind H3/H4 complexes and load them onto the replicating DNA coupled to passage of the DNA replication fork. Coupling DNA replication to chromatin assembly at the DNA replication fork occurs by the direct binding of CAF-1 to PCNA (Zhang et al. 2000). This is another example of how a sliding clamp can tether a protein to the DNA replication fork. Moreover, CAF-1 binds directly to the heterochromatin protein 1 (HP1) that assembles onto heterochromatic foci during passage of the DNA replication fork (Murzina et al. 1999). Recently, ASF1 has been shown to bind RF-C, the clamp loader protein, but it is not known whether this interaction is required to bring this histone chaperone to the DNA replication fork.
fork or whether it is required for another aspect of ASF1 activity in checkpoint control (Franco et al. 2005). Although it is not discussed here, histones and DNA are modified, and the nature of the modification influences whether the region of the genome will be active in transcription or silenced and heterochromatic. Many DNA- and histone-modifying enzymes are tethered to the DNA replication apparatus or to locus controlling elements and ensure that the inheritance of such modifications occurs in the right place in the genome.

OVERVIEW OF FUTURE OF DNA REPLICAION

We have not discussed many aspects of DNA replication in eukaryotes, including termination and replication of specialized structures such as chromosome ends, but these topics are addressed in later chapters. Although the field of DNA replication in eukaryotes is quite mature, with major advances occurring over the past 25 years, there are still many open questions and issues that have not been resolved.

Perhaps most obvious is the lack of understanding of the biochemical nature of the pre-initiation complex and the subsequent mechanism of initiation of DNA synthesis at each chromosomal origin of DNA replication. Many of the components are now known, but the lack of a suitable biochemical assay for initiation of DNA synthesis with purified proteins makes this task a difficult one.

There is still much to learn about how chromosome replication occurs in the nucleus of cells. Of particular interest is the temporal regulation of DNA replication and the dynamic spatial location of the replication apparatus within the nucleus. How this is coordinated with other cellular processes such as gene transcription is still a research area in its infancy. DNA replication is coordinated from many loci throughout S phase, but there are only hints about how this occurs and how DNA replication may influence chromosome structure and gene expression, both globally across the genome and locally at specific loci during development. An interplay between histone modifications and epigenetic inheritance processes is an area of research that has a long way to go.

One intriguing discovery over the past few years has been the role of DNA replication initiation proteins in processes other than DNA replication. ORC subunits have been shown to be required for gene silencing in yeast, for heterochromatin maintenance in Drosophila and human cells, for chromosome segregation and centrosome function during mitosis, and even for cytokinesis (Loo et al. 1995; Triolo and Sternglanz 1996; Pak et al. 1997; Loupart et al. 2000; Prasanth et al. 2002, 2004; Chesnokov et al. 2003; Shareef et al. 2003). This raises the possibility that ORC
and potentially other DNA replication proteins (Tachibana and Nigg 2006) are required for coordinating DNA replication with other key processes in the chromosome duplication and segregation cycle, and even for cell division itself. Such coordination may be subordinate to the cyclin-CDK kinases that orchestrate the timing of cell-cycle events and ensure that one event is complete before proceeding to the next.

The DNA replication field has diverged into two major paths, as have other fields of cell biology. One path involves the detailed understanding of the structure and chemistry of the process of DNA replication, and for many replication proteins and complexes, particularly DNA polymerases and their accessory proteins, this approach is well under way. A second path will involve the biology of DNA replication in cells and organisms, determining how genomes are duplicated and how this meshes with nuclear structure, gene expression, and processes such as DNA repair, recombination, and chromosome segregation. For the immediate future, both areas of investigation will be very active, and they will continue to yield new surprises.

REFERENCES


Copyright 2006 Cold Spring Harbor Laboratory Press.


Copyright 2006 Cold Spring Harbor Laboratory Press.


Copyright 2006 Cold Spring Harbor Laboratory Press.

Copyright 2006 Cold Spring Harbor Laboratory Press.


Copyright 2006 Cold Spring Harbor Laboratory Press.


Copyright 2006 Cold Spring Harbor Laboratory Press.
The ability of Sgs1 to interact with DNA topoisomerase III is essential for damage-induced recombination. DNA Repair 4: 191–201.


