Selective Bypass of a Lagging Strand Roadblock by the Eukaryotic Replicative DNA Helicase

Yu V. Fu, Hasan Yardimci, David T. Long, The Vinh Ho, Angelo Guainazzi, Vladimir P. Bermudez, Jerard Hurwitz, Antoine van Oijen, Orlando D. Schärer, and Johannes C. Walter

SUMMARY

The eukaryotic replicative DNA helicase, CMG, unwinds DNA by an unknown mechanism. In some models, CMG encircles and translocates along one strand of DNA while excluding the other strand. In others, CMG encircles and translocates along duplex DNA. To distinguish between these models, replisomes were confronted with strand-specific DNA roadblocks in Xenopus egg extracts. An ssDNA translocase should stall at an obstruction on the translocation strand but not the excluded strand, whereas a dsDNA translocase should stall at obstructions on either strand. We found that replisomes bypass large roadblocks on the lagging strand template much more readily than on the leading strand template. Our results indicate that CMG is a 3′ to 5′ ssDNA translocase, consistent with unwinding via “steric exclusion.” Given that MCM2-7 encircles dsDNA in G1, the data imply that formation of CMG in S phase involves remodeling of MCM2-7 from a dsDNA to a ssDNA binding mode.

INTRODUCTION

In eukaryotic cells, DNA replication initiates at many chromosomal locations called origins. At each origin, two “sister” replisomes are assembled that move away from the origin in opposite directions. An essential component of the replisome is the replicative DNA helicase, which unwinds parental DNA, generating substrates for leading and lagging strand DNA polymerases. Current evidence indicates that the eukaryotic replicative DNA helicase contains at least three components, a heterohexameric ATPase called MCM2-7 and two cofactors, Cdc45 and GINS (Bochman and Schwacha, 2009; Ilves et al., 2010; Moyer et al., 2006; Pacek et al., 2006; Takahashi et al., 2005). The complex of Cdc45, MCM2-7, and GINS is called CMG (Moyer et al., 2006).

A striking feature of eukaryotic DNA replication is the intricate, bi-phasic assembly of CMG, which underlies the cell cycle regulation of DNA replication (reviewed in Arias and Walter, 2007; Labib, 2010). CMG assembly is best understood in yeast. In the G1 phase, when cyclin-dependent kinase (CDK) activity is low, ORC, Cdc6, and Cdt1 recruit MCM2-7 onto origins of DNA replication, forming a prereplication complex (pre-RC). Within the pre-RC, MCM2-7 complexes bind to double-stranded DNA (dsDNA) as inactive double hexamers (Evrin et al., 2009; Gambus et al., 2011; Remus et al., 2009). At the G1/S transition, when CDK activity rises, numerous additional factors cooperate to convert the MCM2-7 double hexamer into two CMG complexes. In particular, Cdc7-Dbf4 protein kinase (DDK) phosphorylates MCM2-7 (Randell et al., 2010 and references therein). CDK phosphorylates Sld2 and Sld3 (Treslin/Ticrr in metazoans), promoting their interaction with Dpb11 (TopBP1 in metazoans). The Sld3-Sld2-Dpb11 complex enables the stable binding of Cdc45 and GINS to phosphorylated MCM2-7. Once formed, CMG unwinds the origin, allowing replisome assembly. Importantly, the high CDK activity present in the S and G2 phases inhibits the functions of ORC, Cdc6, and Cdt1, such that de novo MCM2-7 recruitment is blocked. As a result, CMG complexes that leave the origin during the first initiation event cannot be replaced, and each origin fires only once. Although the regulation of CMG assembly in metazoans is very similar, it appears to be even more complicated, depending on several additional factors, including Mcm9, DUE-B, GEMC1, Geminin, and CRL4Cdt2.

A crucial question in the field is why so many proteins are required to convert bound MCM2-7 complexes into the active CMG helicase. This question is inextricably connected to another open question, which is how CMG unwinds DNA in S phase. Only by describing the final disposition of CMG on DNA will it be possible to understand how it is assembled and activated. MCM2-7, the core molecular motor of CMG, is composed of six related AAA+ ATPase subunits, Mcm2-Mcm7 (Tye, 1999).
Thus, GINS and Cdc45 have been proposed to bind to the outside of the MCM2-7 complex and to regulate its ATPase activity (Ilyes et al., 2010). Crystallography and electron microscopy reveal that eukaryotic and archaeal MCM complexes form ring-shaped hexamers whose central channel can accommodate single-stranded DNA (ssDNA) or dsDNA (Evrin et al., 2009; Fletcher et al., 2003; Pape et al., 2003; Remus et al., 2008). Thus, DNA unwinding likely involves one or both strands of the DNA passing through the MCM central channel.

Most well-studied replicative DNA helicases, such as E. coli DnaB and Bacteriophage T7 gene product 4 (T7 gp4), function as single hexamers that encircle and translocate along one strand of DNA while excluding the other strand (“steric exclusion”). In biochemical experiments using purified proteins, archaeal and eukaryotic MCM proteins including the CMG translocate along ssDNA in the 3′ to 5′ direction (Bochman and Schwacha, 2008; Ilyes et al., 2010; Kelman et al., 1999; Moyer et al., 2006; Shechter et al., 2000). Using elegant manipulation of DNA templates, steric exclusion by a purified Mcm4/6/7 subcomplex was directly demonstrated (Kaplan et al., 2003). However, the physiological relevance of these studies is not clear since the reactions bypassed the endogenous helicase loading and activation pathway.

More recently, several considerations led to speculation that CMG might unwind DNA by translocating along dsDNA (Laskey and Madine, 2003; Mendez and Stillman, 2003; Takahashi et al., 2005). First, Mcm4/6/7 can translocate along dsDNA with considerable force (Kaplan et al., 2003), Second, MCM2-7 interacts with ssDNA as part of the pre-RC (Evrin et al., 2009; Remus et al., 2009). It is therefore attractive to propose that MCM2-7 does not fundamentally change its interaction with DNA upon assembly into the CMG. Several models propose how CMG might unwind DNA while translocating along dsDNA. One idea is that CMG functions as an obligate dimer that pumps dsDNA toward its dimer interface, and that single-stranded DNA is extruded through lateral channels, analogous to a mechanism proposed for SV40 large T-antigen (Li et al., 2003; Wessel et al., 1992). However, there is no need for helicase dimerization during eukaryotic DNA replication, arguing against this model (Yardimci et al., 2010). Alternatively, a single CMG complex motors along dsDNA. As DNA emerges from the rear of CMG’s central channel, a proteinaceous pin or “ploughshare” bisects the duplex, leading to strand separation (Takahashi et al., 2005). In a variation of this idea, ssDNA exits through side channels midway along the longitudinal axis of MCM (Brewster et al., 2008). In all of these models, DNA enters the MCM central channel as a dimer.

To understand how the replicative helicase interacts with DNA in S phase, we examined the collision of replisomes with various DNA roadblocks in Xenopus egg extracts. We previously showed that when two replisomes converge on a DNA interstrand crosslink (ICL), the 3′ ends of the leading strands initially arrest 20-40 nucleotides (nt) from the ICL (Raschle et al., 2008). Here, we provide evidence that the footprint of CMG on DNA underlies these distal arrest points. We had also speculated (Raschle et al., 2008) that the wide distribution of arrest points might suggest that CMG translocates along dsDNA, since the first replisome to arrive at the lesion could engulf the ICL and impose a more distal arrest on the second replisome. However, we show here that blocking the arrival of one replisome using biotin-streptavidin (biotin-SA) roadblocks has no effect on the leading strand arrest points of the other converging replisome. To directly examine the CMG translocation mode, we employed strand-specific roadblocks. We reasoned that if CMG translocates along dsDNA, it should be arrested by a bulky roadblock on either the leading or lagging strand template. In contrast, if it moves along ssDNA in the 3′ to 5′ direction, it should be arrested by a roadblock on the leading strand template but not on the excluded, lagging strand template. We found that the nascent leading strand stalls ~30 nucleotides from a biotin-SA complex on the leading strand template, consistent with CMG stalling, whereas a biotin-SA complex on the lagging strand template induced little arrest. Similar results were obtained in a single-molecule assay using strand-specific quantum dot (QDot) roadblocks. Together with previous biochemical analyses, our data strongly suggest that CMG unwinds DNA by translocating along ssDNA in the 3′ to 5′ direction, consistent with DNA unwinding by steric exclusion. The data imply a series of discrete molecular events that underlie the conversion of MCM2-7 to an active CMG in S phase.

RESULTS

To understand how the eukaryotic replicative DNA helicase is configured on DNA in S phase, we examined how the replisome interacts with specific DNA lesions. To this end, we employed nucleus-free Xenopus egg extracts (Walter et al., 1998), which support efficient DNA replication of plasmids or λ DNA. In this system, DNA is first incubated with a high speed supernatant of egg cytoplasm (HSS), which chromatinizes the template and also promotes sequence nonspecific MCM2-7 recruitment to the DNA by ORC, Cdc6, and Cdt1 (Arias and Walter, 2004). Subsequently, a concentrated nucleoglysomal extract (NPE) is added, which supports CMG assembly dependent on Cdc7-Drl1, Cdk2-Cyclin E, Mcm10, and Cdc45 (Takahashi and Walter, 2005; Walter and Newport, 2000; Wohlschlegel et al., 2002). The Xenopus CMG travels with the replisome and unwinds DNA throughout S phase (Pacek et al., 2006; Pacek and Walter, 2004). These observations indicate that nucleus-free Xenopus egg extracts promote activation of the replicative DNA helicase by the same events that occur in cells.

CMG Binding Correlates with Leading Strand Arrest 20 Nucleotides from a DNA Interstrand Crosslink

We first sought to detect the CMG footprint on DNA in S phase. To this end, we examined the collision of the replisome with two different lesions: a DNA inter-strand crosslink, which should arrest CMG, and a DNA intra-strand crosslink, which should stall the DNA polymerase but not CMG. Plasmids containing a site-specific cisplatin inter-strand crosslink (plICL \textsubscript{Intra}; Figure 1A, red line) or 1,2 cisplatin intra-strand crosslink (plICL \textsubscript{Intra}; Figure 1B, red bracket) were incubated sequentially in HSS and NPE containing [α-\textsuperscript{32}P]dATP. At different times after NPE addition, DNA was extracted and digested with Stul, which cuts the plasmid once 288 nt to the right of both lesions (Figure 1A and
Leftward leading strands were visualized on a sequencing gel alongside an appropriate sequencing ladder. On pICL\textsubscript{Inter}, the 3' ends of the leading strand initially stalled 20-40 nucleotides from the ICL (Figure 1C, lane 2, and Figure 1A, red strand) (Raschle et al., 2008). Subsequently, DNA synthesis resumed and the leading strand stalled again 1 nt from the ICL (Figure 1C, green arrow and Figure 1A, green strand) (Raschle et al., 2008). In contrast, during replication of pICL\textsubscript{Intra}, the 3' end of the leading strand immediately advanced all the way to the lesion, where it stalled (Figure 1C, lanes 10–16; Figure 1B, red strand). These data imply that leading strands arrest 20–40 nt from an inter-strand crosslink due to the footprint of the replicative DNA helicase, which we denote as CMG.
To test this idea, we performed chromatin immunoprecipitation (ChIP) for Mcm7, a CMG subunit. pICLLead/Lag was replicated, and at different times, the reaction was crosslinked with formaldehyde, sonicated, immunoprecipitated with Mcm7 antibody, and the recovered DNA amplified with ICL-proximal and control primer pairs (Figure 1D). ChIP revealed that soon after replication began, Mcm7 was depleted from the control region, as expected if CMG travels toward the ICL while displacing any latent MCM2-7 complexes (Figure 1D, purple triangles). In contrast, Mcm7 initially accumulated at the ICL, concurrent with the arrival of leading strands at the −20 position (Figure 1D, compare pink circles and blue diamonds). The subsequent disappearance of Mcm7 from the ICL after 10 min correlated with the disappearance of the −20 to −40 leading strand cluster (Figure 1D), and the advance of the leading strand to the −1 position (Figure 1D, gray squares). These data indicate that arrest of the leading strand 20–40 nt from a cisplatin ICL is caused by the footprint of CMG on DNA (and any dead volume of the DNA polymerase) and that dissociation of CMG facilitates resumption of leading strand synthesis toward the ICL.

Converging Replisomes Do Not Interfere with Each Other at an ICL

We postulated that the wide range of leading strand arrest points near a DNA inter-strand crosslink (Figure 1C, red bracket) could reflect heterogeneity in the CMG footprint and/or interference by converging CMG complexes. In the latter view, the first CMG to arrive at the ICL would impose a more distal stoppage point on the second replisome, giving rise to the distribution of leading strand products. Such interference might be particularly pronounced if CMG travels along dsDNA, since an ICL might enter deep into the central channel of CMG, allowing the first replisome to prevent approach of the second replisome (Figure S2A, top).

To test whether replisome interference occurs, we designed a plasmid, pICLLead/Lag, in which the arrival of replication forks on one side of an ICL can be blocked. pICLLead/Lag contains a nitrogen mustard (NM)-like ICL [which causes a −24 position arrest (Raschle et al., 2008)], as well as four biotinylated thymidine nucleotides placed 34–40 nt to the right of the ICL, two on the leading strand template and two on the lagging strand template of the leftward replication fork (Figure 2A). In the presence of SA, the leftward DNA replication fork should not be able to reach the ICL. To test whether this is the case, pICLLead/Lag was preincubated with and without SA and then replicated in the presence of [±−32P]dATP. At different times, DNA was digested with StuI and nascent strands of the leftward replication fork were visualized on sequencing gels (Figure 2B). In the absence of SA, the leftward leading strand initially advanced to within 24 nt of the ICL (Figure 2B, lane 1, red bracket), after which it crept forward, nearly reaching the ICL (Figure 2B, lanes 5, 7, 9, 11, 13, green arrow), as expected based on our previous results (Raschle et al., 2008). In the absence of SA, the −24 cluster was largely absent, and we observed a new set of products starting at the −70 position (Figure 2B, lanes 2, 4, 6, 8, 10, 12, orange bracket), which is 30 nt from the outermost biotin-SA complex at the −40 position (Figure 2B, bottom white arrow on DNA sequencing ladder). The arrest of the leading strand 30 nt from the biotin-SA complex is similar to the arrest observed at cisplatin (−20) and NM ICLs (−24). After the leading strand paused for 10–15 min at −70, it was further extended to within one nucleotide of the outermost biotin-SA complex located at the −41 position (Figure 2B, lanes 4, 6, 8, 10, 12, 14, black arrow). We infer that the leading strand arrest at −70 is due to the footprint of CMG, which has stalled at the outermost biotin-SA complex, and that the advance of the leading strand to the −41 position reflects CMG dissociation and advance of DNA polymerase ε to the lesion. This interpretation is supported by the fact that purified DNA polymerase ε advanced to within a single nucleotide of a biotin-SA complex in primer extension reactions (Figure S2B). In summary, biotin-SA complexes placed to the right of the ICL efficiently arrest the leftward CMG complex for 10–30 min.

To look for interference of the rightward fork by the leftward fork, pICLLead/Lag was replicated in the presence and absence of SA and then digested with AffIII, which cuts 151 nt to the left and 581 nt to the right of the ICL (Figure 2A). Strikingly, the −24 arrest pattern of the rightward fork was unchanged by the addition of SA (Figure 2C, blue bracket, compare ± SA lanes), even though the leftward fork was efficiently arrested by the roadblock (Figure 2C, even lanes, orange arrow). We conclude that the presence of a replisome on one side of the ICL does not affect the position of the second replisome on the other side of the lesion. This result can be interpreted in two ways. First, CMG is a dsDNA translocase, but the ICL cannot enter its central channel, preventing the first replisome from engulfing the ICL. However, this explanation is unlikely since the NM-like ICL is nondistorting (A.G., Z. T., S.C., and O.D.S., unpublished data) and does not expand the diameter of the duplex (Angelov et al., 2008). The second interpretation is that MCM2-7 translocates along ssDNA, such that both replisomes arrest just before the ICL, avoiding interference (Figure S2A, bottom).

Preferential Bypass of a Lagging Strand Roadblock by the Replicative DNA Helicase

To interrogate directly the translocation mode of CMG on DNA, we confronted the replisome with a roadblock on only the leading or lagging strand templates. To this end, we prepared two derivatives of pICLLead/Lag called pICLLead and pICLLead, which contain biotins on the leading or lagging strand templates, respectively, for the leftward replication fork (Figures 3B and 3C). Using these DNA templates, we could address how the biotin-SA roadblocks placed on the leading or lagging strand templates influence the leftward moving CMG without interference from the rightward moving CMG, whose arrival is prevented by the ICL (Figure 2C). If CMG translocates along ssDNA in the 3′ to 5′ direction, it will stall at a biotin-SA complex located on the leading strand template, yielding the same −70 leading strand arrest as seen on pICLLead/Lag (Figures 3A and 3B). In contrast, a biotin-SA complex located on the lagging strand template might not affect CMG progression since this strand would be excluded from the central channel of the helicase (Figure 3C). In this case, the leftward CMG should only stall once it hits the ICL, manifesting as a −24 arrest (Figure 3C). On the other hand, if CMG translocates along dsDNA, it is predicted to stall
Figure 2. Replisomes Converging on an ICL Do Not Interfere with Each Other
(A) Locations of restriction sites, the nitrogen-mustard-like ICL, biotins, and sequencing primers on pICL\textsuperscript{Lead/Lag}. Figure S2A presents two alternative scenarios for how replication forks might interact at an ICL.
(B) pICL\textsuperscript{Lead/Lag} was preincubated with buffer or streptavidin, as indicated, and replicated in egg extracts in the presence of [\(\alpha\text{-}^{32}\text{P}\)]dATP. At the indicated times after NPE addition, replication intermediates were digested with Stu I and separated on a DNA sequencing gel alongside a sequencing ladder generated with primer M. The distance of products from the ICL is indicated on the left of the gel. White arrows on the DNA sequencing ladder indicate the location of biotins. Red bracket, leading strand arrest 24–50 nt from the ICL in the absence of SA. Orange bracket, leading strand arrest 70–80 nt from the ICL in the presence of SA (30–40 nt from the outermost biotin). Green arrow (\(-1\) position), leading strands that have advanced to the ICL. Black arrow (\(-41\) position), leading strands that have advanced to the outermost biotin-SA complex. Figure S2B shows that DNA polymerase \(\varepsilon\) can advance to within one nt of a biotin-SA complex on the leading strand template.
(C) pICL\textsuperscript{Lead/Lag} was preincubated with buffer or streptavidin, as indicated, and replicated in egg extracts in the presence of [\(\alpha\text{-}^{32}\text{P}\)]dATP. At the indicated times after NPE addition, replication intermediates were digested with AflIII (see Figure 2A) and separated on a DNA sequencing gel alongside a sequencing ladder generated with primer S. Products of the rightward fork are shown. The leftward fork was efficiently arrested by the biotin-SA (see Figure S2C).
at the SA whether the obstruction is located on the leading or lagging strand templates since both strands pass through the central channel of the helicase (Figures 3E and 3F; −70 arrest). In short, only if CMG translocates along ssDNA is the leading strand predicted to reach the −24 position on pICL Llag bound to SA (Figure 3C).

pICL Lead/Lag, pICL Lead, and pICL Llag were replicated separately in the presence or absence of SA, and nascent strands were analyzed after digestion with StuI, as in Figure 2B. When SA was bound to pICL Lead, leading strands arrested at the −70 position, almost exactly as seen for pICL Lead/Lag (Figure 3G, compare lanes 2 and 4, 8 and 10, 14 and 16, 20 and 22). In contrast, when
biotin-SA was located on the lagging strand template (pICL_{Lag}), the result was very different. While a small fraction (maximally 28% of total signal) of leading strands transiently arrested near the −80 position (Figure 3G, lane 6), the majority advanced directly to the −24 position (Figure 3G, lanes 6, 12, 18, and 24, red bracket). Therefore, CMG bypasses a lagging strand roadblock much more readily than a leading strand roadblock.

Using gel shift analysis, we verified that pICL_{Lead} and pICL_{Lag} bound equally to SA (Figure S3A). To assess the retention of biotin-SA complexes on DNA after replication in egg extracts, we immunoprecipitated SA and quantified the associated plasmid by real-time PCR. Importantly, pICL_{Lag} was recovered 28% of total signal) of leading strands transiently arrested near the result was very different. While a small fraction (maximally 28% of total signal) of leading strands transiently arrested near the −80 position (Figure 3G, lane 6), the majority advanced directly to the −24 position (Figure 3G, lanes 6, 12, 18, and 24, red bracket). Therefore, CMG bypasses a lagging strand roadblock much more readily than a leading strand roadblock.

Using gel shift analysis, we verified that pICL_{Lead} and pICL_{Lag} bound equally to SA (Figure S3A). To assess the retention of biotin-SA complexes on DNA after replication in egg extracts, we immunoprecipitated SA and quantified the associated plasmid by real-time PCR. Importantly, pICL_{Lag} was recovered 28% of total signal) of leading strands transiently arrested near the result was very different. While a small fraction (maximally 28% of total signal) of leading strands transiently arrested near the −80 position (Figure 3G, lane 6), the majority advanced directly to the −24 position (Figure 3G, lanes 6, 12, 18, and 24, red bracket). Therefore, CMG bypasses a lagging strand roadblock much more readily than a leading strand roadblock.

Using gel shift analysis, we verified that pICL_{Lead} and pICL_{Lag} bound equally to SA (Figure S3A). To assess the retention of biotin-SA complexes on DNA after replication in egg extracts, we immunoprecipitated SA and quantified the associated plasmid by real-time PCR. Importantly, pICL_{Lag} was recovered 28% of total signal) of leading strands transiently arrested near the −80 position (Figure 3G, lane 6), the majority advanced directly to the −24 position (Figure 3G, lanes 6, 12, 18, and 24, red bracket). Therefore, CMG bypasses a lagging strand roadblock much more readily than a leading strand roadblock.

Using gel shift analysis, we verified that pICL_{Lead} and pICL_{Lag} bound equally to SA (Figure S3A). To assess the retention of biotin-SA complexes on DNA after replication in egg extracts, we immunoprecipitated SA and quantified the associated plasmid by real-time PCR. Importantly, pICL_{Lag} was recovered 28% of total signal) of leading strands transiently arrested near the −80 position (Figure 3G, lane 6), the majority advanced directly to the −24 position (Figure 3G, lanes 6, 12, 18, and 24, red bracket). Therefore, CMG bypasses a lagging strand roadblock much more readily than a leading strand roadblock.

Using gel shift analysis, we verified that pICL_{Lead} and pICL_{Lag} bound equally to SA (Figure S3A). To assess the retention of biotin-SA complexes on DNA after replication in egg extracts, we immunoprecipitated SA and quantified the associated plasmid by real-time PCR. Importantly, pICL_{Lag} was recovered 28% of total signal) of leading strands transiently arrested near the −80 position (Figure 3G, lane 6), the majority advanced directly to the −24 position (Figure 3G, lanes 6, 12, 18, and 24, red bracket). Therefore, CMG bypasses a lagging strand roadblock much more readily than a leading strand roadblock.
neither rightward nor leftward forks should be able to bypass the QDot (Figures 4B, and 4Bii).

In agreement with the 3’ to 5’ ssDNA translocation model, we found that all rightward DNA replication forks stalled at the QDot (Figure 4B), whereas 80% of leftward forks bypassed the QDot (Figure 4Biii). We also attached the QDot to the “top” strand, 15 kb from the end of λ DNA (Figure 4C, green dot). On this DNA template, all polarities are reversed. Thus, if CMG is a 3’ to 5’ ssDNA translocase, leftward forks should stall at the QDot (Figure 4Ci) and rightward forks should bypass it (Figure 4Cj). In agreement with this prediction, 93% of leftward forks stalled at the QDot (Figure 4Ci), whereas 74% of rightward forks bypassed it (Figure 4Cj). The fact that 20%–25% of forks stalled at a QDot on the lagging strand template suggests that while CMG is a 3’ to 5’ ssDNA translocase, it can be transiently arrested by a lesion on the excluded strand (see Discussion).

**DISCUSSION**

This study addresses whether the eukaryotic replicative DNA helicase (CMG) travels along ssDNA or dsDNA in S phase, a fundamental, unanswered question in cell biology. Our strategy was based on the supposition that both 3’ to 5’ ssDNA and dsDNA translocases should arrest at an obstruction on the leading strand template (Figures 3B and 3E), whereas only a 3’ to 5’ ssDNA should be able to bypass a roadblock on the lagging strand template (Figure 3C versus Figure 3F). Using the arrest point of the leading strand as a marker for CMG stalling, we observed that the helicase was arrested much more efficiently by a leading strand biotin-SA roadblock than a lagging strand roadblock. Analogous results were obtained using single-molecule assays, in which replisomes were collided with QDots. These data support the idea that CMG translocates along ssDNA in the 3’ to 5’ direction. We cannot formally rule out an alternative interpretation of the data in which MCM2-7 translocates along dsDNA and transient opening of the helicase central channel allows bypass of a lagging strand roadblock. However, this model is improbable, principally because it does not explain how the breached helicase could bypass the lesion (see Figure S4 and legend thereof).

Importantly, other aspects of our data are also consistent with CMG being a ssDNA translocase. First, we observed no interference between converging replisomes at an ICL, as might have been expected in the dsDNA translocation model (see Results). Second, the apparent size of the CMG footprint is consistent with a ssDNA translocation model. Thus, upon initial collision of the replisome with a cisplatin ICL, the 3’ end of the leading strand can advance to within 20 nucleotides of the lesion. Given that the longitudinal axis of the eukaryotic MCM2-7 complex comprises ~115 Å (Remus et al., 2009), MCM2-7 should protect ~34 basepairs of dsDNA (115 Å ÷ 3.4 Å per basepair = 34 basepairs), which likely occurs in the context of prereplication complexes. Importantly, ssDNA can be extended to 1.7 times the length of B-form DNA (Smith et al., 1996; van Oijen, 2007). Therefore, 20 nt (34 × 1.7) of fully extended ssDNA would suffice to traverse the entire MCM2-7 channel. Thus, a ssDNA translocation model is compatible with the ~20 nt arrest point seen for the leading strand if the active site of DNA polymerase is located immediately behind the rear exit channel of CMG. Although stretching ssDNA to this extent is energetically unfavorable, this might occur as a result of interactions between DNA and the CMG central channel and/or the pulling force exerted by DNA polymerase. It is also worth noting that most of the leading strands arrest further than 20 nt from the ICL, indicating that only a small fraction of stalled CMG complexes encircle fully extended ssDNA. The reason why leading strands arrest 4 nucleotides further away from a nitrogen-mustard like ICL versus a cisplatin ICL (Raschle et al., 2008) is likely that the former lesion stabilizes the duplex, preventing full approach to the lesion by CMG. In the case of SA roadblocks, a steric clash between CMG and SA might explain the more distal arrest point (~30 nt) observed for this lesion.

Finally, our conclusion that the replicative helicase moves along ssDNA is supported by extensive biochemical analysis of purified MCM complexes. Archaeal MCMs (Kelman et al., 1999; Shechter et al., 2000), Mcm4/6/7 (Ishimi, 1997), and CMG (Moyer et al., 2006) exhibit 3’ to 5’ ssDNA translocation activity on model DNA templates. In light of our findings, the observation that MCM complexes can slide (Evrin et al., 2009; Remus et al., 2009) or actively translocate along dsDNA (Kaplan et al., 2003) could have other implications. As previously proposed (Remus et al., 2009), the collision of a replicosome with latent MCM2-7 complexes could induce the latter to slide or translocate along dsDNA ahead of the fork. These mobilized MCMs might represent a cadre of reserve helicases that are deployed to rescue stalled replication forks (Remus et al., 2009), or they might serve to increase the local concentration of pre-RCs in unreplicated DNA to potentiate an increased initiation frequency late in S phase (Lucas et al., 2000).

In summary, the data are most consistent with CMG translocation along ssDNA in the 3’ to 5’ direction. The evidence thus indicates that CMG unwinds DNA via steric exclusion, and it strongly disfavors models in which dsDNA passes through any part of the CMG central channel. Future experiments will be required to address the precise molecular mechanism by which CMG translocates along ssDNA and how this leads to strand separation.

**Implications for CMG and Replisome Architecture**

Our results have implications for the overall architecture of the DNA replication fork. As discussed above, the ~20 nt leading strand arrest point matches precisely what is expected if ssDNA is fully extended within the central channel of MCM2-7. Barring rapid dissociation of replication proteins upon collision with the ICL or a major difference in the dimensions of yeast and Xenopus MCM2-7, this observation implies that the C-terminal ATPase domains of the MCM2-7 complex comprise the leading edge of the replisome and that Cdc45, GINS, and any other helicase-associated factors do not substantially alter the footprint of MCM2-7 on DNA. Moreover, the data imply that the active site of DNA polymerase epsilon resides immediately behind the rear exit channel of CMG, which is consistent with our finding that purified pol ε can advance to within 1 nucleotide of a biotin-SA complex on the leading strand template. This configuration is advantageous, as emerging ssDNA will be immediately replicated, minimizing the possibility of strand reannealing or cleavage.
We noted that when a biotin-SA complex was placed on the lagging strand template, there was a low level of transient stalling (Figure 3G, lane 8). Similarly, in single-molecule analysis, QDots placed on the lagging strand template induced ~20% stalling. These results suggest that the excluded strand might interact intimately with the outer face of the CMG, leading to steric clashes with the bulky lesions that cause occasional or transient arrest. Recent structural analysis of the *Drosophila* CMG complex supports this conclusion since it reveals a potential groove/channel on the outer surface of CMG that might interact with the excluded strand (Costa et al., 2011).

**Interactions of the Replisome with DNA Damage**

Our data have implications for the early steps of DNA interstrand crosslink repair, which is initiated by the collision of two DNA replication forks with an ICL in Xenopus egg extracts (Knipscheer et al., 2009; Raschle et al., 2008). It was previously unknown why leading strands pause at the ~20 to ~40 positions and then undergo further elongation toward the ICL, followed by lesion bypass. Our experiments indicate that the delay involves the dissociation of CMG from the site of the lesion. Another implication is that when the replisome first arrives at an ICL, the DNA immediately adjacent to the lesion is single stranded since one strand is sequestered within the CMG central channel while the other strand is excluded.

**Implications for the Mechanism of CMG Assembly**

In G1, latent MCM2-7 double hexamers appear to interact with dsDNA, as evidenced by the absence of ssDNA within pre-RCs (our unpublished results; Bowers et al., 2004; Geraghty et al., 2000) and the ability of latent MCM2-7 complexes to slide along DNA (Evrin et al., 2009; Remus et al., 2009). Our data indicate that in S phase, CMG encircles ssDNA with no duplex DNA remaining inside the central helicase. Furthermore, CMG likely functions as a monomer (Gambus et al., 2006; Ilves et al., 2010; Moyer et al., 2006; Yardimci et al., 2010). Assuming these starting and end points, a number of discrete stages can be envisioned that convert MCM2-7 to CMG, not necessarily in the following order (Figure 5). (1) The MCM2-7 double hexamer is split. (2) The latent MCM2-7 rings straddling dsDNA are opened, perhaps via regulation of the Mcm2-Mcm5 gate (Bochman and Schwacha, 2008). Both events might depend on MCM2-7 phosphorylation by DDK. (3) One strand of the DNA duplex is extruded from the central channel, perhaps by the binding of Mcm10 or Sld2 to ssDNA (Kanter and Kaplan, 2010; Warren et al., 2008). (4) The gate is reclosed and the helicase motor is jump-started by Cdc45 and GINS (Ilves et al., 2010). Whatever the precise sequence of events, the reconfiguration of the MCM2-7 complex from a dsDNA binding mode in G1 to a ssDNA binding mode in S phase helps explain why so many initiation factors are required to assemble CMG. This view of initiation establishes a roadmap for future investigations into the functions of replication initiation factors. Notably, the DNA tumor virus initiator proteins E1 and possibly Large T Antigen initially bind to dsDNA within the viral origin of replication but unwind DNA via steric exclusion (Enemark and Joshua-Tor, 2006). Thus, the transition from dsDNA to ssDNA binding appears to be widely conserved among eukaryotic replicative DNA helicases.

**EXPERIMENTAL PROCEDURES**

**Preparation of Plasmids**

To make pICLLead/Lag, pICLLead, and pICLLAG, various ICL-biotin oligonucleotides (see Supplementary Methods) were purified by polyacrylamide gel electrophoresis, and ligated into pSVRLuc to form the three plasmids (see Figure 3) (Guainazzi et al., 2010; Raschle et al., 2008). pICLControl contains the identical sequence as pICLLead/Lag, except for the four biotinylated thymidine nucleotides. pICLExtr and pICLIntra were constructed as previously described (Raschle et al., 2008; Tremeau-Bravard et al., 2004). Compared to pICLExtr and pICLIntra, the plasmids pICLControl, pICLLead/Lag, pICLLead, and pICLExtr contain a 41 nt insert near the ICL.

**Xenopus Egg Extracts and Replication**

The preparation of Xenopus egg extracts (NPE and HSS) was as described (Walter et al., 1998). For DNA replication, the plasmids (75 ng/μl) were first incubated with an equal volume of Streptavidin (5 μg/μl) (SouthernBiotech, Birmingham, AL, USA) or buffer for 1 hr at room temperature (RT), after which this mixture was added to HSS for 5 min at 22°C. The nascent strand analysis was carried out as described (Raschle et al., 2008).

**Nascent Strand Analysis**

The nascent strand analysis was carried out as described (Raschle et al., 2008). Briefly, purified DNA replication products were isolated and digested with the indicated enzymes. Restriction fragments were separated on 5% or 7% polyacrylamide sequencing gels. Gels were transferred to filter paper, dried, and nascent strands visualized with a phosphorimager. Sequencing ladders using primers S and M (see Figure 2A) were generated using the

---

**Figure 5. Model for Replication Initiation**

Model for helicase activation in which MCM2-7 encircles dsDNA in G1 and ssDNA in S phase. See text for details.
replication would yield average dig-dUTP tracts of 6.7 kb (25 min).

These experiments was 268 bp/minute (Yardimci et al., 2010), uninterrupted.

were labeled with dig-dUTP for 25 min. Since the average fork rate in

describe (Yardimci et al., 2010). SYTOX, fluorescein labeled anti-dig, and

egg extracts and visualization of replicated products was performed as

replication fork arrest near QDots, replication forks

QDot was imaged using 568 nm, 488 nm, and 405 nm laser light,

described (Yardimci et al., 2010). SYTOX, fluorescein labeled anti-dig, and

ACKNOWLEDGMENTS

SUPPLEMENTAL INFORMATION

Supplemental Information contains Extended Experimental Procedures and four figures and can be found with this article online at doi:10.1016/j.cell.2011.07.045.

REFERENCES


EXTENDED EXPERIMENTAL PROCEDURES

Preparation of pICL\textsuperscript{Lead/Lag}, pICL\textsuperscript{Lead}, and pICL\textsuperscript{Lag}

To make pICL\textsuperscript{Lead/Lag}, pICL\textsuperscript{Lead} and pICL\textsuperscript{Lag}, two oligonucleotides (M1: 5'-CCCTGTCACTGGTAG\*ACAGCATTGGAATTC-GAATTCCC

CCCTTTTTCAACAT\*GGAAT\*GCAGACTGGC-3' and M2: 5'-GCACGCCAGTGCACTGAT\*TCCAT\*GTTTAAAAGAAAAGGGGGGAA

ATTCCATGCTG*TCTACCAGTGAC-3') containing a single modified guanosine base, 7-deaza-7-(2,3-dihydroxy-propyl)-guanine (G*), were annealed and processed to generate the ICL (Angelov et al., 2009). M1 (for pICL\textsuperscript{Lead}), M2 (for pICL\textsuperscript{Lag}), or M1 and M2 (for pICL\textsuperscript{Lead/Lag}) were internally biotinylated at two positions (T*) by incorporation of a Biotin-dT phosphoramidite (Glen Research, Sterling, VA, USA). The ICL-biotin duplex was purified by polyacrylamide gel electrophoresis, extracted from the gel by electroelution, and ligated into pSVRLuc to form the three plasmids. Compared to pICL\textsuperscript{Inter} and pICL\textsuperscript{Intra}, the plasmids pICL\textsuperscript{Control}, pICL\textsuperscript{Lead/Lag}, pICL\textsuperscript{Lead}, and pICL\textsuperscript{Lag} contain a 41 nt insert near the ICL (underlined sequence in M1 and M2), such that the StuI site is located at different positions on the two groups of plasmids (–268 versus –309).

Chromatin Immunoprecipitation

After the addition of NPE, 3 \( \mu \)l aliquots of the reaction were crosslinked through the addition of 47 \( \mu \)l 1% formaldehyde in ELB (10 mM HEPES-KOH, [pH 7.7], 2.5 mM MgCl\(_2\), 50 mM KCl, 250 mM sucrose, and 1 mM DTT). After 10 min incubation at RT, 5 \( \mu \)l of 1.25M glycine was added to stop the crosslinking reaction. The crosslinked material was spun through Micro Bio-Spin 6 Chromatography Columns (BIO-RAD, Hercules, CA, USA). The flow-through was diluted to 500 \( \mu \)l with sonication buffer (20mM Tris 7.5, 150 mM NaCl, 2 mM EDTA, 0.5% NP-40, 5 mg/mL Aprotinin+Leupeptin, 2 mM PMSF) and subjected to sonication until the average final size of DNA fragments is \( \sim \)300-500 bp. Immunoprecipitation with the indicated antibodies and quantitative real-time PCR were performed as described in Pacek et al., 2006.

Primer Extension Assay

The primer extension assay was modified from Bermudez et al., 2010. A 5\(^{\prime}\)-\(^{32}\)P-labeled primer (5'-GCGTTGGCCGATTCAT TAATGCA-3') was anneal to a 68-mer oligonucleotide (5'-CTTTTTAACAT\*GGAAT\*GCAGACTGGCGTGCGCGGCCGCGATCC GCTGCATTAATGCACTGCGCCCAAAGC-3'), which contains two biotinylated thymidines (T*), forming the biotin template. The control template was identical to the biotin template except that it lacked biotinylated thymidine. The templates (0.5 \( \mu \)M) were incubated with 8 \( \mu \)M Streptavidin or buffer for 1 hr at room temperature. All primer extension reactions were performed in 10 \( \mu \)l containing 10 nM \( \lambda \)DNA and 100 \( \mu \)M dNTPs. The reaction for Taq polymerase contained 10mM Tris-HCl (pH 8.3), 50mM KCl, 1.5 mM MgCl\(_2\), and 4 units Taq polymerase (New England Biolabs, Ipswich, MA, USA). Reactions were incubated at 72\(^{\circ}\)C for the indicated time. Reactions with recombinant human DNA polymerase epsilon (Bermudez et al., 2011) contained 20mM Tris (pH7.5), 1mM DTT, 100 ug/ml BSA, 10 mM Mg acetate, 30mM potassium glutamate, 5nM RFC, 50 nM PCNA, 5 nM DNA polymerase epsilon and were performed at 37\(^{\circ}\)C. The reactions were halted with 5 \( \mu \)l loading buffer (95% Formamide, 18mM EDTA, and 0.025% SDS) and heated to 95\(^{\circ}\)C for 3-4 min. The extension products were separated on a 7% polyacrylamide sequencing gel and visualized by autoradiography.

Attaching a QDot to \( \lambda \) DNA

To attach a QDot to the top strand 15 kb from the right end of \( \lambda \) DNA, \( \lambda \) DNA was nicked at positions 33779 and 33791 on the top strand with Nt.BstNBI (New England BioLabs, Ipswich, MA, USA). The oligonucleotide 5'-'TTTCAAGAGT\*CTGAC-3' (Biosynthesis Inc., Lewisville, TX, USA), which contains a digoxigenin at a single thymidine base (T*), was annealed and ligated to the nicked \( \lambda \) DNA. Alternatively, for QDot labeling of the bottom strand 19 kb away from the end, we used Nb.BsrDI which nick \( \lambda \) DNA at positions 29675 and 29707 on the bottom strand and the oligonucleotide 5'-'CATGCTGATACCGTACGTTAGCTGAAACGACATA-3' containing a digoxigenin at T*.

SUPPLEMENTAL REFERENCES


Figure S1. Mapping Rightward Leading Strands Near DNA Interstrand Crosslinks, Related to Figure 1

(A) Location of restriction sites, the cisplatin ICL, and sequencing primers on piCL\textsuperscript{inter} and piCL\textsuperscript{intra} (which differ by only a few nucleotides, as indicated by the blue letters in Figures 1A and 1B).

(B) piCL\textsuperscript{inter} was replicated in egg extracts in the presence of [\(\alpha\)-\textsuperscript{32}P]dATP. At the indicated times after NPE addition, replication intermediates were digested with AflIII and separated on a DNA sequencing gel alongside a sequencing ladder generated with primer S. Blue bracket, leading strand stalling between the –20 and –40 positions. Grey arrow, leading strand stalling at the –1 position.
Figure S2. CMG Complexes Converging on an ICL Do Not Interfere with Each Other, Related to Figure 2

(A) Model for the convergence of two replisomes on an ICL. Outcomes of replisome convergence assuming dsDNA translocation (top) or 3' to 5' ssDNA translocation (bottom) are depicted. The yellow line in the top panel represents the pin postulated to split the duplex as it emerges from the central channel.

(B) The /C0 arrest cannot be explained by the footprint of DNA polymerase ε. (top) Cartoon depicting the biotinylated primer-template used for extension with DNA pol ε. The extension product generated by DNA pol ε is shown in blue. Asterisk, location of radioactive label. (bottom) The primer template shown above or an unbiotinylated control template was incubated with streptavidin or buffer and then incubated with Taq DNA polymerase at 72°C or DNA pol ε at 37°C, as indicated. At different times, extension products were analyzed on a 7% sequencing gel. The length of key extension products is indicated on the left of the gel. Blue arrow, 52 nt extension products.

(C) The top half of the autoradiograph shown in primary Figure 2C is presented. Arrest of the leftward leading strand at the /C0/-80 position and the /C0 position are indicated by orange and black arrows, respectively. Green arrow, approximate position of ICL. Light blue arrow, extension product. The data show that the leftward fork is efficiently delayed at the /C0 position by the biotin-SA complex.
Figure S3. SA Binds with Similar Efficiency to pICL\textsuperscript{Lead/Lag}, pICL\textsuperscript{Lead}, and pICL\textsuperscript{Lag} Plasmids, Related to Figure 3

(A) To verify that pICL\textsuperscript{Lead} and pICL\textsuperscript{Lag} bound equally to SA, we performed gel shift analysis. pICL\textsuperscript{Control}, pICL\textsuperscript{Lead/Lag}, pICL\textsuperscript{Lead} and pICL\textsuperscript{Lag} were preincubated with streptavidin or buffer and digested with StuI and AflIII, which yields a 0.46 kb DNA fragment containing the biotin locus (green arrow). The digestion products were separated on a 1% native agarose gel. SA caused efficient gel retardation of the 0.46 kb DNA fragments from pICL\textsuperscript{Lead/Lag}, pICL\textsuperscript{Lead}, and pICL\textsuperscript{Lag} (see lanes 4, 6, 8, red and blue arrows). The SA-induced shift of the 0.46 kb fragment from pICL\textsuperscript{Lead/Lag} (red arrow) was greater than the shift for the fragments from pICL\textsuperscript{Lead} and pICL\textsuperscript{Lag} (blue arrow), likely because the former contains 4 biotins instead of 2.

(B) Quantification of SA bound to plasmids during replication. pICL\textsuperscript{Control}, pICL\textsuperscript{Lead/Lag}, pICL\textsuperscript{Lead}, and pICL\textsuperscript{Lag} were preincubated with streptavidin and replicated in egg extracts. At the indicated times after NPE addition, reaction products were immunoprecipitated with anti-SA antibody. The quantity of plasmid associated with SA was quantified by qPCR. The average of three independent experiments was graphed. Error bars indicate standard deviations.

(C) Same as (B), except that NPE contained 160 \(\mu M\) free biotin trap (the concentration of SA in the extract was 5 \(\mu M\)). The average of three independent experiments was graphed. Error bars indicate standard deviations.

(D) Mapping of leading strands in the experiment shown in (C). pICL\textsuperscript{Control}, pICL\textsuperscript{Lead/Lag}, pICL\textsuperscript{Lead}, and pICL\textsuperscript{Lag} were replicated as in (C) but in presence of [\(\alpha\text{-}^{32}\text{P}\)]dATP. Replication products were digested with StuI, and analyzed on a 7% sequencing gel. The distance of products from the ICL is indicated on the left of the gel. White arrows on the DNA sequencing ladder indicate the location of biotins. Orange bracket, leading strand arrest 70-80 nt from the ICL in the presence of SA.
In order to explain the bypass of a lagging strand roadblock by a dsDNA translocase, the following model would have to be invoked. When CMG collides with a Biotin-SA complex on the lagging strand template, the central channel of the helicase opens along its entire length (A). Subsequently, the breached helicase motors past the roadblock (B). Finally, the helicase channel recloses to enable continued unwinding by the enzyme (C). For the following reasons, this model is improbable. If the pin or “ploughshare” (blue bar) remains lodged between the two strands of the duplex, the bypass will require strand separation. It is unlikely that a helicase with an open channel could still carry out unwinding since interruption of the ring will disrupt coordinated cycles of ATP hydrolysis. In addition, the rotation of the helicase around DNA that occurs during unwinding would be prevented by the presence of a large steric obstacle on one strand. This mechanism also does not readily explain why only a lagging strand roadblock can be bypassed by CMG. A related model to the one presented in the figure is that after the helicase opens upon encountering the roadblock, the pin disengages (not shown). Although this mechanism would facilitate movement past the roadblock because unwinding by the breached helicase is not required, reactivation of the helicase downstream of the roadblock now requires reinsertion of the pin into the duplex. How this could occur is unknown. Indeed, our single molecule analysis showed that lesion bypass is independent of Cdk2 activity (Figure 4), which promotes a late step in replication initiation (Walter, 2000). Therefore, the pin would have to be reinserted via a different mechanism from the one that normally operates during initiation. In summary, neither mechanism discussed here involves a plausible scenario for how a lagging strand roadblock could be bypassed.