A quest for coordination among activities at the replisome

Nitin Kapadia and Rodrigo Reyes-Lamothe
Department of Biology, McGill University, 3649 Sir William Osler, H3G 0B1

1To whom correspondence should be addressed at: rodrigo.reyes@mcgill.ca

Abstract
Faithful DNA replication is required for transmission of the genetic material across generations. The basic mechanisms underlying this process are shared among all organisms: progressive unwinding of the long double-stranded DNA; synthesis of RNA primers; and synthesis of a new DNA chain. These activities are invariably performed by a multi-component machine called the replisome. Detailed description of this molecular machine has been achieved in prokaryotes and phages, with the replication processes in eukaryotes being comparatively less known. However, recent breakthroughs in the in vitro reconstitution of eukaryotic replisomes have resulted in valuable insight into their functions and mechanisms. In conjunction with the developments in eukaryotic replication, an emerging overall view of replisomes as dynamic protein ensembles is coming into fruition. The purpose of this review is to provide an overview of the recent insights into the dynamic nature of the bacterial replisome, revealed through single-molecule techniques, and to describe some aspects of the eukaryotic replisome under this framework. We primarily focus on Escherichia coli and Saccharomyces cerevisiae (budding yeast), since a significant amount of literature is available for these two model organisms. We end with a description of the methods of live cell fluorescence microscopy for the characterization of replisome dynamics.

Introduction
Every cell needs to replicate its DNA for the genetic material to be passed down to subsequent generations. The cellular machinery responsible for this task is a multi-protein complex known as the replisome. Replisomes begin their journey at origins of replication, and along their path they can encounter various obstacles such as lesions, DNA secondary structures, protein-DNA roadblocks, and the transcription machinery, to name a few (1). It is likely that replisomes achieve high processivity, and ensure high fidelity in DNA replication, by making use of some level of compositional and architectural plasticity – through the dynamics of individual replisomal proteins.

Replisomes have been extensively studied, using diverse approaches that include biochemical ensemble approaches and single-molecule fluorescence microscopy, and various model systems ranging from bacteriophages to mammalian cells. While replisomes in different organisms can have different evolutionary ancestries, they all share basic functions: unwinding duplex DNA, followed by priming and elongation of nascent DNA (2). Below, we describe the E. coli replisome and highlight how recent advances in single-molecule microscopy have helped shape our understanding of its dynamic nature. We then describe the Saccharomyces cerevisiae (budding yeast) replisome, as an eukaryotic model, and compare it to its bacterial counterpart. We end with a discussion on how live-cell single-molecule fluorescence microscopy may be used to advance our understanding of DNA replication in eukaryotes.
**E. coli replisome**

The *E. coli* replisome is composed of at least 12 different proteins that arrange into the following stable subcomplexes: the helicase (DnaB), the primase (DnaG), the DNA polymerase III (Pol III), the processivity factor β clamp, the clamp loader, and the single strand binding protein SSB (SSB₄) (3, 4) (Fig. 1). Multiple protein-protein interactions among these subcomplexes result in the formation of a compact protein structure. DnaB and DnaG interact with each other (5). Up to three copies of Pol III and the clamp loader are physically coupled by the interaction between α and τ subunits, forming the Pol III* subcomplex (6). The interaction between Pol III* and β clamp form the Pol III holoenzyme (HE) (7). Finally, the τ subunit mediates the interaction between DnaB helicase and Pol III HE (8). This architecture is thought to facilitate the coordination of activities during DNA synthesis, specially at the lagging strand: primer synthesis, clamp loading, and recruitment of the Pol III – potentially the same copy that synthesized the previous Okazaki fragment – can rapidly succeed each other. Replisome architecture also increases the chances of simultaneous synthesis at both strands, due to the physical connection between multiple copies of Pol III. Coupling of these activities was suggested in the trombone model (9), which likens the growth of the single-strand DNA (ssDNA)-loop between the Pol III and the helicase at the lagging strand to the movement of a trombone’s slide.

![Figure 1 – Model of the replisome](image)

(A) *E. coli* replisome. The diagram shows a model for the architecture of the bacterial replisome. The homohexameric DnaB helicase opens the double stranded DNA, generating a forked DNA structure. The DnaG primase interacts with DnaB. The τ subunit, which is part of the clamp loader, interacts with DnaB and with the DNA Pol III, serving...
as a link between these two subassemblies. Multiple copies of τ in the clamp loader result in the oligomerization of Pol III, allowing for coupled synthesis of the two strands. A copy of the β clamp binds to each engaged copy of Pol III. At the end of each Okazaki fragment, the Pol III at the lagging strand unbinds from β clamp, which remains bound on DNA. SSB covers the single strand DNA that accumulates during Okazaki fragment synthesis. (B) Budding yeast replisome. The diagram shows a model for the architecture of the yeast replisome. The functional helicase (called CMG) is composed of the heterohexamer MCM2-7, the heterotetramer GINS complex, and CDC45. At the leading strand, Pol ε interacts with the CMG helicase through GINS. At the lagging strand, Pol α interacts with the homotetramer CTF4, which in turn interacts with the CMG helicase. In contrast, Pol δ which is expected to work at the lagging strand has no reported connection with the CMG. MCM10 is thought to have a diverse role at the replisome, not discussed here. Mrc1-Tof1-Csm3 (MTC) complex and RFC are omitted for clarity.

The physical link between enzymes at the replisome has led to the assumption that their activities are actively synchronized. However, an emerging perspective rejects this idea, and instead proposes that near-synchronous synthesis can still be achieved with stochastic sampling of activity rates by individual copies of enzymes (discussed in (10)). An example of these two views – deterministic and stochastic – is found in the models that match the DNA synthesis rates at the leading and lagging strand. A number of deterministic models have proposed mechanisms that increase the synthesis rate of the Pol III at the lagging strand, or that decrease the rate at the leading strand, to compensate for the time invested in priming, clamp loading, and Pol III recruitment (10, 11). In contrast, a stochastic model proposes that the rate of DNA synthesis at either strand is assigned stochastically and independently of the rate of the other strand (10, 12). The evidence for the stochastic model comes from visualizing synthesis of single DNA molecules, through single-molecule microscopy, where individual traces indicated that synthesis was interspersed with stochastic pauses, followed by a switch in the rate of synthesis, independent from the previous rate (12). In this model, variance in the rates will frequently cause one of the two strands to trail behind, despite similar overall average synthesis rates. However, switches in the rates of synthesis can potentially reverse this discrepancy, leading to an averaged effect of near-synchronous synthesis.

A second shift in our understanding of the replisome has come from studying the binding stability among subunits, also through single-molecule microscopy. Until recently, it was assumed that once assembled, the helicase and Pol III* would act processively for hundreds of kilobase pairs. Studies of single replisomes using rolling-circle replication – which consists of a double-stranded circular template with a single-stranded tail, where the replisome can cycle potentially an unlimited number of times – demonstrated processivities of over 40kbp, but with individual instances reaching lengths of few hundreds kbp (13-15). Subsequent experiments using live-cell single-molecule fluorescence microscopy have shown that turnover is much more frequent than previously appreciated, with dissociation occurring every few seconds, indicating a much lower processivity of Pol III* (16-18). Frequent turnover can only be observed in vitro when an excess of Pol III* is present in the buffer (17), which explains why it was not detected in previous studies. Turnover is more rapid during active replication, suggesting that DNA synthesis may exert physical strain on the replisome, and result in dissociation of some of its components. This model matches the proposed torsion buildup on DNA as result of the physical connection between active polymerases at the two strands (19). A dynamic composition of the replisome might also be advantageous to facilitate lesion skipping on damaged DNA - suggesting that the replisome has an
The inherent ability to bypass lesions - and to coordinate handover of the elongating strand to specialized polymerases for damage bypass (20).

**Saccharomyces cerevisiae replisome**

Many of the eukaryotic replisome subunits do not share common ancestry with bacteria (2). However, since they accomplish similar functions, it is unknown the extent at which the strategies they used are shared. The core eukaryotic replisome is composed of over 30 different proteins that include the CMG helicase (composed of CDC45, MCM2-7, GINS), Pol ε, Pol δ, Pol α, the structural protein CTF4, the processivity factor proliferating nuclear cell antigen (PCNA), the clamp loader replication factor C (RFC) and the single-stranded binding protein, replication protein A (RPA) (21) (Fig. 2B). Multiple other proteins interact with the replisome (described in (21)), known as the replication progression complex (RPC), but here we focus on the minimal composed replisome for conciseness. Pol ε and Pol δ are the main replicative polymerases, while Pol α is responsible for synthesizing RNA/DNA primers, primarily on the lagging-strand. Multiple sources of evidence suggest that Pol δ and Pol ε synthesize the lagging strand and leading strands, respectively, although alternative models exist which assign a role at the leading strand to Pol δ (22-33). Reminiscent of the bacterial replisome, Pol ε forms a stable interaction with the GINS subunit of CMG helicase, (28, 34) while Pol α binds to CTF4, which in turn binds to GINS (35), suggesting there is some connection between the leading-strand and lagging-strand components. However, there is no reported interaction between Pol δ and the CMG (28, 34, 36, 37). Furthermore, it is believed that the RFC clamp loader only acts transiently at the replisome, without forming stable interactions that could bridge it to the CMG helicase.

The current model for the architecture of the eukaryotic replisome with Pol δ being physically disconnected, suggests replisome dynamics that differ to the bacterial replisome, and leads to question how excess ssDNA exposure is minimized when there is no coupling between the helicase and the lagging-strand polymerase. It also leaves unexplained how synchronous synthesis between the two strands, reported using both *in vivo* and *in vitro* approaches, is achieved (26, 38, 39).

A possible explanation is that the slow progression of the replication fork, estimated at ~25bp/s in budding yeast (40), may provide enough time for the efficient recruitment of RFC and subsequently Pol δ, and that a much faster rate of synthesis by Pol δ can compensate for the delay. In contrast to bacteria, where the same polymerase synthesizes the two strands, faster synthesis by Pol δ over Pol ε could be explained by their difference in structure. This model also requires that the time needed for the recruitment of a new copy of RFC and Pol δ at the replication fork, which is given by their inherent diffusion coefficients and the number of available copies, must be short. However, in physiological conditions, Pol δ synthesizes at a rate of ~50bp/s, which is not sufficiently different from the speed of the replication fork (41). Furthermore, the reported abundance of Pol δ in the cell suggests that it is not high enough for efficient lagging-strand synthesis, especially at the peak of S-phase when hundreds of replisomes are estimated to be active (42-45). Thus, it seems unlikely that the rate of synthesis by Pol δ can compensate for the additional events at the lagging strand. It is nevertheless conceivable that some aspects of DNA replication in cells may minimize DNA synthesis interruptions even in scenarios dependent on the stochastic recruitment of Pol δ. One such possible mechanism is the proposed clustering of replication origins (46, 47), which could help recycle these proteins from neighbouring replisomes and facilitate their recruitment at the fork by increasing their local concentration.
In eukaryotes, DNA is wrapped around histone cores (collectively referred to as nucleosomes) that are spaced apart ~160bp in chromatin (48). This implies that in addition to activities directly related to DNA replication, the eukaryotic replisome must also progress through and help preserve chromatin. Some activities enable progression of the replication fork (the DNA structure at the replisome) as they encounter nucleosomes (e.g. FACT (49), while others facilitate the transfer of nucleosomes – which may carry posttranslational modifications – from the front of the replication fork to the nascent DNA. Multiple replisome subunits have histone chaperone activity, including MCM2, Pol α, CTF4, Pol ε, and RPA (21, 50-54), but the details of how they contribute to these processes is still emerging. Given the recent insight on the role of stochasticity from the bacterial replisome, we think it is unlikely that histones are being transferred down a fixed, deterministic path – one histone chaperone protein to the next – but rather, the multiple histone interaction sites on the replisome serve to retain them locally and increase the likelihood of them binding to nascent DNA. Chromatin also has important roles during the synthesis and maturation of Okazaki fragments. In budding yeast, Okazaki fragments have a narrow range of approximately 160bp, nearly matching the inter-nucleosomal distance (48). In vitro, chromatin is required to maintain this tight distribution of lengths (39, 49), and can still be observed even in conditions where Pol α is not coupled to the CMG – by removing CTF4 from the reaction (49).

Chromatin may also play a role in the dynamics of the leading-strand polymerase, and influence whether it frequently unbinds from DNA like its bacterial counterpart, Pol III. While Pol ε may likely encounter obstacles and DNA topology issues like PolIII, it also has an unusual flexible linker that could allow it to bypass obstacles, while remaining stably connected to the replisome (33) (Fig. 2). It is possible that the flexible linker was selected for to increase the frequency of deposition of PCNA clamps on the leading strand, as seen in vitro (55). PCNA plays a role in assembling new nucleosomes, though its interaction with chromatin assembly factor 1 (CAF-1), which is involved in depositing H3-H4 tetramers (21, 56). Given how frequently new nucleosomes need to be assembled, and the weak interaction of Pol ε to PCNA, it is possible that the flexible nature allows Pol ε to remain attached to CMG helicase when it unbinds from PCNA - giving access to the clamp loader to load new PCNA - and subsequently engage again with the leading-strand. This behaviour can also be exploited when bypassing obstacles on DNA. Another suggested use for the flexible linker is that it allows Pol ε to take over synthesis after Pol δ initiation in models where Pol δ contributes to the synthesis of the leading-strand (32, 33).

**Study of DNA replication using live-cell fluorescence microscopy**

Fluorescence microscopy has been an invaluable method to study the localization, dynamics, and stoichiometries of proteins, especially in living cells. In particular, single molecule approaches have allowed us to obtain distributions of variables rather than simply ensemble averages, giving us a more stochastic and dynamic view of biology, with a stronger emphasis on the probabilistic nature of biological processes. Indeed, the recent discoveries of the dynamic and stochastic nature of replisome in bacteria are all due to these approaches.

*In vitro* studies of *E. coli* DNA replication have provided important information on the factors affecting the processivity and rates of individual replisomes (11, 13, 57). A complementary approach is to study this process in live cells, where enzymes are in their natural milieu. With recent advances in microscopy and genetics, we and others have been able to successfully pursue this avenue. Specifically, single-particle tracking photoactivated localization microscopy (sptPALM), has the advantage of being able to directly observe the dynamics of single copies of replisomal proteins (16, 18, 58). It was through this technique, as highlighted above, that we were
able to directly visualize the dissociation of PolIII*, as indicated by the disappearance of fluorescent foci – much faster than the measured photobleaching rate – using long-capture (Fig. 2) (16).

Figure 2 - Single molecule approaches: (A) Study of DNA binding proteins using single-molecule approaches allow us to obtain information about their binding kinetics, diffusion coefficients and the proportion of DNA-bound copies. (B) Schematic representation of the data obtained in single-molecule fluorescence microscopy of live cells. Individual spots, representing single copies of the protein studied, are detected and localized at each frame (using a 2D-gaussian fitting of their intensities to determine their localization with high precision). Diffusive molecules can be accurately tracked through a time-lapse if the integration times at the camera are in the range of millisecond to low tens of milliseconds (left). Alternatively, longer camera integration times (ranging from hundreds of milliseconds to few seconds) results in average blurring of diffusive molecules, but still allows detection of immobile molecules (right). (C) Example of the analysis of fast-capture data. The distribution of diffusion coefficients and the fraction of bound and diffusive molecules can be obtained from them. (D) Example of the analysis of slow-capture data, where the distribution of fluorescent spot lifetimes can be used to determine the residence times of proteins on DNA.

SptPALM requires tagging the protein of interest (POI) with a photoactivatable (PA) fluorescent protein (FP) (59). These are proteins that are normally in a dark state, but upon excitation with low 405nm wavelength light, a small subpopulation of tagged proteins will stochastically convert to a fluorescent state. A variation is photoconvertible (PC) FPs which
convert to a different a fluorescent state. PA/PC-FPs have enabled tracking of single molecules even for high copy-number proteins (60). Tracking of single-molecules is done by linking their localizations frame to frame, and various algorithms are available for this non-trivial process (58, 61-63). After tracking, one can extract a variety of parameters for the POI such as its diffusion coefficient, proportions of different diffusive states, residence time, search time, and high precision mapping of cellular localization; all valuable for determining kinetics of replisome subunits in live cells (Fig. 3) (16, 18, 58).

Recent advances in genetics have also paved the way to perturb conditions commonly done in vitro, directly in the cell. One example is using degron tags to rapidly deplete the POI, in a controlled manner, thereby lowering concentrations of competitor molecules (64). Likewise, the use of temperature-sensitive mutants of POIs allows one to inactivate the protein rapidly. The advantages of these approaches compared to simply using deletion mutants, is that we can perform them on essential proteins and cells do not accumulate suppressor mutations (64).

Development of budding yeast as a system for single-molecule microscopy has been slower than in bacteria, having only few pioneering contributions (65-71). The main challenges for doing single-molecule work on replisomes in budding yeast are: 1) the size of the yeast cell (4um diameter for haploid cells, compared to 0.7 um for E. coli), resulting in higher background fluorescence, 2) cellular autofluorescence, 3) the higher number of replisomes active at a given time (45, 72, 73), and 4) the light diffraction from the thick cell wall (74). Even though more work is needed to make of budding yeast an amenable model for single-molecule microscopy, new advances in microscope configurations have allowed for selective illumination to minimize background fluorescence while retaining high sensitivity. These include light sheet microscopy (LSM) and highly-inclined and laminated optical sheet (HILO) (75, 76). In addition, the development of brighter and more photostable fluorescent dyes/proteins (e.g. Halo and SNAP tags) have improved detection of single-molecules, while also allowing for them to be observed for longer periods of time (60, 69, 77, 78).

Perspectives

i) Advances in fluorescence microscopy have allowed us to probe the dynamics and architecture of replisomes in living cells.

ii) The recent view of the bacterial replisome as being a dynamic cellular machine, with inherent stochastic processes, would not have been revealed without the advent of single-molecule fluorescence techniques.

iii) It will only be a matter of time before these techniques are applied to study the eukaryotic replisome, providing a complementary approach to traditional biochemical approaches, revealing a richer understanding on how it operates: is it inherently dynamic and stochastic like the bacterial replisome? How is chromatin assembly coupled to DNA synthesis? What are the consequences when it encounters obstacles on DNA?

The capability to study replication complexes in a range of organisms, especially in live cells, will help to illuminate the factors governing their architectures and dynamics, and how they perform one of life’s fundamental processes: DNA replication.
**Funding**

This work was supported by the Natural Sciences and Engineering Research Council of Canada (Discovery Grant, 435521-2013), and the Canadian Institutes of Health Research (Operating Grant, 142473).

**Competing Interests**

The Authors declare that there are no competing interests associated with the manuscript.

**Acknowledgements**

We like to thank Ziad El-Hajj, Maxime Leroux, Nicolas Soubry for their feedback on the manuscript.

**Author Contributions**

Nitin Kapadia and Rodrigo Reyes-Lamothe wrote the manuscript.
References