Protein-Protein Interactions in the Bacterial Replisome

Patrick Schaeffer, Madeleine Headlam and Nick Dixon

Research School of Chemistry, Australian National University, ACT 0020

Replication of DNA in all organisms proceeds in three stages: initiation at origins of replication, DNA synthesis (elongation) at replication forks, and termination. Each of these processes is mediated by multiple stable or transient protein-protein and protein-DNA interactions involving subsets of the 30 or so different replication proteins (**Fig. 1**). The elongation phase occurs within a large nucleoprotein assembly called the replisome (1). Replisomes have never been isolated intact from cells, so an understanding of their structure and function has required their painstaking *in vitro* reassembly from individual subunits. Mechanistic studies have made extensive use of the *Escherichia coli* proteins, in part because they are separately isolatable in large quantities from overproducing strains, and *in vitro* initiation, elongation and termination reactions have been faithfully reconstituted using defined DNA templates. Other systems studied in similar detail include the replication of the DNA of bacteriophages. Some, like T4 and T7, encode their own replication proteins, while others like the small single-stranded (SS) DNA phages make use of subsets of host enzymes. Similar reductionist approaches have more recently been applied to identify many of the replication proteins from eukaryotes. It is clear that protein functions and mechanisms are conserved in all organisms (**Table 1**), though the proteins may have little or no recognisable sequence similarity.

Work in the last ten years has produced a good understanding of how protein-protein interactions mediate replisomal DNA synthesis (**Fig. 2A**), and high-resolution structures of many of the individual proteins (or domains thereof) and some larger complexes are now available. For all of these reasons, it is apparent that *E. coli* DNA replication provides an excellent opportunity to study the chemistry that governs macromolecular interactions in this complex, flexible and dynamic nucleoprotein assembly. The *E. coli* replisome can thus be seen as a tractable model system to develop knowledge and tools that may be applied to other large dynamic multiprotein complexes. Such complexes mediate or control many aspects of development and function in organisms from all domains of life.

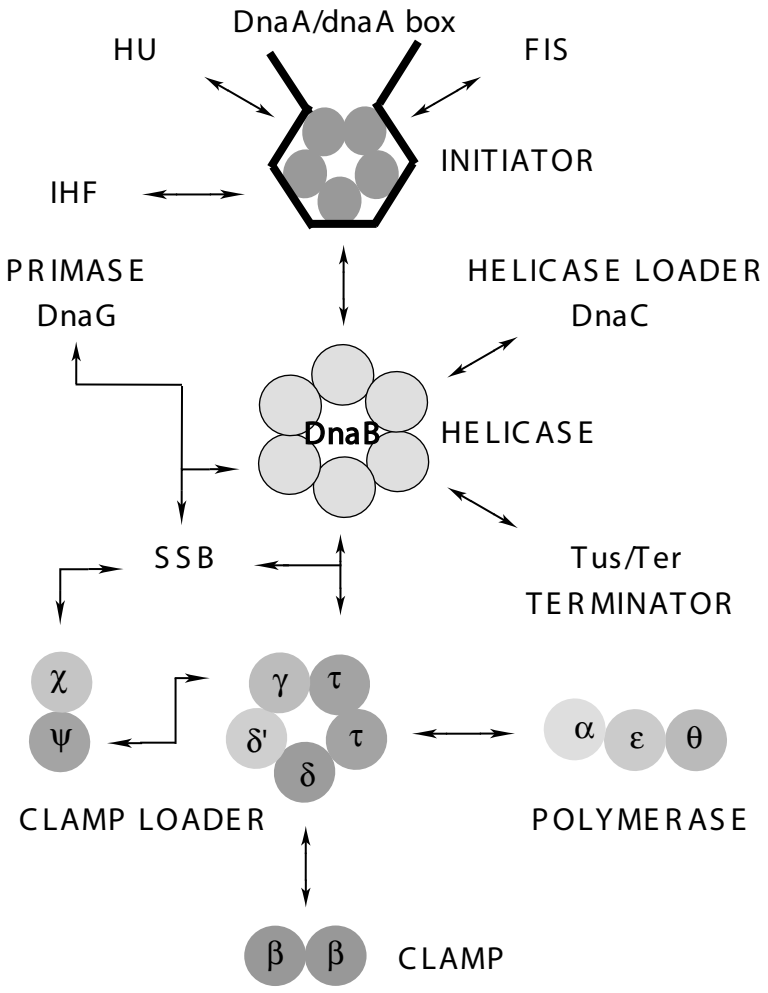

Fig. 1. Schematic representation of the *E. coli* replicative interactome.

Assembly of two replisomes at oriC

How replisomes are assembled is probably the least well-understood process in replication. In the 4.7 million base pairs of the circular *E. coli* chromosome is a 260-base pair sequence that constitutes the unique origin of replication (*oriC*). The origin is recognised by the DnaA replication initiator protein (2, 3). Multiple copies of DnaA bind to five 'DnaA boxes' in *oriC*, and in the presence of small basic histone-like proteins (HU and/or IHF) that stabilise bends in DNA, this leads to separation of the two DNA strands at a nearby AT-rich region to produce a single-stranded bubble.

The ring-shaped hexameric $DnaB_6$ helicase is a molecular motor that hydrolyses ATP to fuel its translocation in the 5′-3′ direction on the strand it encircles, resulting in displacement of the other strand at a replication fork (4, 5). When it is free in the cell, $DnaB_6$ is normally associated with six molecules of the helicase loader protein DnaC. The central channel in the $DnaB_6.(DnaC)_6$ complex is closed (6) so that the helicase cannot associate with DNA until it interacts specifically with the N-terminal domain of DnaA in the origin complex. This leads to opening of the

Table 1. Correspondence among replication proteins from *Escherichia coli*, *Bacillus subtilis*, and bacteriophages T7 and T4.

Function	*E. coli*	*B. subtilis*	Phage T7	Phage T4
Polymerase/exonuclease	αεθ	PolC, $DnaE_{BS}$	Gp5	Gp43
Sliding clamp	β_2	β_2	Trx (*E. coli*)	$Gp45_3$
Clamp loader	$\delta\gamma\tau_2\delta'\psi\chi$	τ_5	–	$Gp44_4Gp62$
Terminator	Tus	RTP_2	–	–
SSB	SSB_4	SSB_4	$Gp2.5_2$	Gp32
Initiator	DnaA, IHF, HU, FIS	DnaA, B, D	T7 RNAP	–
Replicative helicase	$DnaB_6$	$DnaC_6$	Gp4	Gp41
Helicase loader	DnaC	DnaI	–	Gp59
Primase	DnaG	DnaG	Gp4	Gp61

helicase ring, dissociation of DnaC, and directional loading of two $DnaB_6$ molecules within the SS-DNA regions. The helicase molecules extend the single-stranded bubble by translocating in opposite directions on what will become the lagging strands of the two replication forks that invade the duplex DNA in opposite directions on each side of the origin. Each molecule of the helicase transiently associates through its N-terminal region with the C-terminal domain of the DnaG primase. Primase is a specialist RNA polymerase that periodically synthesises short RNA primers on the SS-DNA regions to initiate DNA synthesis by DNA polymerase III (Pol III). DNA polymerases cannot start polynucleotide chains, so each nascent DNA fragment must be primed by DnaG. On the leading strand, the first primer laid down on each strand at *oriC* is extended continuously and processively by Pol III, which may never dissociate from the template. On the other hand, replication of the lagging strand needs to occur discontinuously, producing Okazaki fragments about 1 kb in length that are subsequently processed by DNA polymerase I and joined by DNA ligase. Thus, the lagging-strand Pol III must dissociate and reassociate with a new primer every second or two.

Elongation by DNA polymerase III holoenzyme

The Pol III holoenzyme (**Fig. 2B**) is remarkable not only because it is a very efficient enzyme (turnover number about 1000 s^{-1} per active site), but also because it is almost perfectly faithful (it only incorporates correctly paired nucleotides) and infinitely processive (never dissociates from a template-primer). These attributes require the collaboration of 10 different subunits; the probable composition of the holoenzyme is $(\alpha.\varepsilon.\theta)_2$-$(\delta.\gamma.\tau_2.\delta'.\psi.\chi)$–$(\beta_2)_2$, comprising three separate sub-assemblies (7). The core complex (α.ε.θ) contains the large polymerase subunit (α), the proofreading 3′-5′ exonuclese (ε) that ensures high fidelity by excising nucleotides misincorporated by the polymerase (8), and the small θ subunit that binds to and stabilises ε (9). The ε subunit has two domains, a larger N-terminal portion that has enzymatic activity, and a smaller C-terminal region that associates tightly with α.

The β_2 molecule is a ring-shaped dimer that encircles double-stranded DNA without actually contacting it to form a tight topological linkage (10-12). It acts as a sliding clamp on the newly-replicated leading and lagging strands behind the two Pol III cores, with which it interacts through a C-terminal region of the α subunit. This ensures that the polymerase never dissociates at an inappropriate time from the template-primer. The third sub-assembly $(\delta.\gamma.\tau_2.\delta')$-$(\psi.\chi)$ is the clamp loader or γ complex, which has multiple functions (7, 11). The τ and γ subunits are closely related molecules. They are both translated from the same (*dnaX*) mRNA. About half of the time, the ribosome undergoes a programmed frameshift and terminates translation two-thirds of the way through the message to produce γ. Thus, the N-terminal regions of γ and τ are identical, and it is these regions that associate together and with δ.δ′ in the central five-subunit core of the clamp loading complex (7, 13). The distinct C-terminal regions of the two molecules of τ have two functions. One is to dimerise the core, through interaction with the C-terminal regions of the two α subunits. The other is to interact with the $DnaB_6$ molecule on the lagging strand, ensuring continued communication between the holoenzyme and the helicase (7).

The third function of the γ complex is to act as the clamp loader (11). The δ subunit interacts with the β_2 sliding clamp, probably at the site occupied by α (see the following articles in this Showcase on Research on β_2 by Wijffels *et al.* and its eukaryotic equivalent PCNA, by Kunz *et al.*). On the lagging strand, in a series of conformational changes in the γ complex triggered by the arrival of a new primer terminus and driven by hydrolysis of ATP, δ acts as a wrench to open the β_2 ring and load it onto the template-primer (14, 15). It is not clear yet if the same mechanism is used at *oriC* to load the β_2 molecules that continuously associate with the leading strand polymerase. Indications that β_2 interacts with DnaA (and perhaps other components of the initiatior complex) suggest that they might be loaded in a different way.

SS-DNA regions at the replication fork produced by helicase action are covered by a remarkably versatile protein called SS-DNA binding protein (SSB). SSB_4 has a tetrameric N-terminal DNA-binding domain, and each subunit has a more flexible C-terminal domain that appears to interact with several of the other replication proteins, to stimulate and coordinate their activities at replication forks. The fourth function of the clamp loader complex involves recognition of the SSB_4 in SSB-coated SS-DNA by χ. The small χ subunit of the clamp loader forms a heterodimeric complex with ψ, which in turn interacts with one of the γ and/or τ subunits. By sensing the presence of SS-DNA, χ could facilitate the recognition of a new primer terminus by γ or τ (16).

Some interactions among replisomal proteins seem to provide sensors and switches that regulate and coordinate the actions of the different components (e.g. the interactions of β_2 with α and δ, of SSB_4 with χ and $DnaB_6$, of τ and γ with SS-DNA, and of DnaG with $DnaB_6$ and primer termini). Although great progress has been made in working out how detection of primer termini by τ and γ is communicated to signal it to open and load β_2 (13, 15), unravelling how these switches work remains a challenge.

Termination of replication

Termination of DNA replication in *E. coli* is mediated by the extraordinarily tight association of Tus, a monomeric site-specific DNA-binding protein, to 21-base pair *Ter* sites in the region of the chromosome opposite *oriC* (17). The Tus-*Ter* block is polar, in that forks arriving from one direction pass through unimpeded, while those arriving from the other side are blocked. How polarity is achieved and how the distinctly different termination system of *Bacillus subtilis* operates has provoked recent controversy, and is discussed by Duggin in this Showcase on Research.

Replication restart at stalled forks

If replisome assembly requires DnaB loading at a DnaA-*oriC* complex, then replication forks would be extremely vulnerable to any insult that led to dissociation of the helicase. An alternative way to load DnaB onto SS-DNA was discovered more that 20 years ago during study of duplication of the SS-DNA of coliphage ϕX174, a process that requires only host proteins (1). Four additional proteins, PriA, PriB, PriC and DnaT, were needed to load DnaB from the $DnaB_6$.$(DnaC)_6$ complex at the phage complementary strand origin. Once loaded, PriA and PriB (at least) remained associated with DnaB in what was called the ϕX primosome, and PriA was shown to be a 3′-5′ DNA helicase. It seems very likely that in chromosomal replication PriA (and PriB) continuously associates with DnaB at the replication fork, and that each of the helicases can participate in reloading of the other when required to prevent stalling of forks (18). Recent work has clearly demonstrated roles for PriA and the ϕX primosome (now renamed the replication restart primosome) in several mechanisms for rescue of stalled forks (19).

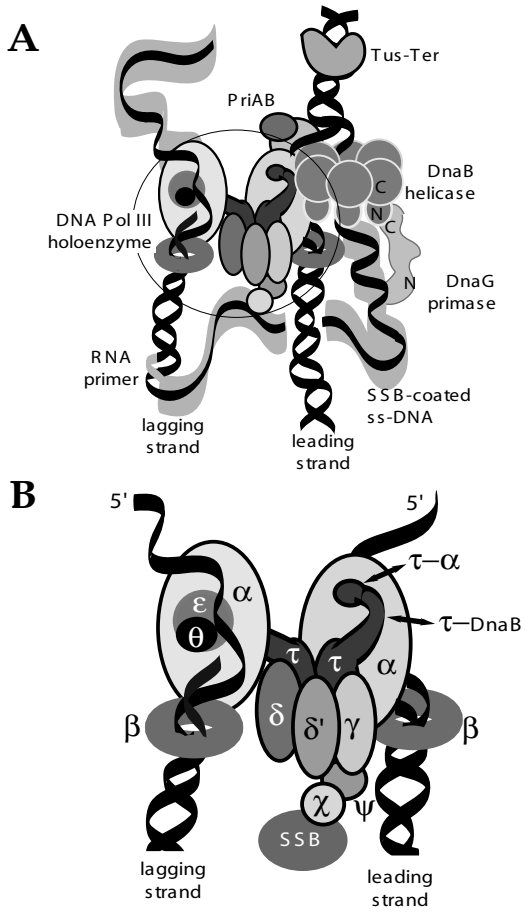

Fig. 2. Protein-protein and protein-nucleic acid interactions in the *E. coli* replisome (A), and in the asymmetric dimer of DNA polymerase III holoenzyme (B).

The replication factory

Reconstituted replication reactions in test tubes fail in several respects to duplicate *in vivo* conditions. Chromosomal DNA is tightly packaged in the crowded environment of the cell. It is also covered by nucleoid proteins, repressors and activators of genes, and is being actively transcribed by RNA polymerases. In addition, it accumulates lesions as a result of environmental abuse, and these are continuously repaired by other large protein complexes. Active underwinding of DNA ahead of replication forks by topoisomerases like gyrase is a prerequisite to translocation of replisomes. The view that replisomes (and perhaps also repair and transcription complexes) move along the chromosome in bacteria is probably incorrect. It has been shown that replisomal proteins of *Bacillus subtilis* are localised to central regions of cells for a significant part of the cell cycle (20) and there is good evidence to suggest that the DNA template is pumped through this replication factory (presumably by back-to-back DnaB hexamers), so that the two new chromosomes are segregated each to one end of the cell (21). It would be surprising if yet undiscovered scaffolding proteins were not involved in organisation of the replisomal complexes.

Thus, although work over the past 25 years has given us satisfying insights into the assembly and structure of replisomes, and good understanding of the roles of the individual proteins, there is still work to be done (even in the prokaryotic systems) to understand signals and switches, and how the replisome works in its cellular environment.

References

1. Kornberg, A., and Baker, T.A. (1992) *DNA Replication*, W.H. Freeman & Co., New York
2. Messer, W. (2002) *FEMS Microbiol. Rev.* **26**, 355-374
3. Erzberger, J.P., Pirruccello, M.M., and Berger, J.M. (2002) *EMBO J.* **21**, 4763-4773
4 Delagoutte, E., and von Hippel, P.H. (2002) *Quart. Rev. Biophys.* **35**, 431-478
5. Delagoutte, E., and von Hippel, P.H. (2003) *Quart. Rev. Biophys.* **36**, 1-69
6. Bárcena, M., Ruiz, T., Donate, L.E., Brown, S.E., Dixon, N.E., Radermacher, M., and Carazo, J.M. (2001) *EMBO J.* **20**, 1462-1468
7. McHenry, C.S. (2003) *Mol. Microbiol.* **49**, 1157-1165
8. Hamdan, S., Carr, P.D., Brown, S.E., Ollis, D.L., and Dixon, N.E. (2002) *Structure* **10**, 535-546
9. Keniry, M.A., Berthon, H. A., Yang, J.Y., Miles, C.S., and Dixon, N.E. (2000) *Protein Sci.* **9**, 721-733
10. Kong, X.-P., Onrust, R., O'Donnell, M., and Kuriyan, J. (1992) *Cell* **69**, 425-437
11. Jeruzalmi, D., O'Donnell, M., and Kuriyan, J. (2002) *Curr. Opin. Struct. Biol.* **12**, 217-224
12. Oakley, A.J., Prosselkov, P., Wijffels, G., Beck, J.L., Wilce, M.C.J., and Dixon, N.E. (2003) *Acta Cryst.* **D59**, 1192-1199
13. Jeruzalmi, D., O'Donnell, M., and Kuriyan, J. (2001) *Cell* **106**, 429-441
14. Jeruzalmi, D., Yurieva, O., Zhao, Y., Young, M., Stewart, J., Hingorani, M., O'Donnell, M., and Kuriyan, J. (2001) *Cell* **106**, 417-428
15. Ason, B., Handayani, R., Williams, C.R., Bertram, J.G., Hingorani, M.M., O'Donnell, M., Goodman, M.F., and Bloom, L.B. (2003) *J. Biol. Chem.* **278**, 10033-10040
16. Kelman, Z., Yuzhakov, A., Andjelkovic, J., and O'Donnell, M. (1998) *EMBO J.* **17**, 2436-2449
17. Neylon, C., Brown, S.E., Kralicek, A.V., Miles, C.S., Love, C.A., and Dixon, N. E. (2000) *Biochemistry* **39**, 11989-11999
18. Allen, G.C., Jr., Dixon, N.E., and Kornberg, A. (1993) *Cell* **74**, 713-722
19. Cox, M.M., Goodman, M.F., Kreuzer, K.N., Sherratt, D.J., Sandler, S.J., and Marians, K. J. (2000) *Nature* **404**, 37-41
20. Lemon, K.P., and Grossman, A.D. (1998) *Science* **282,** 1516-1519
21. Lemon, K.P., and Grossman, A.D. (2000) *Mol. Cell* **6**, 1321-1330