THE REPLICATION FORK AND THE REPLISOME

REVIEWS


For those who want to know the latest, there are more recent reviews from O’Donnell. However, this gives the best big picture overview of how the components at the fork work, cooperate, and deal with DNA lesions. For a “snapshot” view of the replisome see Yao, N.Y. & O’Donnell, M. (2010) Cell. 141:1088. For the latest reevaluation on the dynamics of replisome structure and function (including the idea that there are 3 DNA polymerases at prokaryotic replication forks) see Kurth, I., and O’Donnell, M. (2013). Trends Biochem Sci. 38:195-203.


A little dated, but discusses how hexameric replicative helicases work to unwind DNA.


Latest on the beautiful dissection of how the clamp-loader works to carefully coordinate all its activities.


Extensive review of an important class of oligomeric ATPases that act as switches and do work because their conformation is very sensitive to the binding and/or hydrolysis of ATP. They are used in a diverse set of molecular reactions, not just replication.


Good review of prokaryotic replication fork, and how the next level of questions can be addressed with single molecule studies.

Do not have time to discuss this in the lectures, but this review discusses the third mechanism responsible for replication fidelity, and also a mechanism that polices recombination, making sure it only occurs between DNA with enough sequence identity to likely be from the same species.

**PRIMARY LITERATURE**

In vitro Systems


One of the first uses of in vitro complementation to purify proteins involved in a complex reaction.


Early reconstruction of a replication fork from purified components. Keep in mind that gp 61 was only an impurity in the 99% "pure" preparation of gp32 when this was published. Later another impurity of one of the "pure"components, gp59 the helicase loader, was identified.

Helicases.


GTPase activity of gp41 stimulated by SS DNA. Possibility raised that gp41 is translocating along SS DNA


Demonstration that gp41 behaves as expected of a 5' to 3' helicase.


Structure of the papillomavirus E1 hexameric replicative helicase shows single-stranded DNA binding to protein loops in cavity of hexamer toroid. Sequential hydrolysis then recharging is hypothesized to move these loops in a power stroke that allow translocation down single-stranded DNA.

This single molecule analysis provides the most direct evidence that the eukaryotic replicative helicase is translocating along the leading strand (in a 3’> 5’ direction) in the replisome.

Primases


Description of 5 protein system and evidence that priming requires 41, 61 and rNTP.


GTPase activity of gp41 is required for prolonged primer synthesis. Suggestion that gp41-gp61 behaves as a mobile primosome. Work by others also indicated that the association of primers and helicases often enhance each others activities.


Primase and helicase are part of a single polypeptide in T7 that encircles the lagging strand as a hexamer (in other prokaryotic systems, the two are independent proteins, but physically interact). The primase is much slower and moves in the opposite direction as the helicase. Single molecule analysis of fork movement indicates that during primer synthesis a transient template priming loop is formed between the primer and helicase. This contradicts a previous single molecule study that concluded that the helicase pauses during primer synthesis (Nature 2006. 439: 621-624).

Helix-Destabilizing Proteins.


A classic. The discovery and description of a helix-destabilizing protein. And first use of a DNA affinity column.


Gene 32 protein shown to be very important for getting polymerase through hairpins.

Informative structure suggesting how SSBP preferentially binds to SS DNA over DS DNA, and how it can protect the sugar phosphate backbone and yet facilitate base pairing of the bases.

**Polymerases and Polymerase Accessory Proteins.**


Demonstration that the processivity of the T4 polymerase is enhanced by accessory proteins. ATP hydrolysis required, but not continuously, leading to notion of sliding clamp”.


Demonstration that accessory Protein Beta is topologically linked to DNA. Accessory Protein Gamma and ATP are required to load the clamp onto DNA.


The sliding clamp looks like a doughnut!!! Includes the idea--based on limited sequence similarities--that human PCNA protein and T4 gene 45 protein have a similar structure.

**Processing Okazaki Fragments**


RNaseH and Ligase can fuse okazaki fragments in the T4 in vitro replication system.


Minireview that summarizes evidence that although Pol I is largely responsible for removing RNA primers from DNA, RNase H plays a contributory role.

**Recycling of Lagging Strand Polymerase**

Dilution experiment showing lagging strand polymerase recycles in T4 in vitro replication system. There are a lot assumptions needed in the interpretation of this experiment since (in my view) it wasn't done precisely right.


Evidence that, at the end of an okazaki fragment, the lagging strand polymerase escapes the processivity conferred by the beta sliding clamp by hopping to a new clamp. This observation formed the basis for the collision model of recycling, wherein the lagging strand polymerase knows to release from the clamp for recycling to the next primer by sensing when it has run into the previous okazaki fragment.


The tau subunit of Pol III holoenzyme turns out to be the sensor for DNA structure that triggers the collision mechanism for lagging strand polymerase recycling. When the polymerase runs into the preceding okazaki fragment, the tau protein ejects the polymerase from the clamp. Later in subsequent papers, it is shown that the tau subunit competes with the clamp for the C-terminus of DNA polymerase III, with the DNA structure determining who wins that competition.


An alternative model for polymerase recycling that seems to predominate in the phage T4 system, but may also be important in the E. coli system when lagging strand polymerases run into template lesions, is the “premature release” or “signaling” model. In this model, the lagging strand polymerase can release from its clamp even before it has completed extension of its okazaki fragment when it senses that a new primer (or new clamped primer) has been synthesized. This would leave a gap in the DNA that would have to be filled in later. Here replication by the phage T7 replisome is monitored on a single molecule level, allowing dynamic observation of trombone expansion and collapse. Using the collapse to monitor polymerase release, the authors make a lot of inferences to argue that release is sometimes through the collision model and sometimes through the signaling model. Paper illustrates how single molecule strategies are increasingly needed to address the precise workings of the replication fork.

Clamp Loading and Protein Switching


Competition between the gamma clamp loader and the core polymerase for the same face of the beta clamp is observed. Gamma clamp loader wins in the absence of DNA; core polymerase wins (and
protects the clamp from gamma) on primed DNA template. Help explains the recycling of beta clamp during OF synthesis


Sigma subunit of the gamma complex (γ2σ'χψ) can open the clamp by simply binding to it. This activity is masked when sigma is bound to sigma' in the gamma complex. ATP binding by the gamma subunits induce a conformational change that exposes the sigma and allow it to bind the clamp. Clamp binding inhibits gamma ATPase activity and allows association with a primer-template. Primer-template binding stimulates ATPase activity leading to closure of the clamp around the DNA and dissociation of the gamma complex.


Primase remains on the primer because of association with SSB C-term. Chi subunit of the gamma complex can outcompete primase for the SSB C-term, leading to displacement of primase, freeing the primer-template for clamp loading.


Precise understanding of the conformational changes that allow the clamp loader to pry open the clamp and load it around the primer template junction require atomic level structural analysis.

**SOME QUESTIONS TO THINK ABOUT THE LECTURE**

1) How does a pulse-chase experiment differ from a short pulse versus a long pulse experiment (such as the one performed by Okazaki)?

2) When performing a pulse-chase how do you determine how long to pulse? How is a chase performed, and how can you evaluate how rapidly and completely the chase occurs?

3) How do you know that the reaction product that accumulates when you block a reaction step is a true intermediate and not a dead end byproduct that is derived from an intermediate?

4) If you remove a protein from an in vitro system and steps A, and B of a process occur, but not C, D, E, and F, you could conclude that the protein is necessary for step C. But this approach would only reveal the first step that the protein is required for. It wouldn’t detect whether the protein is also required for a later step, like E. Experiments that try to determine when a protein (or cofactor) has completed all its functions and is no longer needed to complete a reaction are called “execution point” experiments. The determination of when ATP hydrolysis is no longer needed for T4 DNA polymerase
processivity is an example of an execution point experiment. What does one need in order to do an execution experiment for a protein in vitro (or its gene in vivo)?

5) Suppose a four step in vitro reaction requires four different proteins, one for each step. If the proteins acted independently, withholding each protein from the system would lead to a block in a different step. However, suppose all the proteins need to first associate into a macromolecular complex before any can become active for their particular step. What would happen if you withhold any single protein now? What type of perturbation would you now need to assign specific proteins to specific steps?

6) Why is it easier to ask whether a protein or nucleic acid is sufficient for an activity in vitro than in vivo?

7) The connection between SS DNA stimulation of 41 ATPase activity and translocation along SS DNA was made because ATPase stimulation was greatest for circular SS DNA, less for long SS DNA, and poor for short SS DNA. DNA was in excess during these experiments, however, so there was always DNA around for 41 to hop onto. So how would you rationalize the difference in stimulation?

8) Assuming that helicases have to load onto single-stranded DNA before they can translocate to and then separate duplex DNA, how might you design an assay to distinguish between a helicase that translocated along in a 3' to 5' direction from one that translocated in a 5' to 3' direction.

9) In the experiments establishing the topological linkage of beta-dimer to DS DNA, how do you imagine this association was monitored? If this protein-DNA interaction was very stable (i.e. dissociation takes hours) what type of methods were feasible? Why was the second digest performed after the linearized ends were blocked with DNA binding proteins; what possibility was ruled out?

10) How does the ability of SSBP to destabilize DS DNA differ from the activity of helicases?

11) The trombone model for the replication fork was proposed to accommodate the observation that the lagging strand polymerase processively synthesizes many rounds of okazaki fragments without dissociating from the fork. Can you suggest an experiment and results that would demonstrate such processivity? Note the processivity for okazaki fragment synthesis is distinct from the processivity for each nucleotide incorporation.

12) Why doesn't the clamp loader/clamp displaces the primase before the primase has completed its task?

POINTS RAISED IN LECTURE

1) Dissecting mechanisms of biological processes involves identifying and structurally characterizing intermediates in the process. Once identified, the transitions from one intermediate to another define activities required for the process. In addition, one can also define activities that affect the rate and/or fidelity of these steps. Three approaches commonly used to enhance the ability to detect and characterize intermediates are: (1) synchronizing reactions, (2) using pulse-chase labeling, and (3) blocking reaction steps. There are numerous ways to structurally characterize the nucleic acids or
proteins components of an intermediate. Ultimately we would like to identify the protein(s) or nucleic acid(s) that is responsible for each activity, and determine how these components coordinate their activities to act in the right time, place, and order. Both genetics and biochemistry

2) **Structural analysis of in vivo replication intermediates** established that: (1) replication was localized to moving replication forks; (2) synthesis at forks occurs in a semidiscontinuous manner; (3) discontinuous synthesis involves synthesis of short 100-2000 bp okazaki fragments that are ligated to form the daughter strands on one side of the fork. These observations suggest the presence of a number of activities needed at a replication fork in addition to the replicative DNA polymerase.

3) **Awesome power and limitation of genetics.** This was not discussed in lecture, but *E. coli* mutants unable to synthesize DNA were isolated by several schemes; because DNA synthesis is an essential function these schemes relied on the ability to isolate conditional-lethal mutations. Despite little knowledge of the replication process, genetics was able to identify many genes involved in DNA replication. On the other hand, because of little knowledge of replication process and limited ability to characterize replication intermediates in vivo, phenotypic analysis of function was extremely crude. Understanding the precise functions of the genes and their protein products required better biochemical access to the problem through development of in vitro replication systems.

4) **Advantages of dissecting mechanisms in vitro instead of in vivo.** Can study process in isolation. Easier to separate and purify biochemical activities. Easier to perform kinetic analysis with synchronized reaction. Easier to perturb system to establish partial reactions by either blocking steps or introducing defined intermediates. Better ability to probe, isolate, label, and structurally characterize intermediates. Despite huge advances in our ability to study events within the cell, these advantages still hold for in vitro studies.

5) **Complementarity of in vivo and in vitro studies.** It is first important to establish that an in vitro system for a cellular process is indeed a faithful representation of that process. Thus you need to have first characterized the process in vivo so one can have benchmarks to assess the fidelity of the in vitro system. Correlation may not be perfect, but the greater the similarities the greater ones confidence that one is not studying an in vitro artifact. Ultimately one still needs to validate the in vitro findings with in vivo studies.

6) **Biochemical purification of activities for complex processes.** A complex process recapitulated in a soluble system can be broken down into its individual activities even without knowing any details about those activities. As long as you can reconstitute the activity of the full process, you can assay individual activities based on their ability to contribute to this reconstitution. There are two fundamental approaches you can use. One is fractionation reconstitution, where you blindly separate a activities and can pursue the purification of each one based on its reconstitution activity. The other is in vitro complementation, which requires prior genetic isolation of mutants in the process. There you are trying to purify the activity disrupted in those mutants, by complementing defective extracts from those mutants with fractions from wild-type extracts. With the ascendancy of molecular biology, genes that participate in a process can often be readily identified, so these approaches would be accelerated by affinity purifying the gene products and seeing whether they could substitute for any fraction. Keep in mind that we are purifying functions (i.e. activities) not just proteins. However, the ability to rapidly purify proteins may facilitate the purification of these functions, while at the same time allowing a
general function to be assigned to the proteins. Once you can reconstitute the process with purified 
proteins you are ready to dissect the mechanism and define more specific functions for these proteins.

7) **Four fundamental activities work together at the replication fork**: polymerases (with accessory 
proteins), primases, helicases, helix destabilizing proteins. Primase and helicase often work together as a 
unit, the primosome. Know the assays that define protein functions (i.e., what the substrates are, how 
the protein affects the substrate to generate products, how the products are characterized, distinguished, 
and/or separated from the substrate). Biochemical function is determined by biochemical assays, not by 
sequence homology! These four activities were recognized in the following way: replication was 
reconstituted *in vitro* (s.s. phage for Kornberg lab, T4 for Alberts lab); the individual components 
required for replication were purified by a combination of pure biochemistry and *in vitro* 
complementation strategies; once purified, the individual components were studied alone to determine 
what they did or were left out of reconstitution reactions to see what steps failed (the biochemical 
equivalent of genetic deletion). Obviously, some intelligent guessing and imaginative assay design 
narrowed the possibilities tested. Additional activities are responsible for separating the two interlinked 
parental strands, okazaki fragment maturation or increasing the fidelity and/or efficiency of the process.

8) **Protein interactions can modulate protein function.** Many replication proteins interact with each 
other with a wide range of affinities. The functional relevance of these interactions is demonstrated by 
the ability of these interacting proteins to modify each other’s activities or generate new ones. The 
dynamic nature of these interactions allows these proteins to sense and respond to different situations, 
which in turn allows them to carry out a highly ordered and properly timed series of steps. Early 
evidence for this type of interactive and coordinated protein behavior led Bruce Alberts to propose that 
the proteins at the fork act as a "protein machine", i.e. a replisome. Mike O’Donnell has uncovered 
many aspects of how the E. coli replisome machine works, but there remain many unanswered questions 
that offer fertile grounds for Bioreg proposals.

9) **The eukaryotic replication fork** conserves many features of prokaryotic forks but there are some 
differences. (1) Eukaryotes requires three DNA polymerases. Polymerase alpha-primase is required to 
prime all DNA synthesis and extends the primer for a short distance. Polymerase delta and polymerase 
epsilon, each in conjunction with RF-C and PCNA (which are functionally equivalent to the polymerase 
accessory proteins in prokaryotes), provide processive DNA synthesis on the lagging and leading 
strands, respectively. (2) The putative eukaryotic replicative helicase is a complex of the Mcm2-7 
heterohexamer (analogous to the homohexameric prokaryotic helicases) with accessory factors, Cdc45 
and the GINS (Sld5, Psf1, Psf2, Psf3). It translocates along single stranded DNA in a 3’ to 5’ direction 
(instead of the 5’ to 3’ direction of prokaryotic replicative helicases). (3) So far a complete functioning 
eukaryotic replisome has not been reconstituted on a fork template, so very little is known about how 
this protein machine works, an even more inviting target for Bioreg proposals.

10) **The lagging strand polymerase is stably associated with the fork.** The lagging strand polymerase 
is highly processive with respect to the synthesis of multiple okazaki fragments. To explain this 
processivity Bruce Alberts hypothesized that the leading and lagging replicative polymerases were 
attached and moving in the same direction. To accommodate this arrangement he proposed the 
trombone model for the replication fork, where a loop of the parental lagging strand accumulates behind 
the actively synthesized okazaki fragment as the helicase unravels the parental strands. This single-
stranded loop is released as the lagging strand polymerase dissociates from the completed okazaki fragment and serves as the parental template for the next okazaki fragment.

11) **The asymmetry of the replication fork** places different requirements on the enzymatic machinery of the leading and lagging strands. The polymerase on the leading strand must be highly processive whereas that on the lagging strand must dissociate every time it encounters the end of an okazaki fragment (once/sec; this is the time it takes to make one Okazaki fragment of length 1000bp), then recycle to synthesize the new okazaki fragment. For the Pol III holoenzyme, the intrinsic activity is symmetric (both polymerases can operate with regulated processivity appropriate for lagging strand synthesis) and the environments imposes the asymmetry. Assoc with helicase insures that the leading strand polymerase is constitutively processive. Collision of the lagging strand polymerase with the 5' end of the previously synthesized okazaki fragment triggers a conformational change that allows the tau subunit of the holoenzyme to displace the core polymerase from the beta-clamp. The clamp remains to coordinate other replication-related functions such as okazaki fragment maturation, mismatch repair, and in eukaryotes nucleosome assembly. There is also evidence in some systems and under some circumstances that new primer synthesis can somehow “signal” premature release of the lagging strand polymerase before it has fully extended the okazaki fragment. Exactly how the leading and lagging strands coordinate their activity and how the primase and clamp loader on the lagging strand coordinates their activity is still poorly understood and being attacked by single molecule analyses.

12) **Rings can provide a topological solution to processivity.** Sliding clamps form rings around DS DNA and helicases form rings around SS DNA. This processivity contributes significantly to replication speed, since re-engaging a dissociated polymerase or helicase is a slow step. Hexameric helicases are thought to use ATP hydrolysis to translocate unidirectionally down the single strand and to displace a complementary strand that does not enter the central cavity (although there are competing models and possibly more than one way for these helicases to work). Such translocation properties could also allow a “helicase” to displace proteins or alter nucleic acid secondary structures. The topological solution to processivity also provides a topological problem, however; how to get the ring around the DNA in the first place. Clamp loaders have to break the ring and reseal it around a primer-template junction. Helicase loading is one of the key function of the initiation reaction and it can be performed by a ring breakage or by building the ring from monomers.

13) **AAA+ ATPases are used for several replication functions.** These oligomeric proteins have ATP binding pockets formed at the interface between subunits. Thus ATP binding and/or hydrolysis can be coupled to large conformational changes either to do work or to act as switches. The replicative helicases and clamp loader are all AAA+ ATPases as well as initiation proteins. Exactly how helicases couple ATP hydrolysis to translocation is still unclear but recent structures have allowed several models to be hypothesized. ATPase binding of the clamp loader opens the structure and allows a domain to pry open the clamp. ATP hydrolysis by the clamp allows the clamp to close. Key is making sure that the hydrolysis and closing is coupled to binding to a primer-template junction.

14). **The third mechanism for ensuring replication fidelity is mismatch repair.** (not discussed in class) Three things are required for this repair: (1) recognition of the mismatch, (2) identification of the daughter strand, and (3) replacement of the mismatched nucleotide on the daughter strand. Mismatch recognition is achieved by MutS. Crystal structures show that MutS bends the DNA at the mismatch, suggesting that it detects mismatches by their increased “bendability”. Strand identity is established by
the GATC methylation state (the daughter strand is transiently unmethylated after replication) and is recognized by MutH. Nucleotide replacement is initiated on the daughter strand by MutH, which specifically nicks the unmethylated strand at the GATC site closest to the mismatch. Coupling of mismatch recognition and strand bias recognition is mediated by mutL which bridges mutH and mutS. Formation of this complex stimulates mutH to nick the unmethylated daughter strand at the GATC; this nick serves as the ultimate signal for strand bias since it directs the loading of exonucleases and helicase II at the nick, leading to degradation of the daughter strand between the GATC and the mismatch (including the mismatch). The resulting gap is filled in by the replicative polymerase, polIII holoenzyme, with the help of SSB and ligase.