REVIEW

Structural Mechanisms of Hexameric Helicase Loading, Assembly, and Unwinding [version 1; referees: 3 approved]

Michael A. Trakselis
Department of Chemistry and Biochemistry, Baylor University, Waco, Texas, 76798, USA

Abstract
Hexameric helicases control both the initiation and the elongation phase of DNA replication. The toroidal structure of these enzymes provides an inherent challenge in the opening and loading onto DNA at origins, as well as the conformational changes required to exclude one strand from the central channel and activate DNA unwinding. Recently, high-resolution structures have not only revealed the architecture of various hexameric helicases but also detailed the interactions of DNA within the central channel, as well as conformational changes that occur during loading. This structural information coupled with advanced biochemical reconstitutions and biophysical methods have transformed our understanding of the dynamics of both the helicase structure and the DNA interactions required for efficient unwinding at the replisome.
Introduction and context
Cell proliferation relies on the exact replication of an organism’s genetic material in a rapid but precisely controlled and efficient manner. The process and mechanism of DNA replication directs targeted and repetitive enzymatic activities towards long linear polymers of DNA. Interestingly, organisms have evolved a number of toroidal DNA replication and repair protein complexes that can maintain repetitive catalysis by encircling the DNA substrate. These protein-DNA rotaxane-like systems have the intrinsic ability to be processive enzymes due to their topological linkage with the substrate. As such, they provide inherent challenges to the loading and encircling of DNA. The steps required for the loading and encircling of circular protein complexes onto DNA provides for a higher level of regulation, which is required to restrict cell cycle progression and control DNA replication initiation. Because of this, the most highly regulated component within the DNA replisome is the loading and activation of the hexameric helicase, which dictates both the initiation steps and the elongation rate of DNA replication. Even though the general toroidal hexameric helicase structure has been known for more than two decades, the mechanisms for loading, encircling, activating, and unwinding are only just being discovered. These recent advances have been primarily aided by higher resolution structures that include DNA, better mechanistic descriptions of the interactions of the helicase with each separated strand of single-stranded DNA (ssDNA), and higher order in vitro reconstitution of DNA replication systems. It is an exciting time to be a part of the hexameric helicase field as big questions regarding dynamic structure-function relationships with DNA are poised to be revealed.

Hexameric helicase architectural conservation
Although the general architecture of hexameric DNA replication helicases is shared across organismal domains, there is strong evidence that classes of these enzymes have evolved independently for a role in DNA replication. All hexameric helicases are members of the broader P-loop family of ATPases, individual evolution of RecA domains gave rise to the superfamily (SF) 4 helicases including T4 gp41, T7 gp4, bacterial DnaB, and mitochondrial Twinkle, while SF3, including SV40 Large T antigen (SV40 L-Tag) and papilloma virus E1, and SF6, including archaeal and eukaryotic minichromosome maintenance proteins (MCM), helicases came from an ATPases associated with a variety of cellular activities (AAA+) clade (Figure 1 & Table 1). Regardless of the origin, these systems have all converged on a common ring-shaped architecture wherein a central channel is used to repetitively engage and translocate along ssDNA during unwinding. Bacterial and phage SF4 helicases are perhaps the best studied and have contributed most to our understanding of DNA unwinding, but more recent emphases on SF3 and SF6 helicases are providing insight into structure-function relationships across SFs.

Helicase activity requires the presence of both nucleotide triphosphates (NTPs) and Mg$^{2+}$ for unwinding. Although it is not known exactly how ATP hydrolysis directly drives DNA unwinding, it is likely to progress in a sequential manner, with each subunit driving conformational changes throughout the hexamer that contribute to unwinding polarity. RecA-like helicases (SF4) translocate along ssDNA in the 5’→3’ direction, while AAA+ enzymes (SF3 and SF6) translocate in the 3’→5’ direction. The core ATP

Figure 1. Structural conservation of hexameric helicases. Hexameric helicases are shown from different domains (RecA or AAA+) and superfamilies (SFs) with associated unwinding polarity and references. View from the N-terminal domain (NTD) rotated 90° to visualize the lateral length from the C-terminal domain (CTD) to the NTD.
Several high-resolution X-ray and electron microscopy (EM) structures have been reported for the apo and nucleotide-bound forms of hexameric helicases (T7 gp439,80, DnaB31,32, Mito Twinkle33, SV40 Large-T1,34, Sulfolobus solfataricus MCM [SsoMCM]33,35, and Saccharomyces cerevisiae MCM [ScMCM2-7]36). The global shared architecture of the ring-shaped helicases is generally composed of two tiers: an N-terminal DNA-binding domain (NTD) and a C-terminal AAA\(^+\) or RecA motor domain (CTD) (Figure 1). The orientation of most helicases on DNA places the CTD toward the duplex double-stranded DNA (dsDNA) region and the NTD outwards. The exception is E1, where the orientation is reversed, placing the NTD toward the duplex37. Thus, motor domains are often positioned close to the dsDNA duplex, which can leave the NTD regions free to bind, stabilize, or act on the resultant ssDNA.

Across species, hexameric helicase NTDs seem to have evolved differential functions. T7 gp4 can be expressed as either a 56 kDa helicase only form or as a full-length two-domain 65 kDa helicase-primase38. The composition of the T7 gp4 helicase hexamer is thought to be a mixture of the two forms in vivo, controlling the number of primases present for faster replication and less pausing. Other SF4 helicases, including T4 gp41 and Escherichia coli DnaB, interact with a separately encoded primase at the NTD in an analogous configuration. In those cases, the composition and ratio of helicase to primase is <1:1, and more often recognized as 6:339,40. For DnaB, ATP binding by the motor domain can induce conformational changes within the NTD collar that can regulate partner protein (i.e. DnaC or DnaG) selection41. As can be seen in Figure 1, increasing organismal complexity through the SF3 and SF6 helicases (from left to right) generally increases the size of the NTD to where they have evolved additional β-hairpins and zinc-finger motifs for more stabilized binding of the encircled strand and double hexamer formation42,43. The expanded NTD also provides a platform for control of activity through helicase accessory protein binding (in the case of Cdc45 and GINS44) or activation through phosphorylation by either cyclin-dependent kinase (CDK) or Dbf4-dependent Cdc7 kinase (DDK)45–47.

### Helicase loading and the encircling of DNA

The loading of hexameric helicases at replication origins and the associated steps required for encircling only one strand have been the subject of much debate over the years. What is clear is that the loading of the hexameric helicase generally requires the concerted action of accessory initiator proteins to locally melt duplex DNA and facilitate encircling of DNA. However, phage T7, mitochondrial Twinkle, and SV40 L-Tag helicases can load onto circular dsDNA and facilitate encircling of DNA. However, phage T7, mitochondrial Twinkle, and SV40 L-Tag helicases can load onto circular dsDNA on their own46–50. Within the three domains of life, the core ATPase activity and ordered assembly of replication initiation factors seem to be preserved to control the start of DNA synthesis51,52 (Table 1). In bacteria, the initiator, DnaA, forms a multimeric right-hand filament at the replication origin, oriC, to induce unwinding or melting at an A-T rich DNA unwinding element (DUE) (Figure 2A)48–50. Afterwards, the DnaB helicase is loaded on the top and bottom strands by concerted activities of DnaC and DnaA49–50. Once loaded,
the primase, DnaG, interacts with DnaB, displaces DnaC, and activates unwinding. The association of an accessory helicase, Rep, with DnaB may aid in replication fork progression.

In archaea and eukaryotes, the binding of the origin recognition complex (ORC1-6) and Cdc6 to origins of replication is necessary for loading MCM2-7/Cdt1 complexes onto dsDNA to generate a pre-replication complex (Pre-RC) (Figure 2B). The precise structural conformations and dynamics of MCM loading are not fully known, but the steps and components for assembly and activation of the eukaryotic MCM2-7 complex have been recently biochemically reconstituted in vitro, providing significant insight into the process. In archaea, the homohexameric MCM complex exists as a closed ring in solution and would require initiation factors to stimulate opening into a helical conformation onto DNA. Alternatively, increases in temperature for these model hyperthermophilic archaeal MCMs may provide the thermal energy required for the destabilization of a subunit interface required for loading. The eukaryotic MCM2-7 helicase appears to be naturally open, with a labile 2-5 interface that can be trapped by ORC/Cdc6. The ORC-6 complex is arranged in a two-layered cracked ring that encircles DNA and uses the helix-turn-helix domains to engage the MCM2-7 hexamer in a proposed ring-ring interaction, in a manner similar to the loading mechanism of clamp/clamp-loader complexes onto dsDNA. The organization of the ORC complex also appears to be regulated and exists in either an autoinhibited ATP-bound form that precludes DNA binding or a proposed active form that requires a large conformational change in ORC1 that makes the complex competent for encircling DNA. Afterwards, the first MCM2-7 hexamer is loaded through direct interactions of MCM6-Cdt1 with the ORC1-6/Cdc6 complex. The second MCM2-7 hexamer is loaded through contacts between the NTDs of the first loaded MCM2-7 hexamer, rather than through interactions with the ORC1-6/Cdc6 complex. This generated the double hexamer.

**Figure 2. Assembly of Hexameric Replication Helicase at Origins.** A) Loading of the bacterial DnaB helicase by the loader, DnaC, requires the destabilization of a DNA unwinding element (DUE) by the initiator protein DnaA. DnaB complexes with the primase, DnaG, to translocate along the lagging strand unwinding DNA ahead of the replication fork. B) Loading of the eukaryotic MCM2-7/Cdt1 complex requires the initial binding of the ORC complex (ORC1-6) and Cdc6. Interactions with Cdc6 and ORC with the CTD of MCM2-7 directs the adjacent loading of the first hexamer and dissociation of Cdc6 and Cdt1. Subsequent loading of the second hexamer is thought to proceed through direct interactions between the NTD of MCM2-7 to form the double hexamer. Activation of the helicase includes CDK phosphorylation of Sld2 and Sld3 to promote interaction with Dpb11 and stimulate DDK phosphorylation of MCM2/4/6 and recruitment of GINS and Cdc45 to form the CMG complex. Opening of the CMG complex and exclusion of the nontranslocating strand from the central channel activates unwinding and translocation on the leading strand. Gray and black boxes represent major and the foremost, respectively, queries remaining regarding structural conversions of helicases at origins.
hexamer complex, which has been known for years for MCM (and SV40 L-Tag)\textsuperscript{42}, but whether the double hexamer represents an active unwinding unit or an intermediate in the loading process was not known.

After loading of the MCMs, a series of steps are required to form the active unwinding complex (Figure 2B). ATP hydrolysis by Cdc6 and ORC1 causes the dissociation of Cdc6 and Cdt1\textsuperscript{93}. Subsequent phosphorylation of Sld2 and Sld3 by CDK promotes Dpb11 (DNA polymerase B-associated protein) to interact with MCM2-7\textsuperscript{94–96}. Phosphorylation of Sld3, in particular, recruits Cdc45 and GINS (Sld5, Psf1, Psf2, Psf3) to MCM2-7 and stimulates the DDK-dependent phosphorylation of MCM2\textsuperscript{97}, as well as MCM4/6\textsuperscript{98,99}. These phosphorylation events allow opening of the MCM2/5 interface to extrude ssDNA that remains bound to Sld2/Sld3/Dpb11\textsuperscript{100,101}. The active unwinding CMG complex or “unwindsome” is formed through the association of Cdc45 and GINS with MCM2-7 at the labile MCM2-MCM5 interface\textsuperscript{102} along the waist between the NTD and the CTD\textsuperscript{44,103}. Cdc45 in particular blocks the MCM2-5 gate and prevents the loss of DNA from the central channel\textsuperscript{103}. In the reverse mechanism, Cdc45 may also be important for converting MCM2-7 encircled on dsDNA to encircling only a single DNA strand while excluding the other. Formation of the CMG complex widens the gap between the NTD and CTD at MCM2 and MCM5 while concomitantly narrowing the interface at the opposite MCM4 and MCM6 subunits. This induced spiral configuration may contribute to coupled ATP hydrolysis, propagating a conformational change through the MCM2-7 complex to translocate along and unwind duplex DNA\textsuperscript{44,103}.

Interactions with DNA: views of the encircled strand
X-ray structures of hexameric helicases with oligonucleotides bound in the central channel (E1, \textit{E. coli} Rho [EcRho], \textit{Bacillus stearothermophilus} [Bst] DnaB, and \textit{Pyrococcus furiosus} MCM [PfuMCM]) have informed our understanding of the contacts and conformations required for translocation along ssDNA. In these co-crystal structures, ssDNA is bound in a helical conformation in the central channel making direct contacts with each subunit (Figure 3). For E1 and Rho, the hexameric ring is proposed to remain closed, but conformational changes between subunits, coupled with sequential ATP binding and hydrolysis around the ring, direct ssDNA through the central channel through contact with DNA binding loops in a staircase motion\textsuperscript{13,38,104}. Each hexamer subunit interacts with one nucleotide of the oligo, predating a step-size of one nucleotide per ATP hydrolyzed. This is consistent with the measured step-size of T7 gp4 of one base-pair unwound per dTTP hydrolyzed\textsuperscript{105}. For DnaB, the crystal structure resembles more of a lock washer, where similar conformational changes throughout the quaternary structure facilitate movement, with a step-size of two nucleotides per ATP hydrolyzed, maintaining a cracked ring structure\textsuperscript{106}. The ssDNA bound to the archaeal MCM seems to be trapped in a lateral orientation around the interior of the NTD, possibly identifying specific contacts during activation or unwinding, implying a step-size greater than one nucleotide per ATP hydrolyzed during translocation\textsuperscript{107}. The EM structure of the intact eukaryotic CMG complex bound to DNA is in a spiral or lock washer conformation\textsuperscript{44}, more similar to the DnaB/ssDNA complex. The crack in the ring between the MCM2 and 5 subunits is again held in check by the Cdc45 and GINS subunits and helps to stabilize
the spiral configuration (Figure 3). Of course, the impact and absolute degree of spiraling, wrapping, or compaction of the encircled strand will need to be validated experimentally, most likely using single-molecule approaches to measure end-to-end distances during loading and unwinding. Almost certainly the flat ring, the asymmetrical spiral, and the cracked lock washer structures represent intermediates during helicase activation and unwinding, but both conformations will also need to be validated further with additional high-resolution structural studies or rigorous biophysical characterizations to monitor the changes in the conformations.

Interactions with DNA: Impact of the excluded strand

Based on the structures and associated biochemical data, the steric exclusion (SE) model, where one strand is encircled and the other is physically excluded, has become the consensus opinion for unwinding for hexameric replication helicases (Figure 4). One caveat to this model is that it generally ignores any contributions of the excluded strand to unwinding. Electrostatic interactions with the excluded strand on the external surface of hexameric helicases have been noted for archaeal MCM and shown to be important for unwinding, contributing to the development of the steric exclusion and wrapping (SEW) model (Figure 4). Others have also noted that both ssDNA and dsDNA have a binding site on the external surface of other helicases. The dynamic and somewhat stochastic nature of unwinding has been attributed to interactions of ssDNA on the external surface of hexameric helicases E1, T4 gp4, and DnaB. In addition, subunits within the unwindsome complexes of SV40 L-Tag and CMG have been shown to interact with the excluded strand for loading and activation of unwinding. Intriguingly, DNA repair helicases have also been shown to sense damage or modifications on the excluded strand and stall unwinding.

For SV40 L-Tag, initial binding to the origin may be directed by internal β-hairpins making direct contacts with the minor groove and specific contacts of the origin binding domain (OBD) to the major groove. It is not currently understood how this initial dimer contact nucleates assembly of a double hexamer around dsDNA. Once loaded, SV40 L-Tag is proposed to convert from encircling duplex DNA to encircling ssDNA by pumping and extruding one strand out through side-channels. Conformational changes within internal β-hairpins may direct the translocating strand through the central channel, while extruding the opposing strand. Using single-molecule experiments, researchers have shown that DNA unwinding proceeds with a single hexamer of L-Tag in a steric exclusion mechanism that is somewhat conformational mobile and able to bypass bulky adducts during translocation. In comparison, a novel mechanism has recently been proposed for E1 where duplex DNA enters the hexamer before being separated internally and forcing individual strands out through separate exits channels. Of course, the steps and dynamics for how these hexameric helicases convert from encircling duplex DNA to single strand separases by pumping DNA out through side-channels, opening of a gate, or through another unknown mechanism need to be visualized directly with high resolution. A recent EM structure shows the leading strand Pol ε ahead of the yeast CMG complex (at the CTD) and suggests a possible model where the encircled leading strand bends back and threads through a side-channel via the MCM2-5/Cdc45/GINS gate to enter the polymerase active site. Alternative models of replisome-DNA interactions were also proposed in this study.

With this emerging information, excluded or opposing strand interactions shown in the SE model (Figure 4) are poised to play multifaceted roles in loading, encircling, unwinding, and sensing of DNA. In the case of archaeal MCM, the external ssDNA binding path in the SE model serpentine along the lateral length of the homohexamer, spanning the CTD and NTD, and even crossing and wrapping across multiple subunits (Graham & Trakselis, unpublished data). The SE model for archaeal MCM is analogous to a socket wrench, whereby the encircling of one strand represents the socket and external interactions with the excluded strand represent the directional ratchet controlling the speed and stabilization of unwinding. Whether the SE model is conserved in all or most hexameric helicases and/or at what stages of helicase assembly it may occur remains to be determined. Currently, we have found that external interactions and dynamics with the excluded strand in the E. coli DnaB helicase are practically identical to that of SsoMCM, despite their opposing polarities (Carney & Trakselis, unpublished). On the other hand, for T7 gp4 the excluded ssDNA interacts with T7 DNA polymerase to generate a replisome complex, where the helicase and polymerases are within one nucleotide of the fork junction and the helicase can make no external contact with the excluded strand. Next, it will be important to determine which of the eukaryotic MCM subunits (MCM2–7) interact specifically with the

---

**Figure 4. DNA Unwinding Models for Hexameric Replication Helicases.** Steric exclusion (SE) model encircles the translocating strand and physically separates the nontranslocating strand outside of the central channel. The steric exclusion and wrapping (SEW) model takes into account specific interactions within the central channel (compaction or spiraling) as well as external interactions (binding or wrapping) with the excluded strand. Blue arrows indicate the direction of movement of the DNA strands with respect to the hexamer.
excluded strand or whether uniform binding sites have evolved on all subunits. It is intriguing that this external contact may help fill in some of the missing steps highlighted in the gray or black boxes depicted in Figure 2.

Although the structural features of the SEW model may be conserved with various hexameric helicases, both the mechanistic roles and molecular interaction sites may be different. In the case of SsoMCM, disruption of external interactions through mutagenesis reduced unwinding efficiency (3'-5')17, but analogous external mutations on DnaB show a stimulation in unwinding (5'-3') (Carney & Trakselis, unpublished). Modification of the excluded strand to a morpholino oligo similarly stimulates the unwinding rate of T7 gp414. Whether these effects result from opposite unwinding polarities or finely tuned control of unwinding rates and maintenance of the excluded strand requires further testing. However, detection and identification of these novel external interactions may provide a unique opportunity to target specific helicases for inhibition. As none of the different hexameric helicase families exhibit significant sequence homology outside of the center P-loop NTPase fold, novel exterior patches (e.g. between the CTD and the NTD) may provide idealized locations for specific targeting of small molecules that perturb unwinding through disruption of excluded strand contacts and avoid direct inhibition of the internal conserved ATPase site.

Future directions
Although significant advances in our understanding of hexameric helicase assembly, loading, and unwinding have been made over the past few years from quantitative biophysical characterizations and various high-resolution structures, more work is required to reveal specific mechanistic steps and transitions. For example, the essential components for the initial loading of hexameric helicases onto DNA are well described, but the conformational changes that occur within the hexamers during the encircling of ssDNA are still unknown. After all these years, the black box in the whole mechanism is still the structural conversion of the helicase from encircling dsDNA to the encircled ssDNA directing the polarity of translocation and unwinding, primarily for SF3 and SF6 enzymes. Although much is known about the loading and activation mechanism in the Gram-negative E. coli system, far less is known about SF4 helicases in the Gram-positive organisms where DnaI acts as the loader or5,125,126 or in bacteria which lack DnaC/DnaI loader homologs altogether27,128.

Although there is a wealth of structural information on the static hexameric helicases themselves, there is still much debate on the mechanics of helicase action. No longer is the focus directly on the structure of the helicase protein itself. Instead, it has switched from identifying conformational changes, transacting proteins, and post-translational modifications that reveal how duplex DNA is destabilized and the path it takes to be excluded. Finally, although the unwinding mechanism of hexameric helicases was thought to be as simple as excluding one strand from the central channel, new information highlighting the specificity and importance of interactions with the nontranslocating strand have central implications on loading and unwinding mechanisms. It is these dynamic conformational steps from the viewpoint of both the helicase and the duplex DNA that will lead to the next transformational leap in replication helicase discovery.

Abbreviations
AAA+, ATPases associated with a variety of activities; ATP, adenosine triphosphate; Cdc6, cell division cycle 6; Cdc45, cell division cycle 45; CDK, cyclin-dependent kinase; Cdt1, chromatin licensing and DNA replication factor 1; CMG, Cdc45/MCM2-7/GINS complex; CTD, C-terminal domain; DDK, Dbf4-dependent Cdc7 kinase; Dpb11, DNA polymerase B-associated protein; dsDNA, double-stranded DNA; EM, electron microscopy; GINS, go-ichi-ni-san, Japanese for 5-1-2-3 for Sld5-Psf1-Psf2-Psf3; MCM, mini-chromosome maintenance proteins; NTD, N-terminal domain; NTPs, nucleotide triphosphates; ORC, origin recognition complex; Pre-RC, prereplication complex; Psf, partner with Sld5; SE, steric exclusion; SEW, steric exclusion and wrapping; SF, superfamily; Sld, synthetic lethal with dpb11; ssDNA, single-stranded DNA.

Competing interests
The author declares that he has no disclosures.

Grant information
Research in the Trakselis laboratory is sponsored by the American Cancer Society (RSG-11-049-01-DMC) and Baylor University. I confirm that the funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements
I am grateful to the broader helicase field as well as individual members of my laboratory (both past and present) for their inspiration, discussions, and collegiality. Thanks to Sean Carney and Brian Graham for critical reading of the manuscript.

References


Open Peer Review

Current Referee Status: ✔ ✔ ✔

Editorial Note on the Review Process
F1000 Faculty Reviews are commissioned from members of the prestigious F1000 Faculty and are edited as a service to readers. In order to make these reviews as comprehensive and accessible as possible, the referees provide input before publication and only the final, revised version is published. The referees who approved the final version are listed with their names and affiliations but without their reports on earlier versions (any comments will already have been addressed in the published version).

The referees who approved this article are:

1. **Smita Patel**, Department of Biochemistry and Molecular Biology, Robert Wood Johnson Medical School, Rutgers University, Piscataway, New Jersey, 08854, USA
   
   **Competing Interests:** No competing interests were disclosed.

2. **James Berger**, Department of Biophysics and Biophysical Chemistry, Johns Hopkins University School of Medicine, Baltimore, Maryland, 21205, USA
   
   **Competing Interests:** No competing interests were disclosed.

3. **Kevin Raney**, Department of Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, Little Rock, Arkansas, 72205, USA
   
   **Competing Interests:** No competing interests were disclosed.