Mechanism of Origin DNA Recognition and Assembly of an Initiator-Helicase Complex by SV40 Large Tumor Antigen

Y. Paul Chang,1 Meng Xu,2 Ana Carolina Dantas Machado,1 Xian Jessica Yu,1 Remo Rohs,1,3,4,5,* and Xiaojiang S. Chen1,2,3,4,*

1Molecular and Computational Biology Program, Department of Biological Sciences
2Graduate Program in Genetics, Molecular and Cell Biology
3Department of Chemistry
4Norris Comprehensive Cancer Center
5Department of Physics and Astronomy
6Department of Computer Science
University of Southern California, Los Angeles, CA 90089, USA
*Correspondence: rohs@usc.edu (R.R.), xiaojiang.chen@usc.edu (X.S.C.)
http://dx.doi.org/10.1016/j.celrep.2013.03.002

SUMMARY

The DNA tumor virus Simian virus 40 (SV40) is a model system for studying eukaryotic replication. SV40 large tumor antigen (LTag) is the initiator/helicase that is essential for genome replication. LTag recognizes and assembles at the viral replication origin. We determined the structure of two multidomain LTag subunits bound to origin DNA. The structure reveals that the origin binding domains (OBDs) and Zn and AAA+ domains are involved in origin recognition and assembly. Notably, the OBDs recognize the origin in an unexpected manner. The histidine residues of the AAA+ domains insert into a narrow minor groove region with enhanced negative electrostatic potential. Computational analysis indicates that this region is intrinsically narrow, demonstrating the role of DNA shape readout in origin recognition. Our results provide important insights into the assembly of the LTag initiator/helicase at the replication origin and suggest that histidine contacts with the minor groove serve as a mechanism of DNA shape readout.

INTRODUCTION

SV40 large tumor antigen (LTag) transforms eukaryotic cells and is essential for viral DNA replication. SV40 replication involves essential cellular replication proteins, including primase and polymerase proteins (Fanning and Zhao, 2009). To initiate eukaryotic DNA replication, multiple initiator proteins, such as Orc, cdc6, cdt1, and GINS, are required for origin binding and helicase recruitment/activation (Méndez and Stillman, 2003). For SV40 replication, however, LTag alone fulfills the functions of these multiple initiator proteins, i.e., origin recognition, melting, and unwinding (Simmons, 2000). Thus, LTag is an integrated initiator and replicative helicase for DNA replication. LTag has three defined domains for replication: an origin binding domain (OBD), a Zn domain, and an AAA+ domain (Gai et al., 2004; Li et al., 2003; Singleton et al., 2007; Figure 1A).

The SV40 core origin DNA for replication (ori) can be divided into two halves (Figure 1B), with each half containing two of the four 5'-GAGGC pentanucleotides (PEN1–PEN4) and an AT-rich (AT) or early palindrome (EP) region (Deb et al., 1986). Each PEN can be bound by one OBD (Bochkareva et al., 2006; Deb et al., 1987). Each half origin supports the assembly of one LTag hexamer, and the full origin supports double hexamer formation (Mastrangelo et al., 1989; Valle et al., 2006). The assembly of the LTag hexamer/double hexamer at the replication origin is coupled with ori DNA melting and unwinding (Borowiec et al., 1990; Borowiec and Hurwitz, 1988; Gai et al., 2004; Joo et al., 1998; Li et al., 2003; Mastrangelo et al., 1989; Shen et al., 2005; Sreekumar et al., 2000; Valle et al., 2006).

Despite advances in characterizing the LTag helicase domain structure and the structure of individual OBDs interacting with the PEN origin sequence (Bochkareva et al., 2006; Meinke et al., 2007), information is lacking regarding how the OBD, Zn domain, and AAA+ domain (the helicase domain) together recognize each half of the ori during the assembly of an LTag hexamer. Thus, an LTag structure containing OBD, Zn, and AAA+ domains can address the problem of the origin recognition and assembly mechanism in a way that cannot be addressed by studying the separate OBD or AAA+ helicase domains. Here we describe the crystal structure of the EP half origin bound by a dimeric LTag construct that contains OBD, Zn, and AAA+ domains. Our structure reveals several unexpected features in the protein-ori DNA interactions, including the inversion of a domain to contact ori DNA, a previously unidentified ori sequence for OBD recognition, and a particular DNA structural trait that is critical for recruiting the initiator/helicase (i.e., shape readout for DNA-protein recognition). Our results provide detailed mechanistic insights into how LTag initiator/helicase assembles around ori DNA, which should have broad implications for understanding the initiation of replication in other eukaryotic replication systems.
The detailed insights into LTag-DNA binding provided by our structure reveal a critical role of histidine residues in protein-DNA recognition. We observe that the histidine residue of the AAA+ domain interacts with the ori DNA using a mechanism similar to that previously observed for arginine residues (Rohs et al., 2009). On the basis of the analysis of all available crystal structures of protein-DNA complexes, we previously found that arginines can recognize minor groove shape through a shape-dependent electrostatic potential. Here, usingcocrycrystal structures of other protein-DNA complexes, we demonstrate that histidines can play a similar role in DNA shape readout.

RESULTS

Overall Architecture of the Dimeric LTag-dsDNA Complex

The LTag construct (residues 131–627) used for cocystalization with ori DNA contains three separable domains: the OBD, the Zn domain, and the AAA+ domain (Figure 1A, blue box; Figure S1). This LTag131-627 construct (“LTag” hereafter; Figure S2) is crystallized as a dimer in complex with the 32bp EP-half ori DNA (boxed in blue) used for cocrystallization.

The electron density corresponding to the entire 32bp EP-ori DNA is sufficiently well featured to allow unambiguous assignment of the DNA nucleotide sequence (Figures 2A and 2B). As a result, the orientation and register of the EP-ori DNA in the complex are well defined. The average rise and helix twist between adjacent base pairs are 3.4 Å and 35.0°, respectively, and thus are close to those reported for standard B-form DNA (3.32 Å and 35.4°; Olson et al., 1998). The EP-ori DNA interacts with all three LTag domains (OBD, Zn, and AAA+; Figures 2E and 2F, blue arrows), with each domain having unique binding features as described in later sections.

Sequence-Specific Recognition of Ori DNA by OBDs of the LTag Dimer

One of the most surprising observations we made in the LTag-ori DNA complex structure was the manner in which the two OBDs of the dimer recognize the PEN sequences. Instead of the anticipated binding of PEN1 and PEN2 by OBD1 and OBD2 of the two subunits, only PEN1 is bound by OBD1, leaving PEN2 untouched (Figures 2D and 2E; OBD1 in pink, PEN1 and PEN2 labeled). Unexpectedly, OBD2 interacts with a GpC dinucleotide (G16G17/C14G17, termed “hidden site” hereafter; Figures 1B and 2D). Equally surprising was the finding that OBD2 is inverted by almost 180° compared with OBD1 (Figures 2D and 2F), allowing the two OBDs to bind the DNA from opposite faces with orientations inverted relative to each other (Figures 2C and S1) even though the AAA+ domains of the dimer are rotated by only ~60° relative to each other. The flexible linker (residues 258–267) connecting the OBD and Zn domain allows this inversion of OBD2. However, structural modeling indicates that the linker is too short to allow OBD2 in the dimer to reach PEN2, even in its fully extended conformation.

OBD1 interacts with PEN1 in the major groove in a fashion similar to the OBD-PEN interaction reported previously (Bochkareva et al., 2006; Meinke et al., 2007). In particular, the loop region between α1 and β1, and residues just N-terminal of α3 together create a surface that interacts with PEN1 in the major groove (Figures 2E, 3A, and S3A). PEN1-G21A22G23G24 is in contact with R154, S152, and N153 (Figure 3A, highlighted). Additionally, the two consecutive cytosines (C24G25) of the reverse strand of PEN1 interact with the OBD1 protein backbone, whereas the 5‘ guanine (G25) of the complementary strand interacts with R204 (Figure 3A). These contacts, especially the bidentate hydrogen bonds between arginine and guanine bases, lead to highly specific interactions with PEN1, which is frequently used for sequence readout through base-specific contacts (Rohs et al., 2010). Overall, 7 bp of PEN1 are recognized by OBD1 through base readout (Figure S1).

The unexpected OBD2 binding to the hidden site involves substantial interactions with the major groove edges of the C16G17 dinucleotide and its complementary sequence, G16C17 (Figure 3B, hidden site G16C17/C16G17 labeled in Figures 1B and 2D). Therefore, the interactions with the hidden site are sequence specific, but with less specificity than the OBD1-PEN1 interactions. Because of the lack of a full PEN binding site for OBD2, R154 that interacts with the G21 base in OBD1 swings away from the DNA in OBD2 and forms a hydrogen bond with N227 (Figure 3B). Similarly, S152 that binds the A23 base in OBD1 swings away from the DNA in OBD2 (Figure 3B). The loop region containing A149, V150, and F151, which interact with the DNA backbone in the case of OBD1, reorients away from the DNA backbone in OBD2 (Figures 3B and S3B).
Previous biochemical work showed that LTag double hexamers prefer to bind to PEN1 for the EP half ori (and PEN3 for the AT half; Joo et al., 1998). For the LTag double-hexamer assembly, PEN1 (on the EP-ori DNA) and PEN3 (on the AT-ori DNA) alone are sufficient, and PEN2 (EP-ori) and PEN3 (AT-ori) are dispensable (Joo et al., 1998). Our structural data are
consistent with these results in that the LTag dimer assembly binds to PEN1 and not PEN2. It has also been shown that PEN1, PEN3, and the EP region together constitute a strong assembly unit for two head-to-head hexamers, whereas PEN2, PEN4, and the AT region constitute a weak alternative assembly unit (Sreekumar et al., 2000). We were not able to obtain a crystal of dimeric LTag with the AT-half ori DNA under the tested conditions even after extensive exploration, probably reflecting a different strength of the protein interactions with the AT-half region.

Zn Domain Interactions with Ori DNA

Interactions between the Zn domains and the EP-ori DNA occur mainly through the two DNA backbones via charge-charge interactions in the major groove (Figures 2E and 2F). The Zn domain of subunit 1 (Zn-1) contributes most of the binding interface with DNA, with a buried surface of 980 Å² (versus 130 Å² for subunit 2). The Zn domain of subunit 2 (Zn-2) has only one residue that binds the DNA backbone; all other interactions are through Zn-1 (residues in pink for Zn-1 and green for Zn-2 in Figure 3C). A total of 11 residues from Zn-1 contact the DNA backbones in a nonsequence-specific manner (Figure 3C). Such extensive interaction with the DNA backbone through the Zn domain may provide an anchoring point on the DNA that allows the AAA+ domain to bind the AT EP region for assembly.

AAA+ Domain Interactions with DNA

Unlike the interactions of the two OBDs with ori DNA at the major groove, the AAA+ domains of both subunits 1 and 2 interact with the minor groove (Figures 3D–3F). For subunit 1, the four residues at the β-hairpin tip (K512, H513, L514, and N515) and the adjacent helix (F459 on α14) contact DNA (Figure 3E). Together, these residues interact with three phosphate groups, two sugar
Electrostatic Interaction Anchors a Histidine Pair in the Minor Groove

Of particular interest is how the H513 residues of the two subunits interact with EP-ori DNA. It is evident that, for the most part, the two subunits contact the ori DNA differently (compare Figures 2E and 2F). However, the H513 residues on the β-hairpin from both subunits intrude into the minor groove in a nearly identical fashion with both imidazole rings lying in approximately the same plane, following the helical path of the minor groove. The N-N distance between the two H513 imidazole groups is ~2.7 Å, a distance indicative of the formation of a hydrogen bond (Figure 4A), which may further stabilize the interactions within the ternary complex of the LTag dimer and ori DNA.

Importantly, the H513 residues anchor at the position of the ori DNA that was shown to be melted upon hexamer assembly (Borowiec and Hurwitz, 1988). The minor groove region bound by the two H513 side chains is narrower than its adjacent regions (Figure 4B, blue line), with a minimum width of 4.5 Å (versus 5.8 Å for the minor groove width of standard B-DNA). This narrower width of the minor groove could be induced by protein binding, or it could be an intrinsic structural feature of the ori DNA sequence. To distinguish between these two possibilities, we carried out Monte Carlo (MC) simulations (see Experimental Procedures for details) of the DNA structure using the origin sequence. The result illustrates that this DNA region of the H513 contacts is characterized by an intrinsically narrow groove in the absence of protein binding (Figure 4B, green line).

The negative electrostatic potential in the center of the minor groove is enhanced as the groove width decreases (Rohs et al., 2009). The narrowed minor-groove region where the H513 residues bind has an electrostatic potential that is ~2 kT/e more negative than the potential in the wider minor groove of adjacent regions (Figures 4A, red mesh, and 4B, red line). Thus, the binding of H513 residues to the EP-ori sequence is characterized by a shape readout mechanism whereby positively charged protein residues bind to intrinsically narrow regions of the minor groove with enhanced negative electrostatic potential (Rohs et al., 2009, 2010).

After observing the origin recognition mode of the LTag His513 residue in this structure, we sought to determine whether the observation that histidine residues recognize narrow minor groove widths and enhanced negative electrostatic potential is of a more general nature. We analyzed the minor groove width and electrostatic potential for various structures that are part of the IFN-β enhanceosome (Escalante et al., 2007; Panne et al., 2004, 2007). This analysis revealed that conserved histidine residues from IRF-3 (His40) and IRF-7 (His46) that intrude into the minor groove consistently bind regions of narrow minor groove and enhanced negative electrostatic potential (Figure S4).

Conformation of LTag Subunits in the Dimer Structure

Our LTag dimer-DNA cocystal structure reveals how the multidomain of LTag that contains OBD, Zn, and AAA+ domains in a single polypeptide is arranged when it binds to DNA. The structure reveals that, besides the orientation difference between the two OBDs, the two subunits also show different relative orientations between the Zn and AAA+ domains (Figures 5A–5D). In particular, α10 of subunit 1 (magenta in Figure 5A), a long helix connecting the Zn and AAA+ domain, shifts away from the Zn domain by 6 Å compared with the conformations in nucleotide-free, ADP-bound, and ATP-bound states (Figure 5A), whereas α10 of subunit 2 is rotated toward the Zn domain by 13° compared with other LTag states (Gai et al., 2004). These displacements of α10 alter the interactions between the Zn and AAA+ domains (Figures 5B and 5C).
resulting in different degrees of rotation of the AAA+ domain, which brings the β-hairpins of both subunits into close proximity to insert into the minor groove of the EP-region of DNA (Figures 5A and 5H, boxed region). Also, the positions of the β-hairpins move upward toward the Zn domain even more so when compared with those in various nucleotide states (Figure 5A, boxed), which suggests that ori DNA binding induces conformational changes even greater than those caused by nucleotide binding in the LTag hexamer.

Interestingly, some residues buried within the interface between neighboring subunits as shown in previous LTag hexamer structures are now exposed on the dimer’s outer surface and form contacts with DNA. For example, N515 and K271 of subunit 2 are buried at the dimer interface with subunit 1, and interact with subunit 1 in the hexamer. However, N515 and K271 of subunit 1 are on the exposed dimer surface and interact with DNA (Figures 3C and 3D). Conversely, R456 of subunit 1 is part of the dimer interface that interacts with subunit 2 (Figure 5H), but R456 of subunit 2 is located on the exposed surface and utilized for DNA binding (Figure 3F). Thus, these protein residues that are involved in protein-protein interactions for hexamer formation can also be used for DNA binding during helicase assembly. Such dual binding to either another protein subunit or the DNA of these LTag residues may have implications for the assembly of LTag subunits on ori DNA.

Figure 5. Helicase Domain Conformations and Dimer Interface
(A) Superimposition of helicase domains in various nucleotide-bound states (white) and the two subunits in the DNA-bound state from this structure (pink and green), which reveals shifting of the AAA+ domains and β-hairpins, illustrating the ability of multiple conformational switches that are critical for a motor protein such as LTag to melt origin and unwind DNA. Note that α10 adopts different orientations for subunits 1 (green) and 2 (pink), which generates a rotation of the AAA+ domains to bring their β-hairpins (and H513 residues) into close proximity to the DNA minor groove (see the box in H here, and Figure 3D).
(B–D) Change of contacts between Zn (K281/P311) and AAA+ domains (D367/R371/I374 on α10 in the absence (B: close contact) or presence (C: detached; D: altered contact) of DNA.
(E and F) Two views of the protein dimer interface show the tight interface between the two subunits around the β-hairpin regions in the AAA+ domain.
(G and H) Close-up views of the extensive interactions between the two AAA+ domains. The close proximity of the two H513 residues is shown in the inset (H).

Dimer-DNA Complex: An Assembly Intermediate without Melting Activity
In the LTag dimer-ori DNA complex, the DNA is in B-form conformation. The obviously narrower width of the minor groove where the β-hairpins bind appears to be a pre-existing structural feature of the ori DNA (Figure 4). No obvious deformations of the DNA are induced by LTag binding. This observation is similar to previous findings from the crystal structures of archaeal ORC initiator-dsDNA complexes (Dueber et al., 2007; Gaudier et al., 2007), which revealed no severe deformation or melting of DNA. In contrast, biochemical studies of E1 helicase/initiator, a distant homolog of LTag from papillomavirus, suggested that a trimer assembly intermediate is capable of melting ori DNA (Schuck and Stenlund, 2005). No trimer assembly for LTag has ever been observed. However, given the LTag dimer assembly intermediate observed here, we asked whether the stable dimer intermediate could be captured in vitro and, if so, whether such a dimeric intermediate would be similar to E1 in terms of its ability to melt ori DNA.

To address this question, we designed specific mutations to capture a stable dimer intermediate of LTag in solution, as it was previously shown that wild-type (WT) LTag only exists in either a stable monomeric or hexameric form, and no stable dimer or other intermediate oligomers can be observed in solution (Gai et al., 2004; Li et al., 2003). Two mutants (mut1:
L286D/R567E, and mut2: V350E/P417D) are designed to introduce mutations on only one surface of each mutant within the interface, so that intersubunit interactions are disrupted (Figure 6A). We predicted that, when present alone, these two mutants would be monomeric. When mixed together, however, each mutant contributes one native side to form a dimer. The two outer surfaces of the dimer carry mutations to prevent further oligomerization (Figure 6A), thus forming a stable dimer. Gel filtration chromatography results indicated that each individual mutant was indeed monomeric, and the equimolar mixture of the two mutants formed stable dimers (Figure 6B). As a convenient way to distinguish the two mutant proteins within the stabilized dimer intermediate, we added a 14 amino acid C-terminal tail on mut1 (L286D/R567E) to make the protein ~2 kD larger (slower migration on SDS-PAGE), denoted as L286Dt/R567E. SDS-PAGE analysis of such a dimer revealed a 1:1 stoichiometry of mut1:mut2 based on gel quantification (Figure 6B, inset).

To further validate the intersubunit interface of the dimer, we conducted an ATPase assay to test whether the ATP pocket that forms at the dimer interface can hydrolyze ATP. The results showed clearly that each mutant alone had severely disrupted ATPase activity, as expected for a monomeric form, whereas the mixture of the two mutants had WT-level ATPase activity (Figure 6C), which indicates that the stabilized mutant dimer intermediate reconstitutes a WT ATP pocket at the dimer interface, as in the case of a hexamer of the WT protein. A helicase assay showed that the two LTag131-627 mutants individually or as a mixture no longer unwound the ori DNA (Figure 6D, lanes 3–10), consistent with the fact that only functional hexamers of WT LTag can unwind DNA (lanes 11–12).

With the stabilized dimer intermediate at hand, we assayed the origin-melting activity of the dimer intermediate in the presence or absence of ATP. We used the potassium permanganate (KMnO4) reactivity assay on a 92 bp SV40 ori containing dsDNA substrate. The results showed that the dimer intermediate had no detectable DNA melting activity (Figure 6E, lanes 2–5), but the WT LTag131-627 showed the expected ori DNA melting activity (Figure 6E, lanes 6–9). Thus, the dimeric state represents an assembly intermediate that can recognize or bind to ori DNA but is not capable of melting it.
DISCUSSION

In this work we describe the cocrystal structure of a dimeric LTag in complex with the EP-ori DNA, which represents the initial stage of assembly toward a hexameric initiator/helicase complex at the ori DNA for replication initiation. This crystal structure reveals the multiple domain organization of the OBD, Zn domain, and AAA+ domain in one polypeptide of LTag in complex with ori DNA. The detailed molecular interactions of LTag with the EP-ori DNA provide mechanistic insights into origin recognition, recruitment, and assembly of LTag at the initial stage of viral DNA replication in eukaryotic cells.

Importance of the Zn Domain in Assembly of LTag at the Replication Origin

The Zn domain is essential for hexamerization of the helicase domain (Li et al., 2003). In this dimer-DNA structure, the two Zn domains interact with each other in a manner similar to those observed for LTag hexamer structures in various nucleotide states (Gai et al., 2004; Li et al., 2003). Thus, the relative orientation between adjacent Zn domains appears to be independent of the different states in ATP binding, hydrolysis, DNA interaction, and LTag assembly stages.

The LTag dimer-DNA structure reveals that the Zn domain and DNA interface accounts for 25% of the total protein-DNA interface, suggesting that it plays an important role in the initial assembly of the LTag dimer at the origin. Thus, for LTag assembly around the ori DNA, the Zn domain serves two essential functions: (1) to stabilize the intersubunit interface of the dimer, and (2) to reinforce the OBD-initiated, sequence-specific origin binding through interactions with the DNA backbones.

Role of β-Hairpin in Ori DNA Binding

Even though the interactions of the two LTag subunits with the EP-ori DNA are largely different from each other in three domains (i.e., the OBD, Zn, and AAA+ domains), the β-hairpin tip residues (K512 and H513 of the two subunits) form almost identical interactions with DNA in the minor groove of the EP sequence region. These interactions may have several implications.

First, the β-hairpins on the two subunits intrude into the minor groove side by side, following the helical path of the groove. This arrangement is stabilized through electrostatic interactions with the DNA and, likely, a hydrogen bond between the H513 residues of both subunits. Even though the details differ, the helical arrangement of the two β-hairpins in this LTag dimer resembles that in the E1 hexamer structures in which the six equivalent β-hairpins form a helical arrangement (Enemark and Joshua-Tor, 2006; Sanders et al., 2007). An intriguing question is whether additional subunits will continue to position their β-hairpins along the helical path of the minor groove during hexamer assembly around the dsDNA. Normally, duplex DNA in B form is characterized by 10 bp per helical turn (360°). If six LTag subunits bind along six consecutive base pairs and then form a circle, it should distort the duplex DNA during the process and lead to DNA melting.

Second, the two β-hairpin H513 residues insert into a minor groove region with a minimum groove width of 4.5 Å, compared with the standard B-DNA groove width of 5.8 Å (Rohs et al., 2009). The immediate effect of such a narrowed minor groove is that the electronegativity is increased by ~2 kT/e, forming a binding site that attracts the H513 residues more effectively than other regions of the DNA. Our MC simulation results show that the minor groove width of the unbound ori DNA has a minimum groove width of 4.6 Å, matching that in the cocrystal structure. This result indicates that the narrower minor groove is not induced by protein binding but is an inherent structural feature of the particular origin sequence. Thus, we conclude that it is the particular ori DNA structure that is recognized by the β-hairpin H513 residues. This result is consistent with a previous study that showed the importance of the ori DNA sequence by demonstrating that mutation of this DNA region knocks out viral DNA replication (Deb et al., 1987). In those experiments, an A/T base pair was replaced by a G/C base pair, which removed the TpT dinucleotide that supports the narrowing of the minor groove through negative propeller twisting. In addition, histidine residues can better intrude into the minor groove in the vicinity of A/T base pairs because the bulky guanine amino group is absent.

Third, it is known that LTag can translocate along dsDNA and single-stranded DNA (ssDNA), but it is unknown how the direction is determined when LTag hexamers translocate along dsDNA without unwinding. Previous work showed that the β-hairpin is the major lever that shifts back and forth along the central channel upon ATP binding and hydrolysis (Gai et al., 2004), which should be coupled to translocating ssDNA and dsDNA through the central channel. In the complex with ori DNA, the two H513 residues on the β-hairpin insert into the minor groove to interact with both DNA strands. Interestingly enough, the adjacent K512 residues of both subunits interact with only one of the two strands (the strand in yellow in Figure 3D). This tracking on only one strand by two (or possibly more) β-hairpin K512 residues at a given time when binding to dsDNA may provide the directionality needed for translocation not only on ssDNA but also on dsDNA. Consistent with this hypothesis, mutation of K512 completely abolished translocation on dsDNA and ssDNA (Figure S5).

DNA Shape Readout through Histidine Contacts

The electrostatic attraction of histidine to DNA requires its protonation, and histidine residues are frequently protonated in the highly charged environment of DNA (Joshi et al., 2007). In the case of the LTag dimer, the presence of a hydrogen bond is indicative of one histidine being protonated to provide a hydrogen-bond donor, while the second histidine is unprotonated. Thus, the total charge of the histidine pair is likely +1. The H513 pair establishes two crucial interactions within the ternary complex: (1) a hydrogen bond within the LTag dimer and (2) electrostatic attraction of the His513 pair into the minor groove. This readout mechanism of sequence-dependent DNA shape may explain the initial anchoring of the subunits to the DNA, which may be of key importance for LTag assembly.

Histidine was previously observed to bind a narrow minor groove region of the binding site of the Hox protein Sex combs reduced (Scr) but in tandem with an arginine residue (Joshi et al., 2007), forming a hydrogen bond between its guanidinium...
group with the histidine. Similar to the situation in the LTag dimer, this hydrogen bond can only form when the histidine is not protonated, assigning the Arg3-His-12 pair a total charge of +1. This conclusion is supported by the fact that a His-12 mutant of Scr had only a minor effect on binding and in vivo activity compared with a mutant of the charged Arg3 (Joshi et al., 2007). In LTag dimer binding to DNA, however, histidine takes on a key role in protein-DNA recognition without the presence of an arginine side chain. The observation that histidine of LTag on its own uses the mechanism of DNA shape readout to bind DNA is also present in other biological systems based on our analysis of cocrystal structures of the IFN-β enhanceosome (Escalante et al., 2007; Panne et al., 2004, 2007; Figure S4). We conclude that histidine in general uses a readout mechanism for protein-dsDNA interactions in a manner similar to that previously described for arginine (Rohs et al., 2009).

The critical role of histidine in the interaction of LTag with DNA is also apparent from its high sequence conservation, as the H513 residue is highly conserved among all LTag proteins in polyomaviruses and within the distantly related E1 helicase of papillomaviruses (see Figure S2C). Published results show that mutation of H513 on LTag to alanine affects DNA unwinding and origin melting (Kumar et al., 2007; Shen et al., 2005). Additionally, mutations of the equivalent H513 to alanine in E1 helicase disrupted ori DNA binding and unwinding (Liu et al., 2007; Schuck and Stenlund, 2005). The structural data presented here, together with these mutational results in the literature, support the observation that the highly conserved histidine residue plays a crucial role in the interaction of LTag with ori DNA to initiate DNA replication.

**Interactions between OBD and PEN Sequences for Assembly**

It has been a long-standing question how the four PEN repeats (PEN1–PEN4; Figure 1B) of the full ori DNA are utilized in the assembly of a double hexamer to initiate replication (Joo et al., 1998; Sreekumar et al., 2000). Previous studies revealed that each half origin (i.e., the EP-ori carrying PEN1 and PEN2, or the AT-ori with PEN3 and PEN4) can support an efficient hexamer assembly. Here, our dimer-ori DNA structure reveals that only PEN1, and not PEN2, is bound by OBD1. The linker connecting the OBD and Zn domain of subunit 2 in the dimer is not long enough for OBD2 to reach PEN2. However, OBD2 inverts by 180° to bind to a hidden site, yielding a well-positioned OBD2, because the B-factors of OBD2 are comparable to those of OBD1.

The hidden-site sequence of the ori DNA (G16C17/G16C17; Figure S6) revealed by OBD2 revealed in the structure here is consistent with an earlier mutational study by Deb et al. (1986), which showed that G16C17 and C16G17 are important for viral DNA replication. Their data showed that the G16C17 base pair is more critical than the C17G17 base pair, which is consistent with the more extensive interactions of OBD2 with the G16C17 compared with the C17G18 base pair observed in the structure (see the more extensive hydrogen bonds with the G16C17 base pair in Figure S6B). Furthermore, the observation that PEN1 alone is bound by OBD is also consistent with prior biochemical studies showing that the assembly of a double-hexamer requires only PEN1 and PEN3 on a full origin, and PEN2 and PEN4 are dispensable (Joo et al., 1998; Sreekumar et al., 2000). However, because all four PEN sequences are required for viral infectivity (Joo et al., 1998; Sreekumar et al., 2000), further investigations will be required to elucidate the exact role of PEN2 and PEN4 in viral DNA replication in vivo.

**LTag Residues Interact with Both Protein and DNA during Assembly**

At the initial stage of LTag assembly on ori DNA, specific protein-DNA interactions likely are essential for recruiting the first LTag monomer or dimer to the origin. In this dimer structure, several dimer interface residues (such as K271, K512, R456, and N515) are present either on the exposed surfaces of the dimer or in the LTag-DNA interface. Thus, these residues have a dual role in forming contacts at either the protein-protein interface (Figure S1C, middle panel) or the protein-DNA interface (Figure S1C, left panel). The DNA binding by these residues is transient, and contacts will be broken so that these residues can bind to newly arrived protein subunits during assembly (Figure S1C, right panel). Therefore, it is likely that such dual-binding residues are very important for the assembly of LTag around the origin because they can provide stability for the first monomer-ori DNA or dimer-ori DNA intermediate complex and facilitate the subsequent recruitment of new subunits for hexamer assembly (see the model in Figure S1C).

In summary, we have determined the crystal structure of a dimeric LTag in complex with the EP-ori DNA. The structure reveals information about how the OBD in the context of a multi-domain LTag polypeptide recognizes the sequences around PEN1 and PEN2 on the EP half origin. We have also illustrated an origin recognition mechanism by the AAA+ β-hairpin, i.e., minor-groove shape readout through histidines that recognize the minor-groove geometry of the EP region. Because the β-hairpin is critical for origin binding and melting (Kumar et al., 2007; Shen et al., 2005), this shape-readout mechanism is critical for the initial assembly of LTag and the melting of origin. The LTag dimer in complex with the EP-ori DNA likely represents a snapshot of an early-stage assembly intermediate for the hexamer/double-hexamer initiator complex, a process that is common to some replicons in eukaryotic cells.

**EXPERIMENTAL PROCEDURES**

**Cloning, Mutations, and Protein Purification**

The LTag131–627 construct was cloned in pGEX-6P-1 vector as an N-terminal glutathione S-transferase (GST) fusion. The LTag protein was expressed in Escherichia coli BL21 cells at 16–18°C. The protein was purified as follows: After cells were lysed, the protein was purified by glutathione affinity column in a buffer containing 50 mM Tris-HCl (pH 8.0), 250 mM NaCl, and 1 mM dithiothreitol (DTT). After the GST was removed by thrombin protease treatment at 4°C for 16 hr, LTag was purified by Superdex-200 gel filtration chromatography. To generate monomeric mutants for formation of a stable dimer intermediate, V350E/P417D and L286Dt/R567E mutants were made, and mutations were confirmed by sequencing the entire LTag insert.

**Cocrystallization and Data Collection**

To obtain EF-half ori DNA for cocrystallization, oligonucleotides (ordered from Operon) 5’-actacttctggaatagctcagaggccgaggcg-3’ and 5’-cgctcgcggctctgagttcagagtgt-3’ were purified using a Mono Q column. The two
Table 1. Data Collection and Model Refinement Statistics of the LTag-EP Origin Structure

<table>
<thead>
<tr>
<th>Cell Dimensions (Space Group P2₁)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a, b, c (Å)</td>
<td>73.63, 128.31, 166.12</td>
</tr>
<tr>
<td>a, β, γ (°)</td>
<td>90, 89.911, 90</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50–2.8 (2.9–2.8)</td>
</tr>
<tr>
<td>Observations</td>
<td>172,018</td>
</tr>
<tr>
<td>Rmerge</td>
<td>8.3 (59.2)</td>
</tr>
<tr>
<td>l/d</td>
<td>12.1 (1.2)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>93.2 (88.4)</td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50.0–2.8</td>
</tr>
<tr>
<td>No. of reflections</td>
<td>65,350</td>
</tr>
<tr>
<td>Rwork/Rfree</td>
<td>22.69/25.49</td>
</tr>
<tr>
<td>B factor (averaged)</td>
<td>58.554</td>
</tr>
<tr>
<td>DNA</td>
<td>55.396</td>
</tr>
<tr>
<td>Root-mean-square deviations</td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.008655</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.20198</td>
</tr>
<tr>
<td>Highest-resolution shell values are shown in parentheses.</td>
<td></td>
</tr>
</tbody>
</table>

cycles of which were considered as the equilibration period. The simulation protocol was identical to the one described elsewhere (Joshi et al., 2007; Rohs et al., 2005).

ATPase Assay and Helicase Assay

We performed the ATPase assay for LTag WT or mutants by detecting the phosphate generated by ATP hydrolysis as described previously (Greenleaf et al., 2008). For the helicase DNA substrate, Y fork-shaped DNA with 44 nt ssDNA tails and a 44 nt duplex was made by annealing two oligonucleotides of 5'-d(ACGTCTGCACGAGCCGTCGTCACCC) and 5'-d(CGTCTGCACGAGGCAGTCGTCGTCACCTCGC) ACGTGCTGACCGC(dT)₄₄. The helicase assay was performed as previously described (Greenleaf et al., 2008). Briefly, ~10 fmol of [γ-32P]-ATP-labeled substrate DNA was incubated with 300 ng LTag in helicase buffer containing 20 mM Tris–Cl pH 7.5, 10 mM MgCl₂, 5 mM ATP, 1 mM DTT, and 0.1 mg/ml BSA for 45 min at 37°C. The reaction was analyzed on 12% native polyacrylamide gel, and the amount of radioactively labeled oligonucleotide was determined by autoradiography.

K MnO₄ Reactivity Assay

A K MnO₄ reactivity assay to test DNA melting activity was performed as follows: A 92 bp dsDNA substrate that contained EP-half ori was generated by annealing two ssDNA. Only the lower strand was 5'-end labeled with [γ-32P] ATP. Then 10–15 fmol DNA was incubated with 100–800 ng LTag protein in buffer containing 20 mM Tris–Cl pH 7.5, 10 mM MgCl₂, 5 mM ATP, 1 mM DTT, and 0.1 mg/ml BSA for 30 min at 37°C. Then KMnO₄ was added to a final concentration of 6 mM and reactions were incubated at 37°C for 2 min. To stop the oxidation reactions, a stop solution containing 160 mM β-mercaptoethanol, 0.3% SDS to 0.3%, and 10 mM EDTA was added to the reaction mixture. Modified DNA substrates were deproteinized first by digestion with proteinase K (20 µg/ml, 60 min at 37°C) and then binding to QIAGEN QIAEX II beads. Piperdine was added to 20% (v/v) on dried beads to cleave DNA substrates at modified nucleotides (30 min at 90°C).

ACCESSION NUMBERS

The atomic coordinates of the cocrystal structure have been deposited in the Protein Data Bank under id code 4GDF.

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2013.03.002.

ACKNOWLEDGMENTS

We thank the staffs of the USC NanoBiophysics core, Argonne National Laboratory synchrotron beamlines 23ID and 19ID, and Berkeley’s ALS beamlines 5.0.2, 8.2.1, and 8.3.1 for assistance with data collection. This work was supported by NIH grants GM080338 and AI055926 (to X.S.C.) and American Cancer Society grant IRG-58-007-51 (to R.R.). R.R. is an Alfred P. Sloan Research Fellow. Y.P.C. did the crystallography work, structural analysis, and initial writing of the manuscript; M.X. obtained the stable dimer intermediate of LTag and performed melting and related biochemical studies; A.C.D.M. performed the computational analyses; X.J.Y. made LTag K512 mutations and performed functional assays; R.R. supervised the computational work and wrote parts of the manuscript; and X.S.C. coordinated the project and wrote the manuscript.
REFERENCES


