The Human Cytomegalovirus UL44 C Clamp Wraps around DNA

Gloria Komazin-Meredith,1,5 Robert J. Petrella2,3,5 Webster L. Santos,2,6 David J. Filman,1 James M. Hogle,1 Gregory L. Verdine,2 Martin Karplus,2,4,* and Donald M. Coen1,*

1Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA
2Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA
3Department of Medicine, Harvard Medical School, Boston, MA 02115, USA
4Laboratoire de Chimie Biophysique, ISIS, Université Louis Pasteur, 67000 Strasbourg, France
5These authors contributed equally to this work
6Present address: Department of Chemistry, Virginia Tech, 107 Davidson Hall, Blacksburg, VA 24061, USA

*Correspondence: marci@tammy.harvard.edu (M.K.), don_coen@hms.harvard.edu (D.M.C.)

DOI 10.1016/j.str.2008.05.008

SUMMARY

Processivity factors tether the catalytic subunits of DNA polymerases to DNA so that continuous synthesis of long DNA strands is possible. The human cytomegalovirus DNA polymerase subunit UL44 forms a C clamp-shaped dimer intermediate in structure between monomeric herpes simplex virus UL42, which binds DNA directly via a basic surface, and the trimeric sliding clamp PCNA, which encircles DNA. To investigate how UL44 interacts with DNA, calculations were performed in which a 12 bp DNA oligonucleotide was docked to UL44. The calculations suggested that UL44 encircles DNA, which interacts with basic residues both within the cavity of the C clamp and in flexible loops of UL44 that complete the “circle.” The results of mutational and crosslinking studies were consistent with this model. Thus, UL44 is a “hybrid” of UL42 and PCNA: its structure is intermediate between the two and its mode of interaction with DNA has elements of both.

INTRODUCTION

Replicative DNA polymerases synthesize long stretches of DNA without dissociating from their templates. To achieve prolonged association with the DNA template, most replicative polymerases depend on accessory protein subunits known as processivity factors. The best-known processivity factors are the “sliding clamps” (Kuriyan and O’Donnell, 1993), which include the β subunit from Escherichia coli (β clamp), the proliferating nuclear antigen (PCNA) from eukaryotes and archaeobacteria, and gp45 from the T4 and RB69 bacteriophage. The overall structures of the various sliding clamps are similar (Jeruzalmi et al., 2002). They form either dimeric or trimeric ring-shaped structures with a central channel large enough to accommodate double-stranded DNA. These ring-shaped proteins do not bind DNA tightly on their own and ordinarily must be loaded onto DNA in an ATP-dependent process by a protein complex called a clamp loader (Jeruzalmi et al., 2002). Once on DNA, the sliding clamp physically tethers the catalytic subunit of the polymerase to the template, thus preventing premature dissociation (Stukenberg et al., 1991).

Another well-studied processivity factor is UL42 from herpes simplex virus (HSV). Unlike sliding-clamp processivity factors, UL42 binds DNA directly with nanomolar affinity as a monomer in a manner that does not require ATP hydrolysis or accessory proteins (Gottlieb and Challberg, 1994; Randell and Coen, 2004; Weissshart et al., 1999). Substitution of arginine residues on the basic “back face” (surface opposite to that which binds the catalytic subunit) of UL42 with alanines results in reduced affinity of UL42 for dsDNA (Randell et al., 2005). These data and the results of further mutational, biochemical, and crosslinking studies (Komazin-Meredith et al., 2008) strongly suggest that the basic back face of UL42 contacts DNA. There is considerable evidence that this DNA binding by UL42 is important for processive DNA synthesis (Chow and Coen, 1995; Randell et al., 2005).

The DNA polymerase encoded by human cytomegalovirus consists of a catalytic subunit, UL54, and an accessory subunit, UL44 (Erltl and Powell, 1992). Although it has not yet been rigorously demonstrated to increase processivity, UL44 interacts with the C terminus of UL54 (Appleton et al., 2004; Loregian et al., 2004a) and stimulates long-chain DNA synthesis (Ertl and Powell, 1992; Loregian et al., 2004a, 2004b; Weiland et al., 1994). Like HSV UL42, UL44 can bind DNA directly and tightly (Appleton et al., 2006; Loregian et al., 2004b, 2007; Weiland et al., 1994). The crystal structure of residues 1-290 of UL44, which contain all its known biochemical activities, has been solved both alone and in complex with a peptide from the C terminus of UL54 (Appleton et al., 2004, 2006). In both structures, UL44 is a head-to-head homodimer in the shape of a C clamp, consistent with its quaternary structure in solution (Appleton et al., 2004). The fold of each monomer is similar to that of other processivity factors, including HSV UL42 (Zuccola et al., 2000) and the sliding clamp PCNA (Krishna et al., 1994), even though these proteins have no obvious sequence homology. Like HSV UL42, the UL44 monomer has a back face (Figure 1) containing numerous basic residues that can potentially form electrostatic interactions with the phosphate backbone of DNA (Appleton et al., 2004, 2006). In the dimer structure, these faces form parts of the surface of a central cavity, which is large enough to accommodate double-stranded DNA (Appleton et al., 2004, 2006). In addition, each UL44 monomer has a 12 residue flexible loop that is disordered in the UL44 crystal structures (Appleton et al., 2004, 2006).
RESULTS

Overview of the Predicted Structure of the UL44/DNA Complex

The modeling was carried out as described in the Experimental Procedures (additional details in Supplemental Data available online). The structure that is predicted to have the lowest energy with the generalized Born molecular volume (GBMV) solvation model (Brooks et al., 1983; Lee et al., 2003) is shown in Figure 1 (coordinates in Protein Data Bank [PDB] format are provided in the Supplemental Data). Other lowest-energy structures obtained with both GBMV and a distance-dependent dielectric function (rdie) (Gelin, 1976; Gelin and Karplus, 1979; Schaefer et al., 1999) are very similar overall, although the loop conformations vary (see below). In the structure shown, the DNA molecule is positioned in the large central cavity of the dimer, which is composed, in part, of the back faces of the monomers. The DNA is tilted slightly; it is 10° from the perpendicular to the plane of the dimer “C,” with the orientation of the tilt being 35° from the axis of symmetry of the dimer. The gap loops complete a “circle” around the DNA.

Electrostatic Protein-DNA Interactions in the Calculated Structure

The large number (28) of positively charged side chains in the gap loops and in the walls of the central cavity of UL44 suggests that the protein can bind to DNA with more than one arrangement of protein-DNA interactions; this is supported by both experimental and computational results (see below). In all the models, the binding between the DNA moiety and the protein involves electrostatic interactions between the phosphate oxygens in a 10 bp segment of the DNA and charged residues from both the central cavity and the gap loops of the UL44 dimer. In the lowest-energy structure from the GBMV calculations, there are a total of 21 hydrogen bonds between the protein and the DNA molecule. Of the 21 hydrogen bonds, 16 are salt bridges, with charged protein donor groups (ends of arginine and lysine side chains) and charged DNA acceptor groups (backbone phosphates). Some of the salt bridges in this structure, as well as those in the lowest- and second-lowest-energy structures from the rdie calculations, are shown in Figures 2 and 3. Eleven of the 16 salt bridges in the GBMV structure involve the binding between the DNA moiety and the protein. Four involve charged side chains on the loops (three arginines and two lysines; there can be odd numbers of DNA-protein interactions because the complex is not entirely symmetric). Of the remaining five hydrogen bonds, four are between the NH groups of the protein backbone and DNA oxygens; and one is between the NH2 group (H22) of a guanine residue (22) and the Nε atom of the R168 residue on the gap loop of monomer 1, whose side chain inserts into the minor groove. The majority of the hydrogen-bond acceptors on the DNA are backbone oxygens (16/20; 4 are O4’ oxygens on the deoxyribose rings, and most are non-bridging (O1P or O2P) backbone oxygens (15/20), particularly in the central cavity (9/11). Interestingly, as shown in Figure 2,
the charged residues on the protein form spiral tracks for the strands of the DNA phosphate backbone.

**Nonpolar Interactions**

In the GBMV structure shown, the complex buries ≈2900 Å² of (solvent-accessible) surface area: ≈1500 Å² of the DNA and ≈1400 Å² of the protein. The total nonpolar buried surface area is ≈1100 Å²; of this, ≈500 Å² is contributed by the DNA and ≈600 Å² by the protein. A comparison of the ten lowest-energy structures from the GBMV and rdie calculations showed that the buried surface areas are very similar in extent, with values ≈3100 ± 50 Å² for the total buried surface area and ≈1100 ± 50 Å² for the nonpolar buried surface area (see Table S1 for a breakdown of buried surface areas by region in the ten lowest-energy structures from the two calculations). These results are consistent with previously described trends for DNA/protein complexes (Norberg, 2003). The hydrophobic contribution to the free energy of binding, \( \Delta G_{\text{hdyd}} \), which is largely entropic, can be estimated as \( \Delta G_{\text{hdyd}} = \gamma \Delta A_{\text{np}} \), where \( \Delta A_{\text{np}} \) is the nonpolar surface area buried by the complex and \( \gamma \) is in the range of ≈22–50 cal/Å² (Ha et al., 1989; Norberg, 2003; Sharp et al., 1991). Hence, the burial of such a large nonpolar surface area by the DNA/UL44 protein complex, as in the cases of the previously studied DNA/protein complexes (Norberg, 2003), implies a large hydrophobic contribution to the binding, which would be in the range of −25 to −60 kcal/mol of UL44 dimer (this value does not include the entropy loss of the flexible loops upon binding). Recent isothermal titration calorimetry experiments (Loregian et al., 2007) measuring the binding of a 18 bp DNA oligonucleotide to UL44 showed a \( T \Delta S \) of binding of ≈22 kcal/mol and a \( \Delta G \) of binding of −11.5 kcal/mol per UL44 monomer. Because binding was found to be endothermic (\( \Delta H = +11 \) kcal/mol; Loregian et al., 2007), it is entropically driven. This can result from the release of bound ions upon formation of charge-charge interactions, from release of water upon formation of hydrophobic interactions (Ha et al., 1989; Norberg, 2003; Record et al., 1978; Sharp et al., 1991), or both. Extrapolation of plots relating salt concentrations and binding constants (Loregian et al., 2007) to the standard state (log \([\text{Na}^+] = 0\)) indicates that a substantial fraction of the free-energy change of the binding of UL44 to DNA is unrelated to release of bound ions (data not shown). This analysis is consistent with the hydrophobic contribution implied by the modeled structure.

**Gap Loops**

The gap loops (which were disordered in the crystal structure) of the two lowest-energy structures from the GBMV calculations are shown in Figures 4A and 4B, respectively; the two lowest-energy structures from the rdie calculations are shown in Figures 4C and 4D. In these structures, the loops are stabilized not only by loop-DNA hydrogen bonds as mentioned above but also by hydrogen bonds within each loop and between the loops. In the lowest-energy GBMV structure, there are a total of seven intraloop hydrogen bonds (three for loop 1, which is green in the
A cluster analysis of the backbones of the 200 lowest-energy conformations for each loop was performed. For each loop there are five significantly populated cluster centers. The structures in each of the five clusters that are closest to the cluster centers are shown in Figures 4E and 4F, with the lowest-energy cluster for each loop highlighted in yellow. In general, the structures tend to extend across the intermonomer gap in the dimer from either side—that is, across the perimeter of the DNA in a direction roughly perpendicular to its longitudinal axis. The tendency of the flexible loops to “wrap around” the DNA in this manner is similar to that observed for the flexible N-terminal arms of the λ repressor dimer (Beamer and Pabo, 1992; Weiss et al., 1984). The average root-mean-square deviation (rmsd) between all pairs of cluster centers of loop 1 is 6.4 Å and for loop 2 is 3.9 Å, so that the structures of monomer 2 show somewhat smaller structural differences. For each of the monomers, the loop structures can be divided into two general types. For monomer 1, the loop either runs parallel to the DNA backbone, usually on the major-groove side, or crosses over the DNA backbone from the major groove to the minor-groove side. For monomer 2, the loop either runs parallel to the DNA backbone, usually on the minor-groove side, or spans the minor groove. Together, the cluster centers form a curved half-tube, or groove, around one strand of the DNA phosphate backbone. The average energies of the different clusters were within ~15 kcal/mol of the minimum for either loop, which is a narrow energy range for this type of structural study. In rigid-rotation mapping, energy differences between alternative states are exaggerated.
because there is incomplete relaxation (Petrella et al., 1998). An rmsd analysis (see Supplemental Data) of the 500 lowest-energy structures showed that the average rmsd between the loops in any two structures is 5.7 Å (±1.7 Å).

Some of the charged side chains of the loops insert into the minor groove, and less frequently into the major groove. Analyses of the insertion of the side chains in the 40 lowest-energy conformations for each loop showed that K171 and K172 of monomer 2 inserted most frequently, in approximately a quarter to a half of the structures. R168 of monomer 1 inserted in ~10% of the structures, and K172 of monomer 1 inserted in ~5%–7% of the structures.

An analysis of the 200 lowest-energy structures gave the same results.

### Hydrogen-Bonding Analysis of Structural Ensembles

A breakdown of the hydrogen bonds by donor and acceptor types for the ten lowest-energy structures in the GBMV and rdie calculations is given in Table 1. Although the specific hydrogen-bonding patterns differ, the total numbers of hydrogen bonds in the structures are about the same (the maximal difference in total hydrogen-bond counts between any two GBMV structures is 5 out of 525 ± 3). Because of the flexibility of the loops, there are a large number of different hydrogen bonds with donors in the gap loops and acceptors in the DNA; there are 204 different bonds in the 2000 lowest-energy structures, and 94 in the 50 lowest-energy structures. However, as shown in Table S2, a smaller number (roughly 10–20) of these hydrogen bonds, most of which involve charged end groups on the lysine and arginine side chains, tend to be common (frequencies of 15%–50%).

As indicated in both Table 1 and Table S3, there tend to be fewer interloop hydrogen bonds than loop-DNA hydrogen bonds or intraloop hydrogen bonds in the structures. The frequencies of the 30 most common interloop hydrogen bonds are in the range of 1%–12%. For the 2000 lowest-energy GBMV structures, 43.5% of the structures have no interloop hydrogen bonds and 90.3% have two or fewer. For the 50 lowest-energy structures,
Overall, the modeling results indicate that the gap loops, as well as the positively charged side chains of the back faces of donors in a disproportionate fraction (7/30) of these bonds. As is the case with the loop-DNA hydrogen bonds, the donors of the interloop bonds tend to be the side-chain end groups of the charged residues (26 out of 30 most common bonds in the 2000 lowest-energy GBMV structures), but the acceptors are disproportionately nonbasic residues (21/30). Many of the acceptor atoms tend to be backbone carbonyl oxygens (19/30), but the carboxamide oxygen of the N169 residues in the two loops also contribute (10/30); these residues also provide hydrogen-bond donors in a disproportionate fraction (7/30) of these bonds.

Overall, the modeling results indicate that the gap loops, as well as the positively charged side chains of the back faces of the dimer, have an important role in the binding of DNA. In addition, the heterogeneity of loop conformations with similar calculated energies of binding illustrates the flexibility of the loops, which may be important in the mechanisms by which UL44 binds to and diffuses along DNA (see Discussion).

**Table 1. Hydrogen-Bonding Comparisons**

<table>
<thead>
<tr>
<th>Donor</th>
<th>Acceptor</th>
<th>Number of Hydrogen Bonds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loop 1 DNA</td>
<td>Loop 1 DNA</td>
<td>4 3 6 6 6 4 4 3 2</td>
</tr>
<tr>
<td>Loop 2 DNA</td>
<td>Loop 1 DNA</td>
<td>5 4 6 5 7 6 5 7 5</td>
</tr>
<tr>
<td>DNA Loop 1</td>
<td>Loop 1 DNA</td>
<td>1 0 1 2 1 0 0 0 0</td>
</tr>
<tr>
<td>DNA Loop 2</td>
<td>Loop 1 DNA</td>
<td>0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>Loop 2 Loop 1</td>
<td>Loop 1 DNA</td>
<td>1 0 1 0 0 0 0 1 1</td>
</tr>
<tr>
<td>Loop 2 Loop 1</td>
<td>Loop 1 DNA</td>
<td>4 0 0 2 0 1 0 3 1</td>
</tr>
<tr>
<td>Loop 1 Loop 1</td>
<td>Loop 1 DNA</td>
<td>3 7 5 4 3 7 4 3 6</td>
</tr>
<tr>
<td>Loop 2 Loop 2</td>
<td>Loop 1 DNA</td>
<td>4 6 4 4 2 7 7 2 5</td>
</tr>
<tr>
<td>Loop 1 Protein</td>
<td>Loop 1 DNA</td>
<td>0 0 0 0 0 0 0 1 2</td>
</tr>
<tr>
<td>Loop 2 Protein</td>
<td>Loop 1 DNA</td>
<td>1 0 1 1 0 0 1 1 0</td>
</tr>
<tr>
<td>Protein Loop 1</td>
<td>Loop 1 DNA</td>
<td>0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>Protein Loop 2</td>
<td>Loop 1 DNA</td>
<td>0 0 0 0 0 0 0 0 0</td>
</tr>
</tbody>
</table>

Hydrogen-bonding comparisons for the ten lowest-energy structures in the GBMV and rdie calculations. The table gives the number of hydrogen bonds having various donor/acceptor combinations in each structure. "Protein" refers to the parts of the protein not including the gap loops. Results for protein-DNA, protein-protein, and DNA-DNA hydrogen bonds are not shown, because they are very similar for the first two types and the same for the last (the internal structure of the DNA was fixed).

22% have none and 88% have two or fewer. Hence, there is a tendency for the lower-energy structures to have slightly more (although still few) interloop hydrogen bonds. As is the case with the loop-DNA hydrogen bonds, the donors of the interloop bonds tend to be the side-chain end groups of the charged residues (26 out of 30 most common bonds in the 2000 lowest-energy GBMV structures), but the acceptors are disproportionately nonbasic residues (21/30). Many of the acceptor atoms tend to be backbone carbonyl oxygens (19/30), but the carboxamide oxygen of the N169 residues in the two loops also contribute (10/30); these residues also provide hydrogen-bond donors in a disproportionate fraction (7/30) of these bonds.

The apparent K<sub>d</sub> of the wild-type and single-mutant UL44 proteins were all approximately 1 nM, similar to that previously measured for wild-type UL44ΔC290 (Appleton et al., 2004; Loregian et al., 2007). (The K<sub>d</sub>s are apparent, rather than absolute, because for any length of DNA longer than the size of the binding site for UL44, there are multiple potential binding sites.)

As a control for nonspecific effects of the mutations, we used ultracentrifugation to measure the affinities of the UL44 monomer involved in DNA binding, consistent with the model. Possible reasons for why no single substitution mutant exhibited reduced affinity for DNA are explored in the Discussion.

Both UL44 monomers have a 12-residue flexible loop positioned at the intermonomer gap. Each of these gap loops contains five basic residues, which in the UL44/DNA model make contacts with DNA (Figures 1–4). To determine the importance of the gap loops in DNA binding, we substituted each of the five basic residues in this loop with alanines. Each single-alanine substitution mutant exhibited reduced DNA binding affinity with an apparent K<sub>d</sub> > 100 nM, demonstrating that the cavity formed by the basic back faces of each UL44 monomer is involved in DNA binding, consistent with the model.

**Effects of Back Face and Gap Loop Substitutions on UL44 Binding to DNA**

To test the prediction from the calculated UL44/DNA structure that UL44 contacts DNA via both the basic back face (central cavity) and the gap loops (see Figure 1), we made a series of central cavity and gap loop UL44 mutants and measured the affinity of each mutant for DNA. The UL44 termed wild-type in this study, which is the protein that was used for both the crystal structure (Appleton et al., 2004, 2006) and the modeling performed here, lacks the 143 C-terminal residues. Nevertheless, it displays all known biochemical activities of full-length UL44 in vitro (Loregian et al., 2004a, 2004b) as well as activity in origin-dependent DNA replication assays in transiently transfected cells (A. Loregian, G. Pari, and D.M.C., unpublished data). We first substituted four lysines on the back face of UL44 (K35, K158, K224, and K237; see Figure 1) with alanines, one at a time. Two of these lysines (K35 and K224) are predicted to interact strongly with DNA in the calculated lowest-energy structure; the other two (K158 and K237, which in the model are ~9–14 Å from the DNA in either monomer) may interact weakly. None of these single mutants had a measurable difference in affinity for DNA when compared to the wild-type protein in filter binding assays using a 30 bp DNA (Figure 5A). The apparent K<sub>d</sub>s of the wild-type and single-mutant UL44 proteins were all approximately 1 nM, similar to that previously measured for wild-type UL44ΔC290 (Appleton et al., 2004; Loregian et al., 2007). (The K<sub>d</sub>s are apparent, rather than absolute, because for any length of DNA longer than the size of the binding site for UL44, there are multiple potential binding sites.) However, the UL44 mutant in which all four lysines were substituted with alanines exhibited a dramatic decrease in DNA binding affinity with an apparent K<sub>d</sub> > 100 nM, demonstrating that the cavity formed by the basic back faces of each UL44 monomer is involved in DNA binding, consistent with the model.

As a control for nonspecific effects of the mutations, we used isothermal titration calorimetry to measure the affinities of the UL44 mutations that drastically reduce DNA binding do not affect binding to the C terminus of UL54.

As a control for nonspecific effects of the mutations, we used isothermal titration calorimetry to measure the affinities of the
wild-type and the two UL44 mutants with multiple substitutions (K35A/K158A/K224A/K237A and R165A/K167A/R168A/K171A/K172A) for a peptide corresponding to the C-terminal 22 residues of UL54. These two UL44 mutants had exhibited the greatest reduction in affinity for DNA (Figure 5). However, their affinities for the UL54 peptide were similar to that of the wild-type protein, as measured here (Figure 6) and previously (Appleton et al., 2004; Loregian et al., 2004a, 2004b), with $K_d$s within 2-fold of each other. In addition, the mutant proteins behaved as expected for properly folded proteins during purification (data not shown). These results suggest that the alanine substitutions do not globally disrupt folding of UL44, but instead specifically affect DNA binding.

Crosslinking of UL44 to Thiol-Tethered DNA

Disulfide crosslinking has been successfully used to covalently trap protein/DNA complexes (e.g., Banerjee et al., 2005, 2006). We used this strategy to investigate sites on UL44 that contact DNA. We synthesized an 18 bp oligonucleotide with a single thiol tether (Figure 7A) attached to the backbone phosphate at position 11. This thiol can crosslink with a cysteine residue. There are seven cysteines in UL44. In the crystal structure of UL44, two of these (C175 and C117) are exposed and could potentially be crosslinked to the thiol-tethered DNA. Based on our model of UL44-DNA interactions (Figure 1), C175, which is at the end of a gap loop, is ~4 Å from a phosphorus in monomer 1 and ~5 Å in monomer 2, within range of the ~4.1 Å thiol tether assuming modest flexibility. In contrast, C117 would be distant from the DNA binding site (>20 Å; see Figure 1). Nonreducing PAGE analysis of the crosslinking reaction showed that wild-type UL44 was able to crosslink thiol-tethered DNA, as seen from the appearance of a new higher molecular weight band corresponding to the UL44/DNA complex when the C175S and C175S/C117S proteins were used in the crosslinking reaction, and was diminished when the K224C protein was used. In contrast, the C175 protein was able to crosslink thiol-tethered DNA as efficiently as wild-type UL44 (Figure 7B). To test whether the effects of mutations on crosslinking were due to reduced affinity for DNA, we performed filter binding assays. The affinity for DNA of each of the mutant UL44 proteins was similar to that of the wild-type protein (Figure 7C). Taken together, the crosslinking results show that C175 is necessary for making a disulfide crosslink with the thiol-tethered DNA. In light of the modeling results, these findings strongly suggest that, by virtue of its spatial position relative to the DNA next to the C-terminal end of the gap loop, C175 is the site of crosslinking.

DISCUSSION

The structural fold of each monomer of UL44 is shared with those of HSV UL42 and PCNA (Appleton et al., 2004, 2006), but the quaternary structures of these processivity factors differ. UL42 is a monomer and PCNA is a trimeric ring, whereas UL44 is intermediate in structure, forming a C clamp-shaped dimer (Appleton et al., 2004, 2006; Krishna et al., 1994; Zuccola et al., 2000). In this report, we show through computational and experimental approaches that, similarly, UL44 binds to DNA in a manner that is a hybrid between the mechanisms used by HSV UL42 and PCNA. HSV UL42 binds DNA directly as a monomer via a basic back face (Komazin-Meredith et al., 2008; Randell and Coen, 2004; Randell et al., 2005), whereas UL44 is intermediate in structure, forming a C clamp-shaped dimer (Appleton et al., 2004, 2006; Krishna et al., 1994; Zuccola et al., 2000). In this report, we show through computational and experimental approaches that, similarly, UL44 binds to DNA in a manner that is a hybrid between the mechanisms used by HSV UL42 and PCNA. HSV UL42 binds DNA directly as a monomer via a basic back face (Komazin-Meredith et al., 2008; Randell and Coen, 2004; Randell et al., 2005), whereas PCNA has evolved a complex mechanism for binding (i.e., “loading onto”) DNA involving opening and reclosing its trimeric ring with the help of the clamp loaders so that it encircles DNA (Jeruzalmi et al., 2002). Dimeric UL44 interacts with DNA using at least two distinct protein...
regions, the back face, which is similar to the surface used by UL42, and two 12-residue flexible loops, which are located on either side of the intermonomer gap. These loops are shown in the modeling results to bridge the gap, such that DNA is encircled by the UL44 dimer in a manner akin to PCNA. The results of crosslinking and mutational studies are consistent with the model.

**Mechanism of DNA Binding by UL44**

UL44 binds double-stranded DNA more tightly than single-stranded DNA and without sequence preferences (Loregian et al., 2007). These findings, footprinting results from the HSV system (Gottlieb and Challberg, 1994; Randell and Coen, 2001), and the current modeling results together suggest that during DNA replication, UL44 binds DNA on the duplex side of the replication fork. The current modeling results also indicate that the DNA binds preferentially in the central channel of the protein dimer, and that the DNA is tilted slightly, similar to what is observed in the recently determined crystal structure of the *E. coli* sliding clamp (Georgescu et al., 2008). In that structure, which has a somewhat larger central cavity (approximately 30 Å in the smallest diameter, backbone to backbone, compared to ~25 Å for UL44), the DNA is tilted by about 22° from the C2 symmetry axis of the protein, compared to the tilt of about 10° in the current UL44/DNA model.

In the model, the presence of numerous hydrogen bonds and salt bridges between basic residues on the protein and the DNA phosphate groups, and the paucity of such interactions between the protein and the nucleotide bases, is consistent with UL44 binding to DNA without sequence preferences (Loregian et al., 2007). Such a pattern has been observed in other nonspecific DNA-protein interactions (Kalodimos et al., 2004; Masse et al., 2002). Moreover, the spiral track for the DNA backbone formed by the charged residues of the protein is reminiscent of the spiral track formed by the replication factor C (RFC) portion of the RFC/PCNA complex that has been suggested to mediate RFC binding to DNA (Bowman et al., 2004). The flexibility of the gap loops, which is indicated both by their disorder in the crystal structures in the absence of DNA and by the variety of low-energy conformations obtained for the loops in the current modeling studies in the presence of DNA, suggests a possible DNA binding mechanism that involves motion of the loops. Preliminary molecular dynamics simulations of the UL44 dimer in implicit water environments demonstrate some in-plane global motions of the dimer that tend to widen or narrow the intermonomer gap (unpublished data), which could have a role in the binding mechanism. By contrast, different crystal structures of the protein have been found to be related by an out-of-plane “lock washer” type of displacement (Appleton et al., 2004, 2006). Molecular dynamics simulations of PCNA suggest a DNA binding mechanism that involves out-of-plane distortions of the clamp ring (Kazmirski et al., 2005). Further studies will be needed to determine whether or not UL44 employs elements of both types of clamp-opening mechanisms.

---

**Figure 6. Multiple-Alanine Substitution Mutants Retain Binding to the C-Terminal 22 Residues of UL54**

(A) Raw isothermal titration calorimetry data for the titration of UL54 peptide into a sample cell containing wild-type or the indicated mutant UL44ΔC290 proteins.

(B) Amount of heat released per injection in kcal/mol of injectant plotted against the molar ratio of peptide to protein that is present in the cell at the time of each injection. The $K_d$ calculated from the data are shown below the plots.
Hybrid Interaction of Processivity Factor with DNA

Mechanisms of UL44 Diffusion

Processivity factors not only tether catalytic subunits of DNA polymerases to DNA, but they also diffuse with the enzyme along the DNA without impeding translocation. Indeed, UL44 can diffuse on DNA (G.K.-M., R. Mirchev, D. Golan, A. van Oijen, and D.M.C., unpublished data). The current results suggest a possible mechanism by which UL44 achieves strong binding to DNA while still allowing for diffusion. At least two factors are likely to play a role. First, the variability of the loop conformations found in the calculations suggests that they are able to move to help generate a low free-energy pathway for the processivity factor as it moves along the DNA molecule. Second, the results suggest the possibility that the protein moves via a spiral “ionic track” mechanism, in which negatively charged phosphate groups of the DNA are passed from one positively charged flexible protein side chain to the next. The tendency of charged protein residues to interact at numerous points along both of the phosphate backbone strands and the shape of the curved half-tube structural distributions formed by the cluster centers of the flexible loops of UL44 are supportive of the spiral ionic-track mechanism. An ionic-track mechanism has been suggested for the rotation of the γ subunit in F1 ATPase (Ma et al., 2002). The spiral geometry of the ionic track involving UL44’s central cavity and the gap loops could help to explain the mechanism of diffusion: consecutive phosphates along the backbone of the DNA are only ~6.7 Å apart. Because the binding is nonspecific, the processivity factor must stabilize a path of roughly this length for the DNA between equivalent protein/DNA positions. By contrast, phosphates on either side of the minor groove (along the same longitudinal axis) are 12–13 Å apart, and on either side of the major groove are approximately 18–20 Å apart. Hence, pure translational diffusion of the processivity factor along the DNA (i.e., in a direction normal to the “C” plane of the dimer, without rotation, which has been termed “sliding along one face of the DNA” or “2D sliding”; Kampmann, 2004; Viadiu and Aggarwal, 2000), would require stabilization of a path of over 30 Å in length between equivalent protein/DNA positions (the distance across the major and minor grooves, which is 33 Å in canonical B-DNA).

Thus, the reported experimental and modeling results suggest that UL44 can remain associated with DNA for prolonged periods, even in the absence of the catalytic subunit, as is found for sliding clamps (Stukenberg et al., 1991), and that UL44 diffuses on DNA by sliding along a helical path (Berg et al., 1981). These predictions can be tested experimentally (Blainey et al., 2006; Stukenberg et al., 1991).

EXPERIMENTAL PROCEDURES

Modeling Calculations

All calculations were performed with the CHARMM program (Brooks et al., 1983). The initial coordinates for the UL44 structure were taken from the dimer form of X-ray crystal structure PDB ID code 1T6L (Appleton et al., 2004).
Hydrogens were built onto the structure with the HBUILD facility (Brunger and Karplus, 1988) according to the polar hydrogen parameter set with the EEF1 solvation model (Lazaridis and Karplus, 1999) (see below). In the 1T6L structure, there are three regions of disorder in each monomer: the N and C termini, namely residues 1–9 and 271–290, and the gap loop, residues 163–174. The terminal sequences, which are in solvent-exposed regions of the molecule and which were found not to interact significantly with the central cavity, were not included. The DNA moiety used crystal structure PDB ID code 102D (sequence 5′-CGAATTTGGCC-3′). This is a 12 base pair double-stranded DNA oligonucleotide in the B form with a small ligand (propamidine), which was removed.

The all-hydrogen model (PARAM27 parameter set) in the CHARMM program was used for the DNA, and the hydrogens were built according to known internal geometries.

The docking calculations consisted of several phases. In a preliminary phase, a “first-guess” structure was calculated for the flexible gap loops in the absence of DNA, using molecular dynamics simulations. This conformation is compact and close to the protein. Then in the first main phase, the DNA moiety was docked to the UL44 dimer with flexible side chains and with the gap loops fixed in the first-guess conformation; the docking was carried out with a grid-based approach, similar to that used in protein-protein docking studies (Goodsell et al., 1996; Norel et al., 2001; Pierce et al., 2005), and tertiary structure-prediction studies (Petrella and Karplus, 2000). Side-chain relaxation was incorporated as in Xiang and Honig (2001), using rotamer libraries (Lovell et al., 2000). In the second main phase, extensive conformational searching for the flexible gap loops was carried out in the presence of the DNA through grid-based searching, along with the required BYCC list-builder option in the CHARMM program (R.J.P. and M.K., unpublished), which facilitates carried out with a newly developed general search facility called the Z module for the two main phases and the final phase, conformational searching was carried out with a distance-dependent dielectric function or with the GBMV solvent model. For the side-chain insertion analysis, a side chain was said to have inserted if its calculated using a probe radius of 1.6 Å (results using a 1.4 Å radius were similar). Nonpolar portions of the molecules were taken to be nonpolar carbons and their associated hydrogens. For details, see the Supplemental Data. Visual inspection and PROCHECK (Laskowski et al., 1993) analyses were used to verify a number of the lowest-energy structures (all had minimal PROCHECK violations). Coordinates of the lowest-energy GBMV structure are provided in PDB format in the Supplemental Data.

### Plasmids

The pd15-UL44ΔC290 plasmid was described previously (Loregian et al., 2004a). It expresses the N-terminal 290 residues of UL44 with a glutathione S-transferase (GST) protein at the N terminus and a PreScission Protease cut site in between GST and UL44. This plasmid was used as a template for construction of all UL44 mutants, which were obtained using the QuikChange mutagenesis kit (Stratagene). See Table S4 for primer sequences used to create these mutants. All plasmids were sequenced to confirm the presence of the mutation(s) and the absence of undesired mutations.

### Protein Purification

Wild-type and mutant UL44ΔC290 proteins were expressed and purified as described previously, except the SP-Sepharose column was omitted (Appleton et al., 2004). Briefly, all UL44ΔC290 proteins were expressed at 16°C in E. coli. They were purified by using glutathione Sepharose resin (GE Healthcare), followed by cleavage with PreScission protease (GE Healthcare) to remove the GST tag. They were then chromatographed on a heparin HiTrap column (GE Healthcare) and finally on a Superdex200 HR column (GE Healthcare).

### Filter Binding Assays

Filter binding assays were performed as previously described (Appleton et al., 2004). Briefly, 1 fmol of radiolabeled 30 bp DNA was incubated with increasing concentrations of either wild-type or mutant UL44ΔC290. Protein-bound DNA was then separated from free by passing samples through nitrocellulose and DE81 filters. Filters were washed and dried and radioactivity was measured by liquid scintillation counting. Apparent $K_d$ values were calculated using a saturation isotherm analysis from the concentrations of UL44 that resulted in half-maximal binding.

### Isothermal Titration Calorimetry

The UL44-DNA crosslinking

The oligonucleotide (5′-TACCGAAGCCATCAGAGT-3′) with a single thiol tether attached to the backbone phosphate between bases 11 and 12 was synthesized using the method previously described (Banerjee et al., 2006). The complementary oligonucleotide was purchased, PAGE purified, from Integrated DNA Technologies. The double-stranded DNA was formed by mixing the two complementary oligonucleotides 1:1 in a buffer containing 25 mM NaCl and 15 mM Tris-HCl (pH 7.5), heating the mixture to 85°C, and then cooling it slowly to room temperature. Crosslinking reactions were performed by mixing either wild-type or mutant UL44ΔC290 proteins (1 μM) with thiol-tethered DNA (2 μM) in 15 μl of reaction buffer (30 mM Tris-HCl [pH 7.5], 30 mM NaCl, and 10 μM DTT) for 1 hr at room temperature. The reactions were stopped by capping the free thiol groups with S-methylmethionine isothiouate (40 mM), SDS loading buffer was added to the quenched reaction mixtures, and the samples were analyzed on a 10% SDS–PAGE gel under nonreducing conditions. The gel was stained with SimplyBlue Safe stain (Invitrogen).

### Supplemental Data

Supplemental Data include four tables, Supplemental Results, Supplemental References, and PDB format coordinates and are available online at http://www.structure.org/cgi/content/full/16/8/1214/DC1/.

### Acknowledgments

R.J.P. and M.K. performed and analyzed the modeling calculations with assistance from the protein backbone libraries; and the Bauer Center at Harvard for the use of its computational resources. This work was supported in part by the Epplcy Foundation (R.J.P.), and by grants from the National Institutes of Health to D.M. (AI19838), M.K. (GM30804), and G.L.V. (GM44853), and a Ruth L. Kirschstein award to W.L.S.
REFERENCES


herpes simplex virus UL42 support a role for DNA binding in processive DNA synthesis. J. Virol. 79, 12025–12034.


