# C-Terminal Domain of Integrase Binds between the Two Active Sites

Victoria A. Roberts*

San Diego Supercomputer Center, University of California, San Diego, La Jolla, California 92093, United States

ABSTRACT: HIV integrase (HIV-IN), one of three HIV enzymes, is a target for the treatment of AIDS, but the full biological assembly has been difficult to characterize, hampering inhibitor design. The recent crystallographic structures of integrase from prototype foamy virus (PFV-IN) with bound DNA were a breakthrough, revealing how viral DNA organizes two integrase dimers into a tetramer that has the two active sites appropriately spaced for insertion of the viral DNA into host DNA. The organization of domains within each PFV-IN protein chain, however, varies significantly from that found in HIV-IN structures. With the goal of identifying shared structural characteristics, the interactions among components of the PFV-IN and HIV-IN assemblies were investigated with the macromolecular docking program DOT. DOT performs an exhaustive, rigid-body search between two macromolecules. Computational docking reproduced the crystallographic interactions of the PFV-IN catalytic and N-terminal domains with viral DNA and found similar viral DNA interactions for HIV-IN. Computational docking did not reproduce the crystallographic interactions of the PFV-IN C-terminal domain (CTD). Instead, two symmetry-related positions were found for the PFV-IN CTD that indicate formation of a CTD dimer between the two active sites. Our predicted CTD dimer is consistent with cross-linking studies showing interactions of the CTD with viral DNA that appear to be blocked in the PFV-IN structures. The CTD dimer can insert two arginine-rich loops between the two bound vDNA molecules and the host DNA, a region that is unoccupied in the PFV-IN crystallographic structures. The positive potential from these two loops would alleviate the large negative potential created by the close proximity of two viral vDNA ends, helping to bring together the two active sites and assisting host DNA binding. This study demonstrates the ability of computational docking to evaluate complex crystallographic assemblies, identify interactions that are influenced by the crystal environment, and provide plausible alternatives.
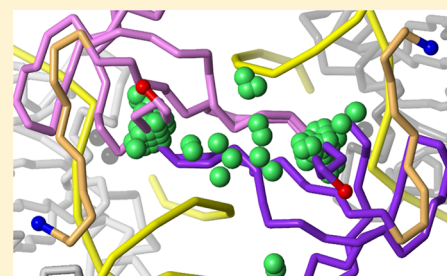
INTRODUCTION

Determining the biological assembly of the retroviral enzyme integrase (IN) from the human immunodeficiency virus (HIV-IN) is of great interest because HIV-IN, one of just three HIV enzymes, is a key target for the development of drugs to treat AIDS.[1−4] IN inserts a DNA copy of the retroviral RNA into host DNA by performing two reactions: 3′-processing of the viral DNA (vDNA) and strand transfer.[5] In the 3′-processing reaction, the two terminal nucleotides are cleaved from the two 3′-ends of the vDNA. The resulting IN/vDNA complex is transported into the nucleus, where the exposed 3′-hydroxyls are joined to opposing strands in the host DNA, producing the integration intermediate. Depending on the retroviral genus, the two vDNA insertion points are spaced four to six base pairs apart in the host DNA. Host enzymes then remove the unpaired nucleotides at the 5′-ends of the vDNA and repair the single-stranded region of the DNA to complete incorporation of the vDNA into the host genome.

The biologically active state of HIV-IN *in vivo* is complex and not fully understood. IN consists of three domains: the N-terminal domain (NTD), the core catalytic domain (CCD), and the C-terminal domain (CTD), with connecting linkers L1 and L2. Binding of vDNA is required for the organization of the catalytically active enzyme, which is believed to be a tetramer.[6,7] Crystallographic structures of the HIV CCD[8−14] and of two-domain HIV-IN constructs (NTD−CCD[15,16] and CCD−CTD[17]) all show the same CCD dimer,[5] but there are no structures of full-length HIV-IN or of any portion of HIV-IN with bound DNA. The related avian sarcoma virus IN (ASV-IN)[18,19] (37% sequence similarity to HIV-IN) has been crystallized with DNA, but only the two-domain CCD−CTD region was seen. Models of the CCD and CTD with bound vDNA were proposed based on computational docking,[20−22] the HIV-IN CCD−CTD structure,[17] and cross-linking data,[23] but these did not address a key question: what interactions determine the spacing between the two vDNA insertion points? The two vDNA insertion points are much closer together than the two active sites of a single CCD dimer, so the biological assembly must require at least two dimers. Models consisting of two or more HIV-IN dimers were proposed,[23−29] but they varied greatly in the relative geometry of the two dimers and the positions of the NTD and CTD.

The recent crystallographic structures of full-length IN from prototype foamy virus (PFV-IN) with bound 19 bp vDNA fragments[30] and host DNA[31] were a major breakthrough, revealing a tetramer of two dimers that each contribute one active site. The active sites of the two inner vDNA-binding subunits (Figure 1A) are correctly spaced to insert the two vDNA ends four base pairs apart in the host DNA. This spacing appears to be determined by the CTD and the L1 and L2 linker regions. Both the 3′-processing and strand-transfer reactions
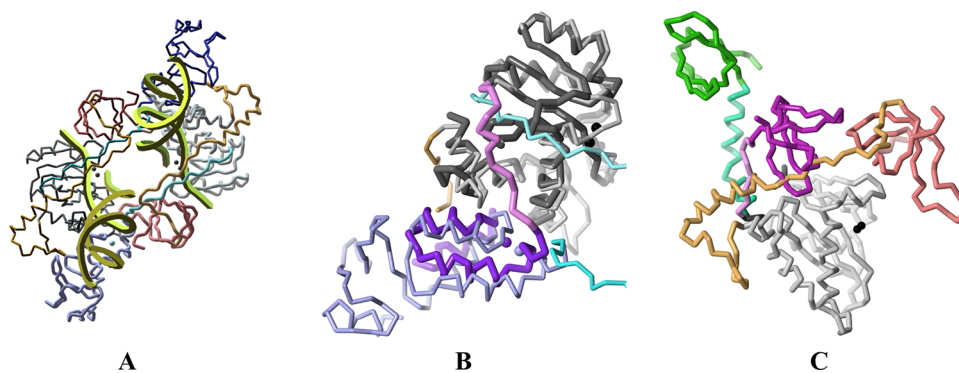
Figure 1. Variations in integrase structures. (A) The two inner subunits of the PFV-IN assembly (PDB ID 3OY9),[30] looking down on the vDNA-binding face. Host DNA binds on the opposite face. The two protein chains (upper with thin tubes, lower with thicker tubes and paler colors) are colored as follows: blue, NTD; light blue, L1; gray, CCD; orange, L2; and red, CTD. The L1 and L2 linker regions extend over the region between the two active sites and contact the vDNA (yellow ribbons). The active-site Mn ions (black spheres) and NTD Zn ion (light blue spheres) are also shown. (B) The L1 linkers in HIV-IN and PFV-IN have very different positions. In the HIV NTD−CCD structure 3F9K,[16] L1 (pink) lies in a groove on the surface of the CCD (dark gray) and connects to the adjacent NTD (purple), creating a compact structure. In PFV-IN (thin tubes), the CCD (gray) and its contacting NTD (lavender) are attached to different protein chains and L1 (light blue) extends through solvent (as shown in (A)) so that the NTD and CCD in the same protein chain do not contact each other. A second HIV NTD−CCD structure[15] (PDB ID 1K6Y, not shown) has no density for L1, and the NTD was assigned to a very different position than that in 3F9K. Subsequent analysis[5] revealed that the crystal lattice includes an NTD that matches the NTD position in 3F9K. (C) Differences in the CTD position among integrases. Superposition of the CCDs (gray, shown for PFV-IN only) shows the dramatic variation in L2 conformation and CTD position among PFV-IN (orange and red), HIV-IN (light green and green, PDB ID 1EX4),[17] and ASV-IN (pink and magenta, PDB ID 1C0M).[18] Further variation is found in the second subunit of the HIV-IN and ASV-IN CCD−CTD structures and in SIV-IN,[33] which has a helical L2 of the same length as that in HIV-IN.

occur within the crystal, and the structures of ground states have been trapped along the reaction pathway.[32]

The PFV-IN crystallographic structures, however, may not fully represent the biologically active assembly in solution. About one-quarter of the total protein is missing in the crystal; only the CCD is observed for the two outer subunits of the PFV-IN tetramer. The long linker regions of PFV-IN could be influenced by crystal packing, altering the arrangement of the domains. Crystal packing may also influence the angle between the two vDNA molecules. This angle is about 90° in the crystal, but it is larger in solution according to small-angle X-ray scattering (SAXS).[34] An angle greater than 90° is also supported by atomic force microscopy on the HIV-IN/vDNA complex.[35]

It is not clear how far the PFV-IN structures can be extended to HIV-IN and ASV-IN. PFV-IN is about 100 residues longer and has low sequence homology with the NTD, CTD, and linker regions of HIV-IN and ASV-IN. The crystallographic positions of the PFV L1 and L2 linkers lie between the CTD and the bound vDNA, blocking CTD/vDNA interactions found by cross-linking studies on HIV-IN and ASV-IN.[36] The PFV-IN[30] and HIV-IN[5,15,16] structures show similar contacts between the CCD and the NTD, but these two domains are connected by the L1 linker in very different ways (Figure 1B). The position of the CTD relative to the CCD[5,19] (Figure 1C) and the contacts of the CTD[17,18,30] vary greatly among the crystallographic structures of PFV, HIV, and ASV INs. Furthermore, the L2 linker, which connects the CCD and CTD, varies greatly in conformation and in length: 50 residues in PFV IN, 21 residues in HIV IN, and just 8 residues in ASV IN. The L2 spans about 48 Å in the PFV-IN structures.[30] It has been proposed that the L2 of HIV-IN can span this distance by adopting a fully extended conformation,[37] but L2 has a helical conformation in the structures of the HIV CCD−CTD[17] and the closely related simian immunodeficiency virus (SIV) CCD−CTD.[33] No conformation of the short eight-residue ASV-IN L2 can span this distance.[19] Therefore, if all three retroviral integrases share a similar role and position for the CTD in the vDNA-bound assembly, then it cannot be the same as that observed in the PFV-IN structures.

To attempt to resolve the apparent contradictions between the PFV-IN structures and structural and experimental data on HIV-IN, computational docking was applied to PFV-IN and HIV-IN. The goal was to identify interactions that are shared by PFV-IN and HIV-IN. Docking should reproduce the interactions among the components of the PFV-IN assembly, since each component has undergone the induced shape fit needed for formation of the complex. Instead, a new position for the PFV CTD in the presence of bound vDNA was found that is consistent with the length and conformation of the HIV-IN L2 linker region. We propose that the CTD forms a dimer between the two vDNA-bound active sites that is the main determinant of their spacing. Each CTD inserts an arginine-rich loop into the space between the two active sites, a region that is unoccupied in the PFV-IN structures. These loops alleviate the large negative charge created by the two vDNA ends and facilitate the binding of host DNA needed for strand transfer.

■ RESULTS

Interactions within the PFV-IN and HIV-IN assemblies were investigated using the macromolecular docking program DOT.[38−41] DOT performs a global, systematic search between two molecules by translating and rotating one component (the moving molecule) around a second component or complex (the stationary molecule), resulting in 100−900 billion configurations. Each configuration is ranked by the sum of intermolecular electrostatic and van der Waals energies. A key advantage of DOT for studying protein−DNA complexes is the calculation of the electrostatic potential of the stationary molecule by Poisson−Boltzmann methods, which take into account dielectric, solvation, and ionic strength effects. The resulting DOT electrostatic term provides a good estimate of the electrostatic energy of the highly polar protein−DNA interaction.[40]

Table 1. Analysis of Docked Components from the PFV-IN Structure 3OY9

run	stationary molecule or complex	moving molecule	hits[a] in top 30		hits[a] in top 2000		best ranked hit (RMSD)
			no.	RMSD (Å)	no.	RMSD (Å)	
	Comparison with the Crystallographic Position						
1	CCD	vDNA-9	22	1.0 to 5.0	248	0.9 to 5.0	1 (3.8)
2	NTD	vDNA-9	0		0	Best 23.6	
3	CTD	vDNA-9	0		0	Best 13.0	
4	CCD + vDNA	NTD	24	0.9 to 3.8	155	0.9 to 5.0	1 (1.3)
5	NTD + CCD	vDNA-9	25	0.7 to 4.0	218	0.7 to 5.0	1 (0.9)
6	CCD + vDNA	CTD	0		0	Best 17.1	
7	NTD + CCD + vDNA	CTD	0		0	Best 16.9	
8	(CCD + vDNA)$_2$	CTD	0		0	Best 12.5	
9	(NTD + CCD + vDNA)$_2$	CTD	0		6	0.9 to 1.7	399 (0.9)
10	(CCD + vDNA)$_2$	NTD	27	1.1 to 3.3	150	0.9 to 5.0	1 (2.1)
11	(CCD + CTD + vDNA)$_2$	NTD	30	1.1 to 3.9	243	1.0 to 5.0	1 (2.1)
12	(CCD + CTDdock + vDNA)$_2$	NTD	30	1.0 to 3.6	365	0.9 to 5.0	1 (2.1)
	Comparison with the Docked CTD Position[b]						
6D[c]	CCD + vDNA	CTD	0		0	Best 6.8	
7D[c]	NTD + CCD + vDNA	CTD	0		0	Best 15.8	
8D[c]	(CCD + vDNA)$_2$	CTD	29	0.0 to 4.3	464	0.0 to 5.0	1 (1.4)
9D[c]	(NTD + CCD + vDNA)$_2$	CTD	28	0.0 to 4.3	532	0.0 to 5.0	1 (1.4)

[a]Placements with an RMSD ≤ 5 Å between the moving molecule and the target position. [b]Symmetry-related CTD docked positions (ranks 2 and 12) from docking run 8, (CCD + vDNA)$_2$ with CTD. [c]Same number run as above, but compared with the docked CTD position.

Two inner subunit protein chains of PFV-IN, which provide all of the observed vDNA and host DNA contacts, and the two bound vDNA molecules were generated from PDB ID 3OY9.[30] The individual PFV-IN domains were extracted from these coordinates. Linker regions were not included because of their high variability in both length and sequence among integrases from different retroviruses. Although each domain has undergone the induced fit needed for complex formation, a single domain may not have a sufficiently complete interaction surface to give a correctly docked complex.[41] As more components are added to complete the binding surfaces, the fit of the docked complex to the crystallographic structure should improve. First, individual PFV-IN domains were searched for distinct DNA-binding sites, and then larger assemblies were built based on the most well-defined interactions.

Calculations were then extended HIV-IN, focusing on the interaction of the HIV NTD and CCD (taken from PDB ID 3F9K) with vDNA. First, the active-site region of the HIV CCD was remodeled based on the PFV CCD to account for structural changes induced by vDNA binding. The interaction of the HIV NTD and CCD with vDNA was examined in the absence and presence of the L1 linker from 3F9K, which is the only HIV-IN structure that shows density for L1.

In all dockings, the 2000 top-ranked placements of the moving molecule were retained. The 30 top-ranked placements were analyzed in detail for clustering and contacts with the stationary molecule. Distinctive clusters among the top 30 complexes were also observed in the top 2000, with 150 or more placements within each cluster. For DNA dockings, the distribution of the molecular centers of the top 2000 placements was also examined, which typically traces out the axis of the bound DNA.[42]

Docking DNA to the Individual PFV-IN Domains. DNA binding to the three individual PFV-IN domains was investigated. Selection of the coordinates for the NTD, CCD, and CTD were based on previous sequence and structure alignment,[30] except that eight residues of linker L2 were added to the C-terminus of the CCD and 11 residues of linker L2 were added to the N-terminus of the CTD (see Methods). The vDNA dockings used an active-site fragment (vDNA-9) consisting of seven nucleotides at the 3′-end of the cleaved strand and nine nucleotides of the uncleaved strand. This vDNA fragment includes all of the CCD and CTD contacts and most of the NTD contacts found in the crystal structure. Since DOT searches over all space, both the vDNA and host DNA binding sites could potentially be identified among the top-ranked dockings. The shape of the bound vDNA might favor the vDNA-binding site. To eliminate bias, a B-form DNA fragment (B-DNA) that had the vDNA sequence, including the two cleaved nucleotides, was also docked.

Computational docking identified the observed vDNA-binding site on the isolated CCD as the most favorable DNA-binding surface. In the docking of vDNA-9 to the CCD, the 30 top-ranked vDNA-9 placements aligned with the crystallographic vDNA. Twenty-eight correctly positioned the DNA minor groove over CCD helix α4 (residues 217−235), and two put the DNA major groove over helix α4. The 22 vDNA-9 placements with RMSDs of 5.0 Å or less (Table 1, run 1) and two more (RMSDs of 5.1 and 6.0 Å) positioned the 3′ hydroxyl within 4 Å of the active-site Mg ion. The top 2000 vDNA-9 placements predominantly clustered over the vDNA-binding site, but some overlapped the adjacent host-DNA-binding site. These were generally aligned with the bound vDNA rather than the bound host DNA.

The distribution of the top B-DNA placements followed that of vDNA-9, with the majority of the top 2000 and all of the top 30 over the vDNA-binding site. Twenty-eight of the 30 top-ranked B-DNA placements showed good alignment with the crystallographic vDNA, but all positioned the major groove, rather than the minor groove, over CCD helix α4. The other two favorable placements had the minor groove over helix α4, but they were rotated about 30° from the crystallographic vDNA. These results are consistent with our previous studies showing that rigid-body docking with B-DNA can identify DNA-binding sites on proteins and the orientation of DNA at
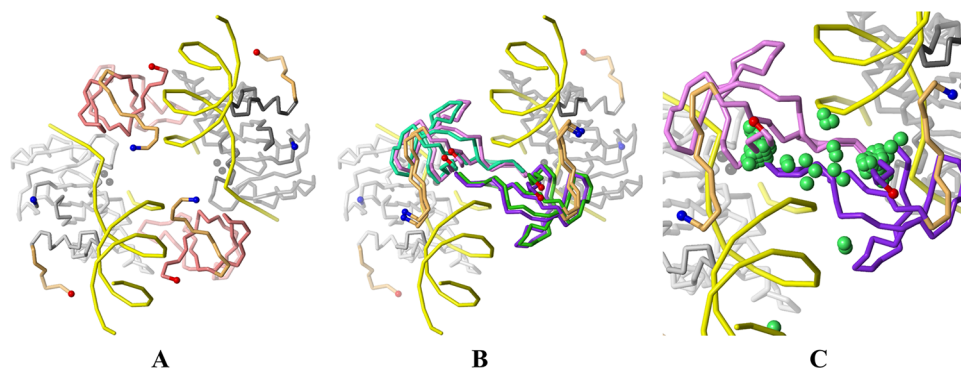
Figure 2. PFV CTD docks between the two active sites. (A) Each crystallographic CTD (red tubes) contacts both vDNA molecules (yellow tubes), but the active-site region between the two 3′-ends of the cleaved vDNA strands is unoccupied. The CCDs (gray with spheres representing the Mn ions) and the CTDs have attached L2 residues (orange) that were included in the docking calculations. N- and C-termini are shown as blue and red spheres. (B) Two symmetry-related pairs of docked CTD placements (ranks 2 and 12, green; ranks 4 and 15, purple) represent the two clusters comprising 29/30 of the top-ranked positions. The docked CTDs occupy the central active-site region between the two vDNA molecules, reducing the distance between the CCD C-terminus and the CTD N-terminus by 12 Å. (C) The 100 top-ranked docked CTD positions, represented by their geometric centers (green spheres), are concentrated in the region over the two active sites. Of these, 78 had an RMSD within 5 Å of the symmetry-related pair (purple).

these sites but that it is less successful at identifying the DNA contacts involved in the interaction.[42] Several of the top 30 B-DNA placements partially overlap the host-DNA-binding site near the active site, but the top 2000 did not show alignment with the bound host DNA.

In the PFV-IN structure, the double-stranded region of the vDNA lies over, and approximately parallel to, NTD helices $\alpha2$ and $\alpha3$.[30] Docking vDNA-9 or B-DNA to the isolated NTD did not distinguish this vDNA-binding mode (Table 1, run 2). Although the 30 top-ranked placements generally docked over the vDNA-binding surface, they were rotated so that NTD helices $\alpha2$ or $\alpha3$ were inserted into the DNA major groove.

In the PFV-IN structure, the CTD has two contacts with vDNA, one with the double-stranded region of one vDNA molecule and one with the two-nucleotide overhang of the uncleaved strand of the second vDNA molecule.[30] In addition, CTD residues Arg 329 and Arg 362 contact host DNA.[31] Computational docking of vDNA-9 and B-DNA neither found a distinct DNA orientation nor distinguished any of these contacts (Table 1, run 3). Instead, the CTD showed a diffuse DNA-binding surface created by its N- and C-terminal ends and the loop between $\beta1$ and $\beta2$ (residues 326−337). This surface contains all of the residues that contact the double-stranded region of the vDNA (residues Ser311 and Arg313 in L2, Arg326 and Arg336 in the loop, and Arg374 at the C-terminus) and one (Arg329) that contacts the host DNA. Thus, computational docking identified DNA-contacting surfaces of the NTD and CTD domains observed in the crystallographic structures of the full, tetrameric PFV-IN assembly, but it did not identify the orientation of the DNA over these surfaces.

Multicomponent PFV-IN Dockings. The combination of a CCD from one inner subunit, the contacting NTD from the other inner subunit, and a bound vDNA molecule creates a compact structure. This three-component interaction was probed starting with the complex of the CCD with vDNA, the interaction that was reliably reproduced by computational docking. The CCD was combined with full-length vDNA (17 bp plus the two-nucleotide overhang of the uncleaved strand) to make the CCD + vDNA complex. Docking the NTD to this complex identified the NTD crystallographic position, with 24 hits among the 30 top-ranked configurations (Table 1, run 4).

In the other six favorable configurations, NTD helices $\alpha2$ or $\alpha3$ were inserted into the DNA major groove, as found in the docking of vDNA-9 to the isolated NTD. Docking vDNA-9 to the combination of the NTD with the CCD (NTD + CCD) improved the clustering among the top 30 vDNA-9 placements (Table 1, run 5) compared to that with CCD alone. In all 30, the vDNA-9 was aligned with the crystallographic DNA axis, with the correct orientation over the NTD surface and the DNA minor groove over CCD helix $\alpha4$. In 28 vDNA-9 placements, the cleaved 3′ end was close to the active-site, with 25 within the 5 Å RMSD cutoff. Thus, in complex with the CCD, the crystallographic vDNA-binding mode of the NTD was clearly identified.

In contrast to the results for the NTD, docking the CTD to CCD + vDNA (Table 1, run 6) produced no match with the crystallographic interaction. Instead, the CTD varied widely in position and orientation relative to the CCD. The 30 top-ranked placements had little or no contact with the CCD, but they contacted the vDNA with the same surface (loop 326−337 and the N- and C-terminal regions) found in the CTD/vDNA-9 dockings. Adding the contacting NTD to make the three component complex NTD + CCD + vDNA did not improve CTD docking (Table 1, run 7). Thus, the CCD + vDNA construct was sufficient to identify and distinguish the binding mode of the NTD but not the CTD.

Dockings Involving the PFV-IN Catalytic Core. In the PFV-IN structure, each CTD molecule contacts two CCDs, two vDNA molecules, and one NTD (Figure 1A), suggesting that a larger assembly would be needed to define the CTD-binding site. The starting point was the dimeric catalytic core, $(CCD + vDNA)_2$, which contains the CCDs from the two inner subunits and their two bound, full-length vDNA molecules. The two CCD + vDNA subunits do not contact each other; the relative positions found in the crystallographic structure were used. Docking the CTD to $(CCD + vDNA)_2$ found no CTD placements near the two crystallographic positions (Table 1, run 8, and Figure 2). Instead, the 30 top-ranked solutions lay in the region over the two active sites (Figure 2B,C), with 29 in two symmetry-related positions. Comparison with ranks 2 and 12, representative of these two positions, revealed a large cluster for each symmetry-related

placement (Table 1, run 8D). Each docked CTD contacted both vDNA molecules using primarily the same vDNA-binding surface as that found in the other CTD dockings. The two symmetry-related CTDs are adjacent to each other, suggesting formation of a CTD/CTD dimer. Although CTD loop residues 326−335 overlap, this large flexible loop could adopt a different conformation that would eliminate steric conflicts.

The previous dockings of the CTD to CCD + vDNA and NTD + CCD + vDNA were analyzed for matches to the symmetry-related docked positions, but none were found (Table 1, runs 6D and 7D). Thus, both vDNA molecules are required to create the CTD-binding site.

In the PFV-IN crystallographic structure, each CTD has a small contact with the NTD of the same protein chain.[30] To test the importance of this interaction, the CTD was docked to (NTD + CCD + vDNA)₂, in which the two NTDs were added to the catalytic core. More than one-quarter of the 2000 top-ranked CTD placements matched the symmetry-related docked position (Table 1, run 9D). Only a few poorly ranked placements matched the crystallographic CTD (Table 1, run 9).

In contrast to the CTD, docking the NTD to the dimeric catalytic core identified the two crystallographic positions, with 27 hits in the top 30 (Table 1, run 10). The other three placements lay near the catalytic sites, overlapping the region in which the CTD had docked. When the two crystallographic CTDs were added to the catalytic core, all top 30 NTD placements matched the crystal (Table 1, run 11). The CTDs might improve docking by increasing the NTD-binding surface or by partially blocking the active-site region. To distinguish these two possibilities, the two symmetry-related, docked positions of the CTD (ranks 2 and 12) were added to the catalytic core to make (CCD + CTDdock + vDNA)₂. Neither CTD contacts the crystallographic NTD position. Docking the NTD to this construct provided the best results; all 30 top-ranked NTD placements and over 18% of the top 2000 matched the crystal (Table 1, run 12). Thus, the CTD improves NTD docking by blocking the active-site region rather than by providing additional NTD contacts.

Rebuilding CTD Loop 325−340. We explored alternate conformations for loop residues 326−337, which overlap in the two symmetry-related docked CTDs (Figure 2B, C). When the HIV CTD (from PDB ID 1EX4) and the ASV CTD (from PDB ID 1C0M) were superposed onto the docked PFV CTDs, the corresponding loops of the HIV and ASV CTDs extended between the two active sites. The PFV loops were rebuilt to follow the conformation of the HIV loops (see Methods). Each PFV loop contains four arginine residues (326, 329, 334, and 336), resulting in the insertion of eight positively charged side chains between the two active sites (Figure 3). Arg329 lies at the most extended part of the loop and could contact the host DNA. Other positively charged CTD residues (Arg350, Arg362, Lys339, and Lys345) may also contact the vDNA. The corresponding HIV CTD loop inserts residues Lys236, Arg228, and Arg231 between the two active sites (Figure 3), with Arg231 being capable of contacting the host DNA. HIV CTD Arg263, Lys258, and Lys264 may provide additional vDNA contacts.

HIV-IN Docking. Our predicted position for the CTD clashes with L1 and L2 linkers in the PFV-IN crystallographic structures. The docked CTD position moves the L2 away from the active site, eliminating steric conflicts. It is unclear if the L1 linkers, especially the shorter ones of HIV-IN and ASV-IN, are
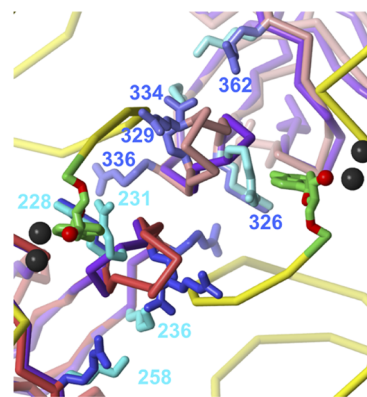

Figure 3. CTD loop 326−337 fills the region between the two PFV-IN active sites. The view is looking down on the host-DNA-binding face of the IN assembly. The loops of the two docked CTDs (red, lower left; light red, upper left) put eight Arg residues (dark blue side chains and labels) between the two 3′-nucleotides (green with red oxygen atoms) of the vDNA (yellow) bound at the active sites (indicated by Mn ions, black spheres). The corresponding loops of the HIV CTD (purple) position four Arg and two Lys residues (light blue side chains and labels) in the same area. PFV CTD Arg362, which corresponds to HIV CTD Lys258, may also contribute to vDNA binding near the active site.

sufficiently long to pass around the docked CTD. Instead, the connectivity between the NTD and CCD may be different than that observed in the PFV-IN structures.

An alternate conformation and connectivity is observed for L1 in the structure of the HIV-2 IN NTD−CCD bound to LEDGF (PDB ID 3F9K).[16] In 3F9K, the NTD is connected to its contacting CCD, with L1 lying on the CCD surface (Figure 1B). The compatibility of this L1 position with vDNA binding was examined.

First, models were made of the HIV CCD and vDNA based on 3F9K that included aspects of the PFV-IN structure induced by complex formation. The vDNA model took its structure from the PFV-IN complex, but the nucleotide sequence was adjusted to bind HIV-IN.[43] To make the HIV CCD model, the PFV CCD active-site loop (residues 209−219) was grafted into the HIV CCD, replacing HIV-IN residues 140−150, and the second Mn ion was added to the active site (see Methods).

To test the quality of these models, the vDNA-9 model was docked to the HIV CCD model. The reference vDNA position was determined by superposition of the HIV CCD onto the PFV CCD (see Methods). In 26 of the top 30 vDNA-9 placements, the vDNA minor groove lay over helix α4, as found in the PFV-IN structure. The nine hits (Table 2, run 1) put the 3′-hydroxyl oxygen 3.1−6.0 Å from the Mn ion. The other placements within this group were either shifted one to two base pairs along the DNA axis or rotated 180°, putting the 3′-end of the uncleaved strand near the active site.

Adding the HIV NTD (NTD + CCD) tightened the clustering, giving 27 vDNA-9 placements with the 3′-hydroxyl within 6.0 Å of the Mn ion. Only 16 of these were within the RMSD cutoff (Table 2, run 2) because of variation in the position of the two-nucleotide overhang of the uncleaved strand (Figure 4A). These two nucleotides contact the active-site loop in the PFV-IN structure, but the HIV active-site loop may have fewer hydrogen bonds with the vDNA. The side chain of PFV-IN Thr210 forms a hydrogen bond with the second phosphate group of the uncleaved strand,[30] but corresponding Val141 of HIV-2 (or Ile141 of HIV-1) cannot

Table 2. Analysis of Docked Components from the HIV-IN NTD−CCD Structure 3F9K

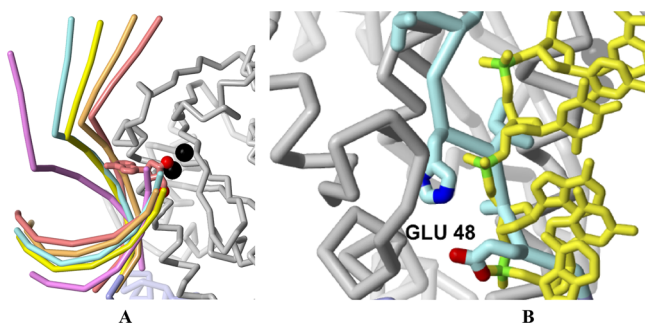| run	stationary molecule or complex	moving molecule	hits in top 30		hits in top 2000		best ranked hit (RMSD)
			no.	RMSD (Å)	no.	RMSD (Å)	
1	CCD	vDNA-9	9	2.0 to 6.0	169	0.8 to 6.0	1 (3.9)
2	NTD + CCD	vDNA-9	16	2.0 to 6.0	250	0.8 to 6.0	1 (3.9)
3	NTD−CCD	vDNA-9	0		3	2.7 to 4.0	230 (4.0)
4	CCD + vDNA	NTD	5	2.2 to 4.9	221	0.9 to 6.0	8 (3.9)

Figure 4. L1 of HIV-IN NTD−CCD structure clashes with bound vDNA. (A) Docking vDNA to the HIV NTD + CCD. In the absence of the L1 linker, the vDNA fits well at the CCD active site (gray tubes with two active-site Mn ions as black spheres), but it shows a wide range for the two-nucleotide overhang (upper left). Docked vDNA are rank 14 (red, RMSD = 2 Å), rank 1 (orange, RMSD = 3.9 Å), rank 2 (yellow, RMSD = 6.0 Å), rank 13 (light blue, RMSD = 8.6 Å), and rank 14 (purple, RMSD = 11.7 Å). (B) The HIV L1 (light blue) blocks the vDNA-binding site on CCD. L1 residues Glu48, Ala49, and Ile50 occupy the groove that binds the phosphate backbone of the uncleaved vDNA strand (yellow). The Glu48 side chain lies in a pocket occupied by a phosphate group (green P atom).

form this hydrogen bond. PFV-IN His213 forms a hydrogen bond with the third phosphate group of the uncleaved strand,[30] but corresponding HIV Asn144 is too short to form the same hydrogen bond in the modeled loop. Even with the reduced contact of the two-nucleotide overhang, the docked vDNA showed good agreement with the active-site contacts found in the PFV-IN structure.

Adding L1 to make the full HIV NTD−CCD construct dramatically changed the results (Table 2, run 3). All 30 top-ranked vDNA placements positioned the DNA major groove over helix α4, moving the 3′-hydroxyl far from the active site. Only three poorly ranked hits appeared among the 2000 top-ranked vDNA-9 placements (Table 2). These hits fit well at the active site, but the opposite end of the vDNA-9 moved away from the CCD surface, reducing contacts with both the CCD and the NTD.

Superposition of the PFV and HIV IN structures revealed the source of the poor vDNA fit to NTD−CCD (Figure 4B). The HIV L1 in 3F9K lies adjacent to helix α4, filling the CCD surface groove that binds the phosphate backbone of the uncleaved vDNA strand in the PFV structures. HIV L1 residues 48−50 overlap the bound phosphate backbone, with the Glu48 side chain in a phosphate binding pocket. Thus, the L1 position in 3F9K, the only HIV-IN structure that shows density for the L1, appears to be incompatible with vDNA binding.

■ DISCUSSION

Computational docking with DOT has proved to be particularly effective for defining the DNA-binding surface of a protein in the absence of a crystallographic structure of the

complex.[42] DOT has been applied to protein−DNA interactions of DNA-repair enzymes,[44−47] transcription factors,[40] the nucleosome,[48] and the CCD of HIV-IN.[21] Large, favorably ranked clusters from the dockings successfully identify active sites,[21,47] map out large DNA-binding regions,[40,47] and reveal multiple points of DNA contact with a protein.[48] Key for predicting protein−DNA interactions is the DOT electrostatic energy term, which gives a good representation[40] of the distinctive electrostatic properties that DNA-binding proteins need to pull the highly charged DNA substrate out of solution.[49−51] With this electrostatic term, the majority of the 2000 best ranked DNA placements surround the bound DNA position, clearly distinguishing the DNA-binding surface on the protein.

Here, DOT was applied to a novel use: dissection of the crystallographic structure of the PFV-IN assembly, a multichain, multidomain, DNA-bound complex. The crystallographic interactions among the NTD, CCD, and vDNA were reproduced, but the crystallographic CTD position was never identified as a favorable energy cluster. Instead, two dominant symmetry-related positions were found for the CTD over the two vDNA-occupied active sites, a region partially occupied by the linker regions in the PFV crystallographic structures.

The CCD by itself showed a well-defined vDNA-binding site. Docking vDNA to the PFV CCD identified the key features: insertion of CCD helix α4 into the DNA minor groove and placement of the 3′-hydroxyl of the cleaved DNA strand near the active-site metal ion. The same vDNA-binding site was found for the HIV CCD, using a model based on the unbound HIV-IN CCD structure and the active-site geometry of the DNA-bound PFV CCD. Previous DNA docking studies have also identified this region as the dominant DNA-binding site on the CCD.[20,21] Thus, the CCD alone has the elements to define the vDNA-binding surface and the orientation of the DNA over that surface. In contrast, docked DNA fragments only partially overlapped the host-DNA binding site and showed no distinct orientation over this site. Since each CCD has only a partial binding site for the host DNA, an effective host-DNA-binding site may require formation of the full assembly organized by the bound vDNA.

The consistent reproduction of the interactions among the CCD, its contacting NTD, and bound vDNA in both PFV-IN and HIV-IN suggest that this three-component unit may be the basic building block of the assembly, but the two three-component complexes in the PFV-IN tetramer have no contact with each other. Instead, the spacing between them, and hence between the two active sites, must be determined by other components in the assembly. In the PFV-IN structures, the four linker regions and two CTDs of the inner subunits appear to determine the spacing. We propose that the CTD is the main determinant of the active-site spacing, with a CTD dimer binding between the two active sites, replacing L1 and L2 in the PFV-IN structures. The docked PFV CTD clusters representing this dimer are comparable in size to those from dockings
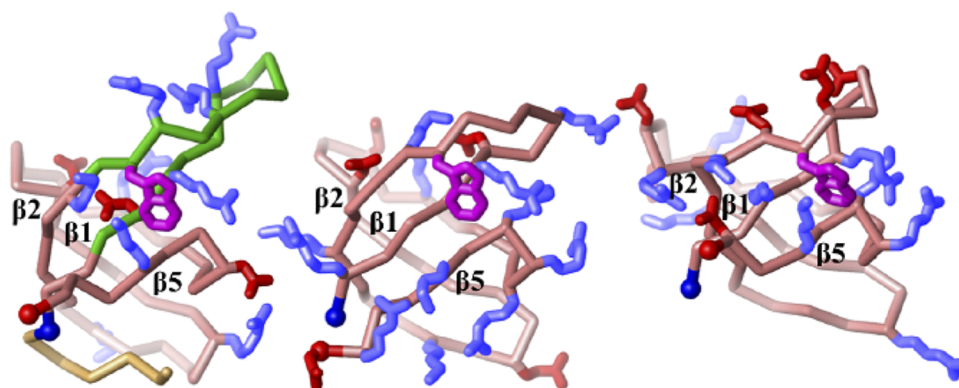
Figure 5. Distribution of the charged side chains in the CTDs from PFV, HIV, and ASV INs. All of the Arg and Lys residues (blue) and Asp and Glu residues (red) are shown. In the PFV CTD (left), the largest cluster of Arg and Lys residues is on and around the loop between strands β1 and β2. This loop (green backbone) is shown in the conformation built to follow the corresponding loop of HIV-IN. The HIV CTD (center) also has a cluster of Arg and Lys residues on the loop between β1 and β2, but it has another cluster on β5 and the preceding loop. In contrast, the ASV CTD (right) has a cluster of Asp and Glu residues (red) on the loop between β1 and β2, with Arg and Lys residues clustered mainly on β5 and the helical loop preceding β5. All CTDs are shown in the same orientation with the conserved Trp (magenta) at the start of β2 (residue 337 in PFV-IN, 235 in HIV-IN). The second β sheet of the CTD is in back, is relatively hydrophobic, and consists of β3, β4, and the second half of β2.

that reproduced crystallographic interactions (Table 1), even though the CTD has not undergone an induced fit for this site. The docked CTD has extensive contacts with both vDNA molecules, consistent with the significant positive charge (+6 for residues 315−375) created by nine Arg and Lys residues distributed over the CTD surface (Figure 5). The PFV CTD loop between β-strands 1 and 2, rebuilt based on the conformation of the corresponding loop in the HIV CTD, lies between the two active sites, a region that is unoccupied in the PFV-IN crystallographic structures. This loop has a cluster of positively charged side chains, so the CTD dimer inserts eight Arg side chains between the two active sites, neutralizing the substantial negative charge created by the two vDNA molecules and the incoming host DNA.

The predicted CTD position is more consistent with the length and conformation of the L2 linkers of HIV-IN and ASV-IN than the CTD position in the PFV-IN structure. The docked CTD decreases the distance spanned by L2 from 48 to 36 Å, close to the length of the helical L2 in the HIV CCD−CTD structure[17] (Table 3). Like the PFV CTD, the HIV CTD

Table 3. Distances Spanned by the L1 and L2 Linkers

system	region spanned	distance (Å)
L2: From the Last Cα in the CCD to the First Cα in the CTD		
PFV-IN (3OY9)	A270 Cα to A321 Cα	48.4
PFV-IN, docked CTD	A270 Cα to A321 Cα	36.4
HIV-IN (1EX4)	201 Cα to 223 Cα, chains A and B	33.8, 34.4
ASV-IN (1C0M)	213 Cα to 222 Cα, chains A and B	19.2, 24.2
L1: From the Last Cα in the NTD to the First Cα of the CCD		
PFV-IN (3OY9)	A102 Cα to A123 Cα	42.6
HIV-IN (3F9K)	A46 Cα to A59 Cα	30.1
HIV-IN (3F9K)	A46 Cα to B59 Cα	32.5

has a cluster of positively charged side chains on the loop between β-strands 1 and 2 (Figure 5), suggesting a similar orientation at the CTD-binding site. The ASV CTD, with its very different charge distribution (Figure 5), is likely to bind in a different orientation at the CTD-binding site, which could significantly shorten the distance between the ASV CTD N-terminus and the CCD. The closest PFV CTD backbone atom

is just 22 Å from the C-terminus of the PFV CCD, a distance easily spanned by the eight-residue ASV-IN L2[18] (Table 3). With a unique orientation for the ASV CTD, vDNA-contacting residues would not correspond to the same structural positions on the HIV and PFV INs. Therefore, experimental data implicating a specific ASV CTD residue in vDNA binding[36] may not be directly transferable to the HIV and PFV CTDs.

The position predicted for the CTD moves L2 away from the active site, but the CTD would still conflict with the L1 conformation found in the PFV-IN structure. Structural analysis revealed that the only other observed L1 conformation, that from HIV-IN structure 3F9K,[16] interferes with vDNA binding. In addition, this configuration would create a vDNA-binding site with the NTD and CCD from the same inner subunit, contradicting the finding that the HIV NTD functions in *trans* with the CCD.[52,53] We propose a third possibility: L1 connects the CCD of an inner subunit to the NTD of the outer subunit belonging to the same CCD dimer. The L1 would span a shorter distance than that in the PFV-IN structure (Table 3) and no longer clash with the docked CTD.

The proposed CTD and L1 and L2 linker positions are shown in a model of the vDNA-bound HIV-IN dimer (Figure 6). The HIV CCD and CTD were superposed onto the PFV CCD and the docked CTD. The HIV L2 helix was inserted so that the three L2 lysine side chains extend toward the vDNA, as suggested by Stroud et al.[17] In the model shown in Figure 6, the CCD, L2, and CTD of one inner subunit wrap around the vDNA. The L2 and CTD could also be contributed by the outer subunit of the dimer because the C-terminal residues of the CCDs comprising the dimer are just 5.5 Å apart (arrow in Figure 6B). This is consistent with studies on combinations of HIV-IN mutants showing that the CTD can function both in *trans* and in *cis* with the CCD.[52,53] The vDNA-binding surface is completed by the NTD from the outer subunit of the CCD dimer (Figure 6B), consistent with the NTD functioning in *trans* to the CCD.[52,53] The three positively charged side chains on the CTD loop between β1 and β2 (Lys228, Arg231, and Lys236) and adjacent residue Lys258 lie between the two bound vDNA ends (Figure 3). CTD residue Arg231 is on the most extended part of the loop and is therefore likely to contact host DNA. Consistent with a role in host-DNA binding, mutation of Arg231 decreased strand-transfer activity but
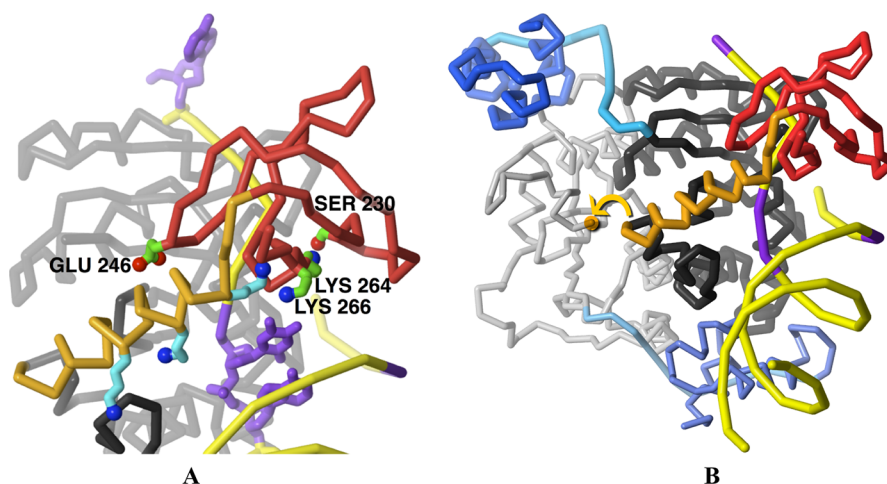
Figure 6. Model of the HIV-IN dimer. (A) The CCD (gray), L2 (orange), and CTD (red) from one inner subunit wrap around the vDNA (yellow with purple indicating DNA sites that cross-link with the CTD). (B) The NTD functions in *trans*, and the CTD can function in both *cis* and *trans*. The chain of one inner subunit (thick tubes) and the NTD−CCD region of the outer subunit (thin tubes) that form the dimer are shown. The NTD (blue, lower right) that contacts the CCD (dark gray) of the inner subunit to complete the vDNA (yellow) binding surface comes from the outer subunit (light blue L1 linker, lower left; light gray CCD, left). Similarly, the NTD (blue, upper left) of the inner subunit contacts the CCD of the outer subunit. Due to the proximity of the C-termini of the two CCDs (indicated by the yellow arrow and an orange sphere), the vDNA-contacting L2 and CTD could also be supplied by the outer subunit.

retained 3′-processing activity.[54] CTD residue Lys258 (equivalent to PFV-IN Arg362) is more deeply buried between the two active sites and therefore may be key for orienting the CTD. Mutation of HIV Lys258 to Glu or Ser results in loss of both activities.[54] CTD side chains Ser230, Glu246, Lys264, and Lys266 are near the bound vDNA (Figure 6A), consistent with cross-linking studies.[5,7,23,55,56]

Why does the predicted CTD dimer not occur in the crystal if it is the biological interaction? In the PFV-IN crystal, the CTD does not contact other tetramers, so it is not directly influenced by crystal packing. However, the CTD position may be indirectly affected by the vDNA, which contributes to formation of the crystal lattice.[34] The PFV-IN structures show an angle of about 90° between the two bound vDNA ends, but SAXS studies on the PFV-IN/vDNA complex in solution find a larger angle.[34] The smaller angle in the crystal decreases the area between the two bound vDNA ends, which could prevent formation of the CTD dimer.

Models constructed before the first vDNA-bound PFV-IN structure in 2010[30] reflected the incomplete, and sometimes contradictory, structural and biochemical data. Although it was known that two dimers were needed to provide the two active sites, the only constraint for orienting the two CCD dimers was the distance between the two vDNA insertion points. The HIV-IN CCD−CTD structure 1EX4[17] showed two positions for the CTD, neither of which matched any CTD position in ASV-IN (Figure 1C) or SIV-IN.[5] The HIV-IN NTD−CCD structure (PDB ID 1K6Y) in 2001[15] gave an NTD position distant from that later reported in 2009[16] (PDB ID 3F9K), which is close to the PFV-IN NTD (Figure 1B). The varied positions of the NTD and the CTD, attached by long, flexible linker regions, resulted in such a large variation among the HIV-IN models that no role for either the NTD or CTD could be clearly identified. Models that combined the NTD−CCD from 1K6Y with the CCD−CTD from 1EX4 proposed a variety of NTD and CTD contacts, but they provided no specific structural mechanism for the spacing of the two active sites.[23,26−28] Other

models[25,29] proposed contacts between the two CCD dimers, resulting in more compact protein tetramers.

The 2010 DNA-bound PFV-IN structures[30,31] reveal a very different assembly and suggest specific functions for the observed NTD and CTD. The two CCD dimers extend away from each other with no direct contact, forming an approximately planar protein structure in which the two vDNA fragments extend away from each other on one face (Figure 1A) and host DNA binds on the opposite face. The NTD, along with its contacting CCD, forms a continuous vDNA-binding surface. The CTD, through its contacts both vDNA ends, is the only domain directly involved in spacing the two active sites. Two HIV-IN models based on the PFV-IN structures maintain the positions of the folded domains, but differ in the L2 conformation. In one,[37] the L2 was fully extended to span the long distance between the CCD and CTD. In the other,[57] the helical conformation of HIV-IN L2 found in 1EX4 was retained, but this required unfolding CTD residues 223−237, adding them to L2 as an extended region and then incorporating the 18 C-terminal residues into the CTD. Our model is based on the PFV-IN structures, but the assembly of PFV-IN domains was first explored by computational docking. The discovery of a new, favored CTD position preserves the HIV-IN L2 helical conformation and the HIV-IN CTD fold observed by both crystallography[17] and NMR.[58,59] Furthermore, the proposed model identifies the role of the CTD dimer to be the central determinant of the geometry of the two active sites.

■ CONCLUSIONS

Given the diverse sequences and structures among retroviral integrases, it is unclear if properties found for other integrases apply to HIV-IN, the key clinical target. Here, an investigation of whether the assembly found in the PFV-IN crystallographic structures, which are the only IN structures that contain bound vDNA, applies to the biological assembly of HIV-IN was performed. Computational docking reproduced the interactions of the PFV-IN NTD and CCD with vDNA and found

corresponding interactions for HIV-IN, supporting a similar grouping for these three components in both assemblies. Computational docking did not reproduce the CTD position in the PFV-IN structures, but it did identify a new position for the CTD that better accommodates the large variation in the L2 sequence, length, and conformation among ASV, HIV, and PFV INs. Therefore, we propose that the same CTD-binding site is shared among the three INs, but it is not the site observed in the PFV-IN crystallographic structures. The two symmetry-related docked clusters suggest that a CTD dimer is formed, with each CTD contacting both vDNA molecules to align the active sites. Two positively charged loops, one from each CTD, extend between the two active sites, a region that is unoccupied in the PFV-IN structures. The loops help to neutralize the large negative potential created by the proximity of the two vDNA ends and assist the binding of the host DNA. The differences in size, sequence, and charge distribution among the retroviral CTDs suggest that each forms a unique CTD dimer, which would explain why each retroviral IN has a unique spacing between the two vDNA insertion points in the host DNA.

This study demonstrates that rigid-body, systematic search macromolecular docking is an effective tool for evaluating crystallographic structures of complex assemblies. For integrase, computational docking combined with structural analysis has provided a model that resolves apparent contradictions between biochemical and structural data. This approach may be especially useful for systems containing protein−DNA interactions or multidomain chains with long, flexible linkers that can be influenced by crystal packing.

■ METHODS

Preparation of Coordinates for the Docking Calculations. The PFV IN coordinates were taken from the crystallographic structure of PFV IN with bound viral DNA (PDB ID 3OY9),[30] which contains the NTD, L1, CCD, L2, and CTD of an inner subunit (chain A), the CCD of the contacting outer subunit (chain B), which forms a CCD−CCD dimer with chain A, and the bound vDNA fragment. The biological tetramer was generated by symmetry operations, and the two inner subunits (chain A and symmetry-related chain A′) and the two bound vDNA molecules were extracted.

The PFV NTD, CCD, and CTD (Table 4) were defined based on previous sequence and structure alignment,[30] with the following exceptions. Residues 271−278 of L2 were added to the C-terminus of the CCD. These residues contact the NTD and have B-factors similar to the CCD domain. Keeping these L2 residues as part of the CCD significantly improved dockings

in which the NTD was the moving molecule. These residues are part of loop 271−286, which packs between the NTD and CCD before extending toward the CTD; there is no equivalent loop in the HIV and ASV structures. The N-terminus of the CTD was extended to include L2 residues 310−320. Residues 315−320 have no equivalent in HIV-IN and ASV-IN, but they are part of the globular PFV CTD, shielding the hydrophobic side chains of Ile366, Val343, Val346, and Val352 from solvent. The Trp315 side chain faces the PFV CTD interior, contributing to its hydrophobic core. Residues 310−314 from L2 include vDNA-contacting side chains Ser311 and Arg313[30] and therefore may contribute to vDNA binding. Retaining a few residues preceding the CTD can improve docking results by eliminating placements that would clash with the L2.

The full-length vDNA in the PFV structures has an uncleaved strand of 19 nucleotides (chain C) with a two-nucleotide overhang (residues 1 and 2) at the 5′-end and a cleaved strand of 17 nucleotides (chain D) that ends with the active-site 3′-hydroxyl (residue 17). When DNA is the moving molecule in computational docking, it is best modeled by fragments 8−12 bp in length. Longer fragments give energetically similar placements shifted along the DNA axis, obscuring the correctly docked position. Therefore, when vDNA was the moving molecule, a truncated active-site vDNA fragment was used, vDNA-9, which contained the first nine nucleotides of the uncleaved strand and the last seven nucleotides of the cleaved strand (Table 4). This fragment includes all of the contacts with the CCD and the CTD. A linear 10 bp B-DNA fragment with the vDNA sequence, including the two cleaved nucleotides, was built with the Nucleic Acid Builder (NAB) program[60] and docked to the individual PFV-IN domains. A second B-form DNA fragment with a completely different sequence was also docked (results not reported), which gave the same distribution of the top 30 and top 2000 placements, showing that the B-DNA dockings were not dependent on the DNA sequence.

Multiple component constructs were created by combining individual components. These large assemblies were assigned as the stationary molecule for computational efficiency in the docking calculation.[41] Full-length vDNA, which extends about 2 bp beyond protein contacts in the crystal, was used in these constructs to ensure a good electrostatic model near the blunt end.[40,41] CCD + vDNA contained the CCD (chain A) and its contacting vDNA (chains C and D). In NTD + CCD, the NTD from chain A′ was added to CCD (chain A) to give a continuous vDNA-contacting surface. (CCD + vDNA)$_2$ contained both vDNA molecules and the CCDs from chains A and A′ to create the catalytic core of the IN tetramer. In (NTD + CCD + vDNA)$_2$, both NTDs were added to the catalytic core. In (CCD + CTD + vDNA)$_2$, both CTDs were added to the catalytic core. In (CCD + CTDdock + vDNA)$_2$, a symmetry-related pair of CTDs found by computation docking was added to the catalytic core.

HIV-IN coordinates were taken from the crystallographic structure of the NTD−CCD construct bound to lens epithelium-derived growth factor (PDB ID 3F9K).[61] Chain A was selected, which includes coordinates for linker L1. Coordinates of the NTD, L1, and CCD of HIV-IN were selected based on Hare et al.,[30] except that L2 residues 202−207 were included with the CCD (Table 4).

We rebuilt the HIV-IN to reflect the induced fit upon vDNA binding. In 3F9K, the active-site loop (residues 140−150) is very different from that in the vDNA-bound PFV-IN structure

Table 4. Selected Coordinates of IN Domains Used in Dockings

component	coordinates used in dockings	from Hare et al.[30]
PFV-IN, PDB ID 3OY9		
NTD	Chain A: 8−102, Zn	1−102
CCD	Chain A: 123−278, 2 Mn	123−270
CTD	Chain A: 310−375	321−374
vDNA	Chain C: 1−19, Chain D: 1−17	
vDNA-9	Chain C: 1−9, Chain D: 11−17	
HIV-IN, PDB code 3F9K		
NTD	Chain A: 4−46, Zn	1−46
CCD	Chain A: 59−207, 2 Mn	59−201
NTD−CCD	Chain A: 4−207, Zn, 2 Mn	

3OY9. 3F9K has only one metal ion bound in the active site, whereas the PFV-IN structures show two. The Cα atoms of PFV residues around the Mn-binding ligands and the loop ends (residues 127−129, 184−186, 207−209, 220−222) were superposed onto the corresponding 3F9K atoms (residues 63−65, 115−117, 138−140, and 151−153) using the molecular graphics program Insight (Accelrys, Inc.). PFV-IN active-site loop residues 209−219 and the second metal ion from 3OY9 were transplanted into 3F9K. Four of the 11 side chains were replaced to provide the HIV-IN sequence. To complete the active site, the 3F9K Glu152 side-chain conformation was adjusted to match the conformation of the corresponding Mn-binding ligand in 3OY9, resulting in a oxygen-Mn bond length of 2.2 Å.

To build the model of the vDNA for the HIV-IN dockings, the nucleotide sequence of the PFV-IN vDNA-9 coordinates was changed to match the sequence used in HIV-IN strand-transfer assays.[43] To obtain the reference position for the vDNA, the PFV CCD with bound vDNA was superposed onto the model of the HIV-IN CCD by the structural elements that contact the vDNA and form the active site: HIV CCD residues 61−66 (β1), 73−78 (β2), 83−86 (β3), 140−150 (modeled active-site loop), and 151−157 (α4). The phosphate backbone of the vDNA model for HIV-IN was then superposed onto the PFV-IN vDNA.

The DOT Calculation. Docking calculations were performed with the program DOT,[38,39,41] part of the DOT2 Suite[41] distributed by the Computational Center for Macromolecular Structure at the San Diego Supercomputer Center (http://www.sdsc.edu/CCMS). In DOT, one molecule (the moving molecule), represented by its atomic positions with partial atomic charges, is systematically moved within the shape and electrostatic potentials of a second molecule or complex (the stationary molecule), providing an exhaustive translational and rotational search. Interaction energies for all configurations of the two molecules are evaluated as correlation functions, which are efficiently computed with fast Fourier transforms. Molecular properties for the integrase protein and DNA molecules were calculated using utilities in the DOT2 Suite, including the program REDUCE,[62] to add hydrogen atoms, determine His side chain protonation states, and correct the geometry of Asn, Gln, and His side chains; the program MSMS[63] to calculate molecular surfaces that encompass the volumes defining the shape potential of the stationary molecule; the AMBER library of heavy atoms with added polar hydrogens[64] to assign partial atomic charges; and the program APBS[65,66] to calculate the electrostatic potential of the stationary molecule by finite difference methods to solve the linearized Poisson−Boltzmann equation.

Docking calculations used a cubic grid with 1 Å grid spacing ranging from 128 to 256 Å on a side. The dimensions of the grid were determined by the sizes of the molecules to ensure that the moving molecule fit within the grid when it was close to the stationary molecule.[41] The moving molecule was centered at each grid point in 54 000 orientations (a 6.0° rotational spacing). A grid size of 128 Å^3 gave about 108 billion placements of the moving molecule about the stationary molecule.

The 2000 placements with the most favorable interaction energies were kept for each docking. The interaction energy is the sum of the electrostatic and van der Waals intermolecular energy terms. The van der Waals term is a count of the number of atoms of the moving molecule that overlap a 3 Å thick favorable layer around the excluded volume of the of the stationary molecule, with each atom contributing −0.1 kcal/mol to the van der Waals energy.[41] If a moving molecule atom overlapped the excluded volume of the stationary molecule, then the configuration was eliminated. The electrostatic energy is calculated as the set of atomic point charges of the moving molecule placed in the electrostatic potential of the stationary molecule.[38] The electrostatic potential was calculated with the program APBS,[66] using a dielectric of 3 for the protein, a dielectric of 80 for the surrounding environment, an ion exclusion radius of 1.4 Å, and an ionic strength of 150 mM. The resulting electrostatic potential was clamped to make the electrostatic energy term compatible with the soft van der Waals energy term by modulating the potential near the protein surface.[41]

Analysis of Docking Results. The 2000 top-ranked placements were analyzed by calculating the root-mean-square deviation (RMSD) between the docked and crystallographic positions of the moving molecule, holding the stationary molecule fixed. For the PFV-IN dockings, where all components come from the DNA-bound assembly, an RMSD of 5 Å or less was considered a hit (Table 1), typically reproducing the residue−residue contacts seen in the crystal. RMSD calculations against the docked CTD position also used a 5 Å cutoff. In the HIV-IN dockings, a more lenient RMSD cutoff of 6 Å was used because both the HIV CCD and the vDNA were models rather than coordinates from a bound complex. RMSD calculations were done for all Cα atoms in protein domain evaluations and for all C3′ atoms in DNA evaluations. The 30 top-ranked placements of the moving molecule were analyzed in detail with computer graphics, and the 2000 top-ranked placements were visualized by examining the distribution of their geometric centers over the stationary molecule.

Rebuilding CTD Loop 324−339. PFV CTD loop residues 324−339 were rebuilt to follow the conformation of the corresponding HIV CTD loop (residues 226−237). The Cα atoms of HIV CTD residues 225−227 and 236−238 were superposed onto the Cα atoms of PFV CTD residues 323−325 and 338−340 of a symmetry-related pair of docked CTDS (ranks 4 and 15 from run 8, Table 1). HIV CTD residues 226−237 were then transplanted into the PFV CTDs. The resulting loop is four residues shorter than the PFV CTD loop. Visual inspection indicated that an insertion between residues 327 and 334 (sequence VARPASLR) was most likely to give a loop fitting between the active sites. In our grafted model, the distance between the Cα atoms of residues 327 and 334 was 5.7 Å. The BriX database[67] was searched for eight-residue loops with a distance of 5.6−5.8 Å between the initial and final Cα atoms and with proline in the fourth position. PDB ID 2GHS residues A128−A135 (sequence RMHPSGAL) were transplanted by superposition of the main-chain atoms of A128 and A135 onto PFV CTD atoms 327 and 334. The residue sequence was corrected, and the side chains of residues Arg326, Arg329, Arg334, Arg336, Lys339, and Leu333 were adjusted to relieve steric clashes with the vDNA.

Modeling HIV-IN. To build the model of the HIV-IN dimer with bound vDNA, the starting point was the HIV CCD model that had been fit to the PFV CCD. HIV CTD residues 224−226 (β1), 235−239 (β2), 248−252 (β3), and 257−262 (β4) from chain A of PDB structure 1EX4 were then superposed onto the Cα atoms of the corresponding residues of the docked PFV CTD (rank 4 from Table 1, run 8). The L2 (residues

A202 to A222) from 1EX4 was inserted between CCD residue 201 and CTD residue 223, and side-chain geometries were adjusted to avoid clashes. L1 (residues 47 to 58) was built as a Cα trace between the last residue of the NTD from the inner subunit (residue A46) and the first residue of the CCD from the outer subunit (residue B59). A copy of the NTD−CCD with the rebuilt L1 was superposed onto the CCD of the inner subunit to complete the dimer shown in Figure 6B.

■ AUTHOR INFORMATION

Corresponding Author
*Tel: 858 784-8028. E-mail: vickie@sdsc.edu.
Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ ABBREVIATIONS

RMSD, root-mean-square deviation; PFV, prototype foamy virus

■ REFERENCES

(1) Hazuda, D. J.; Felock, P.; Witmer, M.; Wolfe, A.; Stillmock, K.; Grobler, J. A.; Espeseth, A.; Gabryelski, L.; Schleif, W.; Blau, C.; Miller, M. D. *Science* **2000**, *287*, 646−650.
(2) Grinsztejn, B.; Nguyen, B.-Y.; Katlama, C.; Gatell, J. M.; Lazzarin, A.; Vittecoq, D.; Gonzalez, C. J.; Chen, J.; Harvey, C. M.; Isaacs, R. D. *Lancet* **2007**, *369*, 1261−1269.
(3) Shimura, K.; Kodama, E.; Sakagami, Y.; Matsuzaki, Y.; Watanabe, W.; Yamataka, K.; Watanabe, Y.; Ohata, Y.; Doi, S.; Sato, M.; Kano, M.; Ikeda, S.; Matsuoka, M. *J. Virol* **2008**, *82*, 764−774.
(4) Walmsley, S. L.; Antela, A.; Clumeck, N.; Duiculescu, D.; Eberhard, A.; Gutiérrez, F.; Hocqueloux, L.; Maggiolo, F.; Sandkovsky, U.; Granier, C.; Pappa, K.; Wynne, B.; Min, S.; Nichols, G. *N. Engl. J. Med.* **2013**, *369*, 1807−1818.
(5) Jaskolski, M.; Alexandratos, J. N.; Bujacz, G.; Wlodawer, A. *FEBS J.* **2009**, *276*, 2926−2946.
(6) Li, M.; Mizuuchi, M.; Burke, T. R., Jr; Craigie, R. *EMBO J.* **2006**, *25*, 1295−1304.
(7) Michel, F.; Crucifix, C.; Granger, F.; Eiler, S.; Mouscadet, J.-F.; Korolev, S.; Agapkina, J.; Ziganshin, R.; Gottikh, M.; Nazabal, A.; Emiliani, S.; Benarous, R.; Moras, D.; Schultz, P.; Ruff, M. *EMBO J.* **2009**, *28*, 980−991.
(8) Dyda, F.; Hickman, A. B.; Jenkins, T. M.; Engelman, A.; Craigie, R.; Davies, D. R. *Science* **1994**, *266*, 1981−1986.
(9) Bujacz, G.; Alexandratos, J.; Zhou-Liu, Q.; Clement-Mella, C.; Wlodawer, A. *FEBS Lett.* **1996**, *398*, 175−178.
(10) Goldgur, Y.; Dyda, F.; Hickman, A. B.; Jenkins, T. M.; Craigie, R.; Davies, D. R. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 9150−9154.
(11) Goldgur, Y.; Craigie, R.; Cohen, G. H.; Fujiwara, T.; Yoshinaga, T.; Fujishita, T.; Sugimoto, H.; Endo, T.; Murai, H.; Davies, D. R. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 13040−13043.
(12) Maignan, S.; Guilloteau, J.-P.; Zhou-Liu, Q.; Clement-Mella, C.; Mikol, V. *J. Mol. Biol.* **1998**, *282*, 359−368.
(13) Greenwald, J.; Le, V.; Butler, S. L.; Bushman, F. D.; Choe, S. *Biochemistry* **1999**, *38*, 8892−8898.
(14) Molteni, V.; Greenwald, J.; Rhodes, D.; Hwang, Y.; Kwiatkowski, W.; Bushman, F. D.; Siegel, J. S.; Choe, S. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2001**, *57*, 536−544.
(15) Wang, J.-Y.; Ling, H.; Yang, W.; Craigie, R. *EMBO J.* **2001**, *20*, 7333−7343.
(16) Hare, S.; Shun, M.-C.; Gupta, S. S.; Valkov, E.; Engelman, A.; Cherepanov, P. *PLoS Pathog.* **2009**, *5*, e1000259.
(17) Chen, J. C.-H.; Krucinski, J.; Miercke, L. J. W.; Finer-Moore, J. S.; Tang, A. H.; Leavitt, A. D.; Stroud, R. M. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 8233−8238.
(18) Yang, Z.-N.; Mueser, T. C.; Bushman, F. D.; Hyde, C. C. *J. Mol. Biol.* **2000**, *296*, 535−548.
(19) Shi, K.; Pandey, K. K.; Bera, S.; Vora, A. C.; Grandgenett, D. P.; Aihara, H. *PLoS One* **2013**, *8*, e56892.
(20) Perryman, A. L.; McCammon, J. A. *J. Med. Chem.* **2002**, *45*, 5624−5627.
(21) Adesokan, A. A.; Roberts, V. A.; Lee, K. W.; Lins, R. D.; Briggs, J. M. *J. Med. Chem.* **2004**, *47*, 821−828.
(22) Zhu, H. M.; Chen, W. Z.; Wang, C. X. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 475−477.
(23) Gao, K.; Butler, S. L.; Bushman, F. *EMBO J.* **2001**, *20*, 3565−3576.
(24) Heuer, T. S.; Brown, P. O. *Biochemistry* **1998**, *37*, 6667−6678.
(25) Podtelezhnikov, A. A.; Gao, K.; Bushman, F. D.; McCammon, J. A. *Biopolymers* **2003**, *68*, 110−120.
(26) De Luca, L.; Pedretti, A.; Vistoli, G.; Barreca, M. L.; Villa, L.; Monforte, P.; Chimirri, A. *Biochem. Biophys. Res. Commun.* **2003**, *310*, 1083−1088.
(27) Karki, R. G.; Tang, Y.; Burke, T. R., Jr.; Nicklaus, M. C. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 739−760.
(28) Wang, L.-D.; Liu, C.-L.; Chen, W.-Z.; Wang, C.-X. *Biochem. Biophys. Res. Commun.* **2005**, *337*, 313−319.
(29) Wielens, J.; Crosby, I. T.; Chalmers, D. K. *J. Comput.-Aided Mol. Des.* **2005**, *19*, 301−317.
(30) Hare, S.; Gupta, S. S.; Valkov, E.; Engelman, A.; Cherepanov, P. *Nature* **2010**, *464*, 232−236.
(31) Maertens, G. N.; Hare, S.; Cherepanov, P. *Nature* **2010**, *468*, 326−330.
(32) Hare, S.; Maertens, G. N.; Cherepanov, P. *EMBO J.* **2012**, *31*, 3020−3028.
(33) Chen, Z.; Yan, Y.; Munshi, S.; Li, Y.; Zugay-Murphy, J.; Xu, B.; Witmer, M.; Felock, P.; Wolfe, A.; Sardana, V.; Emini, E. A.; Hazuda, D.; Kuo, L. C. *J. Mol. Biol.* **2000**, *296*, 521−533.
(34) Gupta, K.; Curtis, J. E.; Krueger, S.; Hwang, Y.; Cherepanov, P.; Bushman, F. D.; Van Duyne, G. D. *Structure* **2012**, *20*, 1918−1928.
(35) Kotova, S.; Li, M.; Dimitriadis, E. K.; Craigie, R. *J. Mol. Biol.* **2010**, *399*, 491−500.
(36) Peletskaya, E.; Andrake, M.; Gustchina, A.; Merkel, G.; Alexandratos, J.; Zhou, D.; Bojja, R. S.; Satoh, T.; Potapov, M.; Kogon, A.; Potapov, V.; Wlodawer, A.; Skalka, A. M. *PLoS One* **2011**, *6*, e27751.
(37) Krishnan, L.; Li, X.; Naraharisetty, H. L.; Hare, S.; Cherepanov, P.; Engelman, A. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 15910−15915.
(38) Ten Eyck, L. F.; Mandell, J. G.; Roberts, V. A.; Pique, M. E. In Proceedings of the 1995 ACM/IEEE Supercomputing Conference, San Diego, CA, December 4−8; Hayes, A., Simmons, M., Eds.; IEEE Computer Society Press: Los Alamitos, CA, 1995; p 22. www.sdsc. edu/CCMS/Papers/DOT_sc95.html.
(39) Mandell, J. G.; Roberts, V. A.; Pique, M. E.; Kotlovyi, V.; Mitchell, J. C.; Nelson, E.; Tsigelny, I.; Ten Eyck, L. F. *Protein Eng., Des. Sel.* **2001**, *14*, 105−113.
(40) Roberts, V. A.; Case, D. A.; Tsui, V. *Proteins: Struct., Funct., Genet.* **2004**, *57*, 172−187.
(41) Roberts, V. A.; Thompson, E. E.; Pique, M. E.; Perez, M. S.; Ten Eyck, L. F. *J. Comput. Chem.* **2013**, *34*, 1743−1758.
(42) Roberts, V. A.; Pique, M. E.; Ten Eyck, L. F.; Li, S. *Proteins: Struct., Funct., Genet.* **2013**, *81*, 2106−2118.
(43) Wang, Y.; Klock, H.; Yin, H.; Wolff, K.; Bieza, K.; Niswonger, K.; Matzen, J.; Gunderson, D.; Hale, J.; Lesley, S.; Kuhen, K.; Caldwell, J.; Brinker, A. *J. Biomol. Screening* **2005**, *10*, 456−462.

(44) Hopfner, K.-P.; Karcher, A.; Craig, L.; Woo, T. T.; Carney, J. P.; Tainer, J. A. *Cell* **2001**, *105*, 473−485.

(45) Fan, L.; Fuss, J. O.; Cheng, Q. J.; Arvai, A. S.; Hammel, M.; Roberts, V. A.; Cooper, P. K.; Tainer, J. A. *Cell* **2008**, *133*, 789−800.

(46) Hammel, M.; Rey, M.; Yu, Y.; Mani, R. S.; Classen, S.; Liu, M.; Pique, M. E.; Fang, S.; Mahaney, B. L.; Weinfeld, M.; Schriemer, D. C.; Lees-Miller, S. P.; Tainer, J. A. *J. Biol. Chem.* **2011**, *286*, 32638−32650.

(47) Roberts, V. A.; Pique, M. E.; Hsu, S.; Li, S.; Slupphaug, G.; Rambo, R. P.; Jamison, J. W.; Liu, T.; Lee, J. H.; Tainer, J. A.; Ten Eyck, L. F.; Woods, V. L., Jr. *Nucleic Acids Res.* **2012**, *40*, 6070−6081.

(48) Fan, L.; Roberts, V. A. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 8384−8389.

(49) Nadassy, K.; Wodak, S. J.; Janin, J. *Biochemistry* **1999**, *38*, 1999−2017.

(50) Jones, S.; van Heyningen, P.; Berman, H. M.; Thornton, J. M. *J. Mol. Biol.* **1999**, *287*, 877−896.

(51) Jones, S.; Shanahan, H. P.; Berman, H. M.; Thornton, J. M. *Nucleic Acids Res.* **2003**, *31*, 7189−7198.

(52) Engelman, A.; Bushman, F. D.; Craigie, R. *EMBO J.* **1993**, *12*, 3269−3275.

(53) van Gent, D. C.; Vink, C.; Oude Groeneger, A. A. M.; Plasterk, R. H. A. *EMBO J.* **1993**, *12*, 3261−3267.

(54) Serrao, E.; Krishnan, L.; Shun, M. C.; Li, X.; Cherepanov, P.; Engelman, A.; Maertens, G. N. *Nucleic Acids Res.* **2014**, *42*, 5164−5176.

(55) Heuer, T. S.; Brown, P. O. *Biochemistry* **1997**, *36*, 10655−10665.

(56) Esposito, D.; Craigie, R. *EMBO J.* **1998**, *17*, 5832−5843.

(57) Johnson, B. C.; Metifiot, M.; Ferris, A.; Pommier, Y.; Hughes, S. H. *J. Mol. Biol.* **2013**, *425*, 2133−2146.

(58) Lodi, P. J.; Ernst, J. A.; Kuszewski, J.; Hickman, A. B.; Engelman, A.; Craigie, R.; Clore, G. M.; Gronenborn, A. M. *Biochemistry* **1995**, *34*, 9826−9833.

(59) Eijkelenboom, A. P. A. M.; Sprangers, R.; Hard, K.; Puras Lutzke, R. A.; Plasterk, R. H. A.; Boelens, R.; Kaptein, R. *Proteins: Struct., Funct., Genet.* **1999**, *36*, 556−564.

(60) Macke, T.; Case, D. A. In *Molecular Modeling of Nucleic Acids*; Leontes, N. B., Santa Lucia, J., Jr., Eds.; American Chemical Society: Washington, DC, 1998; Vol. *682*, pp 379−393.

(61) Cherepanov, P.; Maertens, G. N.; Proost, P.; Devreese, B.; Van Beeumen, J.; Engelborghs, Y.; De Clercq, E.; Debyser, Z. *J. Biol. Chem.* **2003**, *278*, 372−381.

(62) Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. *J. Mol. Biol.* **1999**, *285*, 1735−1747.

(63) Sanner, M. F.; Olson, A. J.; Spehner, J.-C. *Biopolymers* **1996**, *38*, 305−320.

(64) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765−784.

(65) Gilson, M. K.; Davis, M. E.; Luty, B. A.; McCammon, J. A. *J. Phys. Chem.* **1993**, *97*, 3591−3600.

(66) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10037−10041.

(67) Vanhee, P.; Verschueren, E.; Baeten, L.; Stricher, F.; Serrano, L.; Rousseau, F.; Schymkowitz, J. *Nucleic Acids Res.* **2011**, *39*, D435−D442.