

Alexei A. Podtelezhnikov¹

Kui Gao²

Frederic D. Bushman²

J. Andrew McCammon^{1,3}

¹Howard Hughes Medical
Institute,

University of California

San Diego,

9500 Gilman Dr.,

La Jolla,

CA 92093-0365

²Infectious Disease

Laboratory,

Salk Institute,

10010 North Torrey Pines Rd.,

La Jolla,

CA 92037

Modeling HIV-1 Integrase Complexes Based on Their Hydrodynamic Properties

³Department of Chemistry and

Biochemistry,

Department of Pharmacology,

University of California

San Diego,

9500 Gilman Dr.,

La Jolla,

CA 92093-0365

Received 30 April 2002;

accepted 14 May 2002

Abstract: We present a model structure of a candidate tetramer for HIV-1 integrase. The model was built in three steps using data from fluorescence anisotropy, structures of the individual integrase domains, cross-linking data, and other biochemical data. First, the structure of the full-length integrase monomer was modeled using the individual domain structures and the hydrodynamic properties of the full-length protein that were recently measured by fluorescence depolarization. We calculated the rotational correlation times for different arrangements of three integrase domains, revealing that only structures with close proximity among the domains satisfied the experimental data. The orientations of the domains were constrained by iterative tests against the data on cross-linking and footprinting in integrase–DNA complexes. Second, the structure of an integrase dimer was obtained by joining the model monomers in accordance with the available dimeric crystal structures of the catalytic core. The hydrodynamic properties of the dimer were in agreement with the experimental values. Third, the active sites of the two model dimers were placed in agreement with the spacing between the sites of integration on target DNA as well as the integrase–DNA cross-linking data, resulting in twofold symmetry of a tetrameric complex. The model is consistent with the experimental data indicating that the F185K substitution, which is found in the model at a tetramerization interface, selectively disrupts correct complex formation in vitro

Correspondence to: Alexei A. Podtelezhnikov; email:
apodtele@mccammon.ucsd.edu

Contract grant sponsor: NIH, NSF supercomputer centers,
NBCR, and W. M. Keck Foundation

Biopolymers, Vol. 68, 110–120 (2003)

© 2002 Wiley Periodicals, Inc.

and HIV replication in vivo. Our model of the integrase tetramer bound to DNA may help to design anti-integrase inhibitors. © 2002 Wiley Periodicals, Inc. Biopolymers 68: 110–120, 2003

Keywords: time-resolved fluorescence anisotropy; cross-linking; oligomerization; anti-HIV drug design

INTRODUCTION

The continuing emergence of resistant strains of the human immunodeficiency virus type 1 (HIV-1) makes the search for new anti-HIV drugs an imperative. One attractive area of anti-HIV research is HIV-1 integrase. HIV encodes three enzymes, reverse transcriptase, protease, and integrase, but only the first two have been exploited as targets for antivirals.

The viral-encoded integrase protein catalyzes integration of cDNA copy of the viral RNA genome into human genome. The integration step completes the formation of a provirus, which contains all the information necessary for the synthesis of the viral RNAs and proteins for the formation of new virions (for review, see Ref. 1).

The DNA breaking and joining reactions that mediate retroviral cDNA integration are outlined in Figure 1. A linear form of the viral cDNA serves as the immediate precursor of the integrated provirus. Prior to integration, integrase removes two nucleotides from the 3' end of each LTR (long terminal repeat) DNA, exposing recessed 3' hydroxyl groups (steps 1 and 2). This terminal cleavage step may serve to remove heterogeneous extra nucleotides occasionally added to the cDNA ends by reverse transcriptase.^{2,3} Integrase then catalyzes attack by the recessed 3' hydroxyl groups on phosphodiester bonds on each target DNA strand, resulting in joining of each viral DNA 3' end to protruding 5' ends in the target (step 3). The sites of joining on each strand of the target DNA are separated by 5 base pairs. Unfolding of this integration intermediate yields gaps at each junction between viral and host DNA, and a 5' two-base flap derived from the viral DNA. Gap repair and connection of the remaining DNA strands are probably carried out by host DNA repair systems (steps 4 and 5).^{1,4}

This paper focuses on the organization of the HIV-1 integrase complex with DNA. Experimental data strongly indicate that integrase functions as an oligomer of the monomeric protein. The monomer comprises the N-terminal domain (NTD), catalytic core domain (CCD), and C-terminal domain (CTD).¹ The complete structure of integrase and its complexes with DNA has not yet been resolved. Here we present a model of these oligomers based on experimental

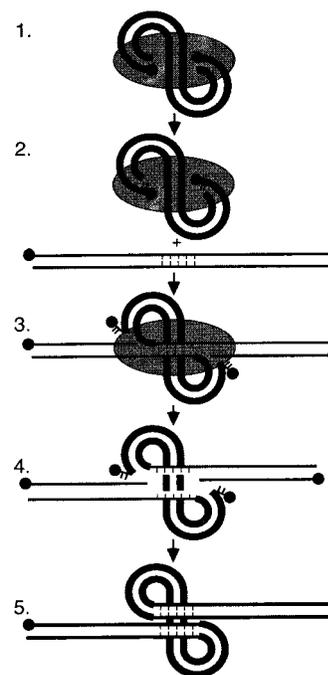


FIGURE 1 DNA cutting and joining reactions mediating cDNA integration. The HIV cDNA and the target DNA are shown by thick and thin lines, respectively. HIV integrase, a part of preintegration complex, is represented by the gray oval. The DNA 5' ends are represented by large circular dots. The explanation of the stages of integration is in the text.

evidence. The proposed model may provide a more complete structural basis for guiding drug discovery, including inhibition of complex formation.

As a starting point for construction of the oligomeric models we used the structures of the three domains of integrase. These were previously determined by x-ray crystallography and NMR.^{5–12} Some information on mutual arrangements of the domains can be deduced from the data on integrase–DNA cross-linking^{13–17} and protein footprinting¹⁸ of integrase and its complexes with DNA. Other structural constraints result from the observed 5-base-pair separation between the two sites of integration in the target DNA¹ and the symmetry in the concerted reactions of joining at each cDNA end. Several models of the HIV-1 integrase–DNA complexes have been previously suggested.^{12,13,17}

In this work we propose an improved model that takes advantage of the recent measurements of the time-resolved fluorescence anisotropy (TFA) for the HIV-1 integrase and its oligomers.^{19,20} From the measured rotational correlation times, the authors analyzed the shape of monomeric HIV-1 integrase and its oligomerization under different conditions. Here we used computer simulations of hydrodynamic properties for structural interpretation of the TFA results.

If the structure of a molecule is known, the hydrodynamic properties can be calculated and values of the rotational correlation times can be deduced.²¹ In this work we attempted to solve the inverse problem of obtaining structural information from the experimental rotational correlation times. The solution could only be modeled at low resolution. We were, however, able to find relative arrangement of the three integrase domains in a model monomer that satisfied the TFA data.

Our model structure of an integrase dimer was based on the crystallographically observed interfaces between the catalytic core domains. We also propose a symmetrical tetrameric structure of integrase capable of concerted insertion reactions. The proposed dimeric and tetrameric assemblies of these monomers were checked against the TFA measurements. To fit the model to the cross-linking data, we adjusted the orientation of the integrase domains relative to the DNA substrates and to each other. The model was verified to satisfy the majority of cross-linking and footprinting data available to us. In addition, the model placed the F185 residue at the tetramerization interface. We showed here that the F185K substitution disrupts correct integration *in vitro* under fastidious assembly conditions, and F185K is known to disrupt viral replication,²² consistent with the idea that the model contains an authentic tetramerization interface.

METHODS

TFA: Theory and Simulations

TFA experiments pursue the measurement of the correlation times of molecular motions such as the overall tumbling of a macromolecule as well as its internal fluctuations.²³ For tumbling, the rotational correlation times are related to the rotational diffusion coefficients of the molecule, which are, in turn, determined by its size and shape. For a spherical molecule, when the diffusion coefficients for rotations in all directions are the same, the anisotropy obeys a single-exponential decay²³:

$$r(t) = r_0 e^{-t/\tau}$$

The rotational correlation time, τ , is given by the Perrin equation:

$$\tau = \frac{\eta V}{kT} = \frac{4\pi r^3 \eta}{3kT}$$

where η is the solution viscosity, k is the Boltzmann constant, T is the temperature, V is the volume of the rotating molecule, and r is the radius of the molecule. Thus, one can extract the volume of the spherical molecule from the experimentally measured rotational correlation time.

Proteins are rarely spherical. Theoretically, for molecules with complex shapes, the anisotropy, $r(t)$, obeys a five-exponential decay:

$$r(t) = a_1 e^{-t/\tau_1} + a_2 e^{-t/\tau_2} + a_3 e^{-t/\tau_3} + a_4 e^{-t/\tau_4} + a_5 e^{-t/\tau_5}$$

where the coefficients a_i depend on orientations of chromophores in the molecule and the rotational correlation times τ_i are defined by the rotational diffusivity matrix of the molecule.^{23–25} For randomly oriented chromophores, the coefficients a_i are identical. If the number of chromophores per molecule is large and their orientations are diverse, the coefficients a_i are comparable to each other and all five correlation times should be detectable in TFA experiments.²⁵ Reliable extraction of the correlation times τ_i from noisy fluorescence data is possible using the maximum entropy method.²⁶ In practice, when the rotational correlation times are close to each other, only the range of the correlation times can be approximately obtained.

One way to analyze the TFA-measured rotational correlation times in structural terms relies on the assumption of the shape of an ellipsoid of revolution for the molecule. Approximate expressions for the rotational correlation times can then be obtained analytically.²³ Using these expressions, one can interpret the measured values in terms of the lengths of long and short axes of the ellipsoid representing the molecule.

In this work we used an alternative procedure of interpreting TFA results based on the numerical technique developed by Garcia de la Torre and co-workers.^{21,24,25} This technique allows one to calculate the five rotational correlation times from the atomic-level structure of the molecule for given temperature and solution viscosity. In these calculations the structure of the molecule is assumed to be rigid. Its van der Waals surface is mimicked by a shell of identical touching beads with less than atomic radii.²¹ The hydrodynamic interactions between the beads are approximated by Rotne–Prager tensor,²⁷ which corresponds to stick hydrodynamic boundary conditions as more appropriate for proteins in water. The rotational component of the Brownian diffusivity matrix for the entire molecule is then obtained from the tensor of hydrodynamic interactions.²⁴ The five rotational correlation times are calculated from the eigenvalues of the rotational diffusivity matrix.²⁵ A public-domain program for shell building and calculating the rotational correlation times, HYDROPRO, is available at <http://leonardo.fc.u.um.es/macromol/software.htm>.

To solve the inverse problem of obtaining structural information from the experimental rotational correlation times using the described procedure, one can vary unknown parameters of the structure and calculate the rotational correlation times. The values of the parameters can then be found by fitting the calculated rotational correlation times to the measured ones. Note that, theoretically, only limited structural information can be obtained given just five values of the measured correlation times.

Modeling of Hydrodynamic Properties of HIV-1 Integrase

The integrase monomer was modeled by combining the structures of the three integrase domains taken from Protein Data Bank (PDB). The structure of NTD of HIV-1 integrase, amino acids from 1 through 47 resolved by NMR, was taken from PDB structure 1WJA.⁵ Only one chain of two identical chains of the NTD dimer was used. The CCD structure was taken from PDB structure 1BIS.⁹ We used chain B of the structure with all amino acids from 56 to 209 resolved by x-ray crystallography. The structure of CTD was taken from PDB structure 1IHV.^{10,23} Only one of the two identical chains of the dimeric structure was used. In this structure, amino acids 219 through 270 are resolved by NMR.

Connections between the domains in the trial conformations of the full-length integrase were assumed to be rigid, as justified in this work. We varied only separations and orientations between the domains. The atomic structure of each domain remained invariable and corresponded to the published structures. The model structures of the integrase oligomers were constructed by combining the structures of several model monomers. All the monomers in a given oligomeric structure were identical. We assumed that the oligomers are formed by rigid connections between the monomers. To calculate the rotational correlation times of the model monomeric integrase and its oligomers, we used unmodified HYDROPRO. The calculations were done at the temperature of 20°C and the viscosity of 0.01 poise.

Our modeling accounts for 253 of 288 amino acids of HIV-1 integrase monomer. The remaining 35 amino acids belong to the linkers between the domains and are at the C-terminal of the protein. We omitted them from our models since they comprise a small portion of the protein and should not affect the simulation results significantly. We, however, placed NTD and CTD so that they could be connected to CCD by plausible linkers.

Molecule Manipulation

To prepare trial conformations of the integrase domains at different separation and orientation between them, we used InsightII 2000 (Accelrys, Inc., San Diego, CA). The manipulations of the molecules in preparing dimeric and tetrameric structures as well as the fitting of DNA into the structure were also done using InsightII 2000. All the structures were checked to avoid overlapping atoms.

The orientations of the domains were adjusted to satisfy cross-linking and footprinting data that define the segments of the domains facing viral and host DNA upon binding. The experimentally cross-linkable parts of the integrase domains and DNA were placed at distances less than 10 Å, which corresponds to the length of the cross-linking agents. This improved the accuracy of the model based on hydrodynamic calculations.

Integration In Vitro

Integration in vitro was carried out essentially as described earlier.^{28,29} Plasmid pHIIN1¹³ was used as target DNA. Oligonucleotides of sequence HU534T: 5' G TGA CTA ATA AGG GTC TGT GGA AAA TCT CTA GCA; HU540B: 5' ACT GCT AGA GAT TTT CCA CA GAC CCT TAT TAG TCA CGT AC modeled the viral LTR ends. Oligonucleotides were end-labeled by treatment with kinase and γ -³²P ATP, and reaction products were visualized by autoradiography. Integrase proteins, modified to contain hexahistidine tags, were over expressed in *Escherichia coli* and purified by chromatography on nickel-chelating sepharose (Novagen, Inc., Madison, WI) essentially as described earlier.²⁸ NC (nucleocapsid) protein refolded with Zn was the kind gift of Robert Gorelick (NCI, Frederick, MD).

The model structure of the HIV-1 integrase tetramer complexed with DNA is available from authors upon request.

RESULTS AND DISCUSSION

Hydrodynamic Properties of HIV-1 Integrase

The method of hydrodynamic calculation that we used in this work required the molecule under investigation to be rigid (first section under Methods). In the case of a flexible molecule, one would observe two correlation modes: one corresponding to the overall tumbling of the molecule and the other to its internal flexibility.²³ The experimental data for integrase monomers, however, show correlation times in the very narrow range around 20 ns.²⁰ The dimer of HIV-1 integrase was also characterized by the rotational correlation times in the narrow range around 40 ns.²⁰ Thus, the integrase monomer and its dimer can be considered as stiff structures.

In addition, if the integrase monomers were flexible in the junctions between the domains, one would expect to see the correlation times similar to the values for the individual domains in the spectrum of correlation times. We investigated the hydrodynamic properties of each integrase domain separately from the others. Table I shows the correlation times predicted for each domain. All the values are much lower

than 20 ns and were not observed in the TFA experiments. This confirms that the integrase monomer is a sufficiently stiff molecule and our treatment of its hydrodynamic properties is valid.

The structures of NTD and CTD determined by NMR and the CCD structure determined by x-ray crystallography do not overlap but rather have to be connected by linkers of 8–9 amino acids. This leaves great freedom for the placement of the integrase domains into a full-length monomeric structure. We investigated a number of the trial placements with different distances and orientations among them. The structures of each individual domain were preserved as originally determined. The distances between the domains were limited by the length of the plausible linkers between them. We also avoided overlapping of the domains.

To select candidates for an appropriate integrase model among the trial conformations, we compared the ranges of the five calculated rotational correlation times for each trial conformation with the ranges reported for HIV-1 integrase in the TFA experiments.²⁰ Note that this comparison is only valid if all the rotational correlation times were experimentally detected. The experimental results were obtained using the fluorescence of tryptophanyl residues of HIV-1 integrase. The rotational correlation time distributions were reliably recovered using the maximum entropy method.²⁶ The integrase has seven tryptophans in diverse orientations. The fluorescence of multiple tryptophans must have contributed to the distributions since four distinctive excited-state lifetime classes were found in the experiment. Therefore, it is safe to assume that the ranges of all five rotational correlation times should have been detected in the experiments (first section under Methods).

The simulated rotational correlation times for four representative conformations are given in Table II. Our results show that the hydrodynamic properties of the model monomer structures strongly depend on the relative distances between the domains. The rotational correlation time gradually decreases with decreasing separation between the domains. For the most of the

Table I The Rotational Correlation Times for the Integrase Domains Separate from Each Other^a

Integrase Domain	Rotational Correlation Time, ns
CCD	9.5–11
CTD	4.5–5.3
NTD	4.0–4.7

^a The ranges between the minimal and maximal calculated correlation times are given.

Table II Dependence of Rotational Correlation Times on the Separation Between the Integrase Domains^a

Model	Domain Separations, Å	Rotational Correlation Time, ns
M0	48, 42, 57	30–45
M1	39, 42, 48	25–32
M2	33, 33, 28	23–27
M3	20, 28, 33	19–20
Experimental		17–23

^a The separations between the centers of mass of the domains are given in the order: NTD–CCD, CCD–CTD, CTD–NTD. The ranges between the minimal and maximal calculated correlation times are given.

trial conformations (M0–M2) the calculated correlation times are notably larger than the experimental values. Only when the integrase domains were placed as close to each other as possible (M3) could we obtain the rotational correlation times in agreement with the experimental values. From this test we concluded that the domains should be tightly compacted together.

This model is consistent with data from protein footprinting. It was proposed that CCD is protected against proteolytic attack by the presence of NTD and CTD, probably because the terminal domains sterically prevent the protease access to the core.¹⁸

The close proximity among the domains proposed here is inconsistent with the recently published crystal structure of the integrase containing CTD and CCD.⁶ In this structure, CTD and CCD are joined by an extended α -helical linker and separated by 42 Å. We calculated that the correlation times for the monomer from this structure are 18–35 ns, which is larger than the values from the TFA measurements. We also found that the rotational correlation times for the dimer from this structure, 35–80 ns, were also larger than the experimental values of 35–50 ns.²⁰ Possibly, the disagreement may be due to integrase adopting a particularly extended conformation in the presence of detergent used in the preparation of the crystals. In Table II, model M0 corresponds to the structure of an active integrase monomer from the model suggested by Gao et al.¹³ that was based on this two-domain crystal structure. The rotational correlation times for the integrase monomer and dimer in this model were twice as large as the experimental values. This comparison suggests that both the crystal structure and the integrase model based on it are inconsistent with hydrodynamic data. We note that in one monomer in the extended structure the α -helix between domains is kinked, suggesting that the monomer could become more compact by bending around this hinge point.

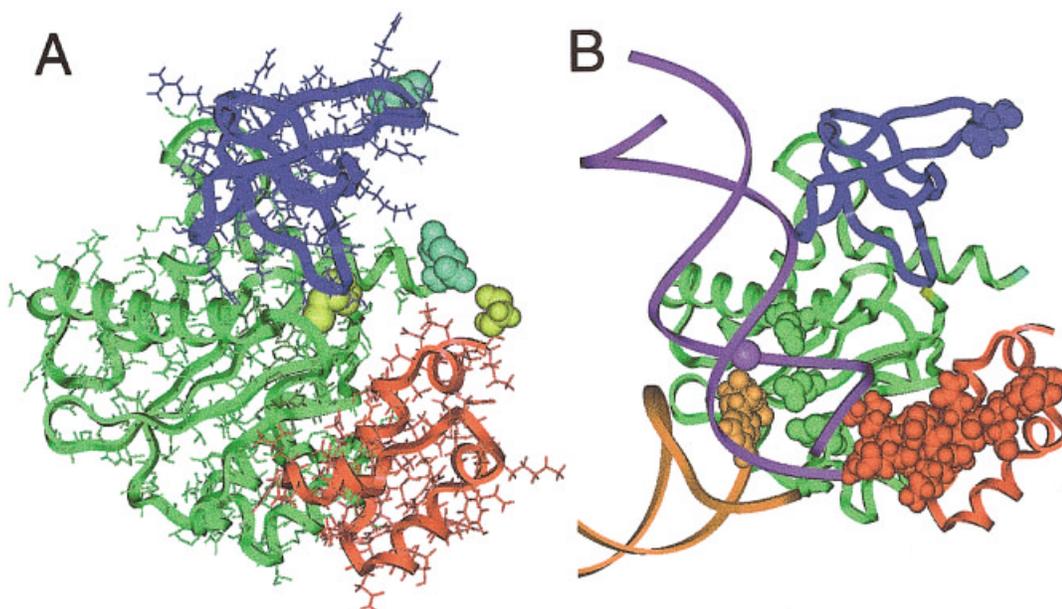


FIGURE 2 A model structure of an integrase monomer. (A) Three domains of the integrase monomer are shown in different colors: NTD is in red, CCD is in green, and CTD is in blue. The distances between the domain termini to be connected into a full length monomer are 16 Å for NTD-CCD linker and 20 Å for CCD-CTD linker. The boundaries of the linkers are shown by bulk light blue and yellow residues. (B) Partial verification of the model against cross-linking data. The backbone of the viral LTR is in orange, the segment of host DNA is in purple. Adenosine at the recessed 3' end (in orange) is located near the catalytic triad D64, D116, and E152 (in green). The site of the joining reaction is marked by purple phosphorus atom. The amino acids 1-12 of NTD (in red) are in contact with minor groove of host DNA in the vicinity of base pairs 2 and 3 downstream from the site of integration. The residue 270, shown in blue, represents the boundary of the peptide 271-288, which was shown to cross-link with the host DNA upstream from the site of joining.

We also calculated the rotational correlation times for different orientations of the domains at fixed separations between their centers of mass. We found that the relative orientation of the integrase domains does not affect the hydrodynamic properties strongly. Independent of the domain orientation, the rotational correlation times always fell in the same range of values at constant spacing between centers of mass. Since only the range of the rotational correlation times was determined in the TFA experiments, it is impossible to model the integrase structure at higher resolution based solely on the hydrodynamic simulations. Therefore, to determine the orientations of the integrase domains, we used the data from cross-linking and footprinting experiments.

Model Monomer

After placing the integrase domains in a compact arrangement, we manually refined the orientations of the domains to satisfy the data from cross-linking between the integrase and DNA substrates near the

site of integration.¹³⁻¹⁷ These data define the segments of the integrase domains that are in close contact with host and viral DNA in the complex. The model of the integrase monomer that we were able to construct is presented in Figure 2.

Initially, we only considered cross-linking with DNA within 3 base pair (bp) from the site of integration because these data are more likely to define segments of a single active integrase monomer that are in contact with DNA (Figure 2B). In this model, viral DNA binds only to CCD at the terminal 3 bases (here and thereafter we will number the LTR base pairs from the recessed A of the 5'-CA-3' end of the cleaved LTR). Our model is in agreement with the cross-linking of the second and third base pairs of LTR with the peptides containing amino acids 49-69 and 139-152.¹⁶ In addition, the LTR 3' end is near the residues K156 and K159, in accordance with photo-cross-linking data,¹⁴ and close to the catalytic triad D64, D116, and E152. The orientation of the DNA strands with respect to the catalytic core domain is in agreement with the experimental predictions.¹⁴

Host DNA forms extensive contacts with all three integrase domains in our model (Figure 2B). This is in agreement with the ability of CCD and CTD to non-specifically bind DNA.³⁰ The putative zinc finger of the NTD is also in contacts with host DNA in the model. More specifically, the second and third base pair downstream from the site of integration are close to the peptide containing amino acids 1–12, in agreement with the earlier photo-cross-linking data.¹⁶ The part of the host DNA upstream from the site of integration is close to CTD. The cross-linking and footprinting data indicate close contact between the host DNA and amino acids 271–288.^{16,18} These amino acids are not resolved in the CTD structures available to us. In our models, however, the residue 270 is reasonably close to the host DNA.

For the model, the rotational correlation times are in the range from 18 to 22 ns, in agreement with the experimental values. The model has a characteristic oblate shape. This agrees with the ratio of 2.5 for the axes of an oblate ellipsoid representing the molecule in the earlier interpretation of the TFA results.²⁰

The model presented in Figure 2 should not be considered as a unique solution to the problem. It is representative of a class of HIV-1 integrase structures with this particular pattern of domain arrangements. For the model in Figure 2, we found that shifting of NTD and/or CTD by several angstroms around the catalytic core would produce a different model of the same class that would still satisfy the TFA measurements for the integrase monomer. The length of linkers between the domains (8 or 9 amino acids) also permits formation of similar models with slightly shifted domains. Models with slightly shifted domains can also satisfy the cross-linking data that define the contacts between the integrase domains and DNA. The length of the cross-linking agents is about 10 Å so the accuracy of the model is likely in this range.

A different class of models can be suggested. In this class, CTD and NTD have switched positions comparing to those of the model in Figure 2. This model satisfied data from TFA, cross-linking, and the known 5 bp spacing in the target DNA, but did not display a satisfactory tetramer interface or parallel to the Tn5 structure,³¹ as with the favored model (unpublished data and see below). Therefore, we omit these alternative models from our discussions.

Model Dimers

Dimerization of HIV-1 integrase has been shown in numerous reports: each of the integrase domains has been shown to form dimers in crystal or NMR structures.^{5–12,32,33} The dimer of the catalytic core domain

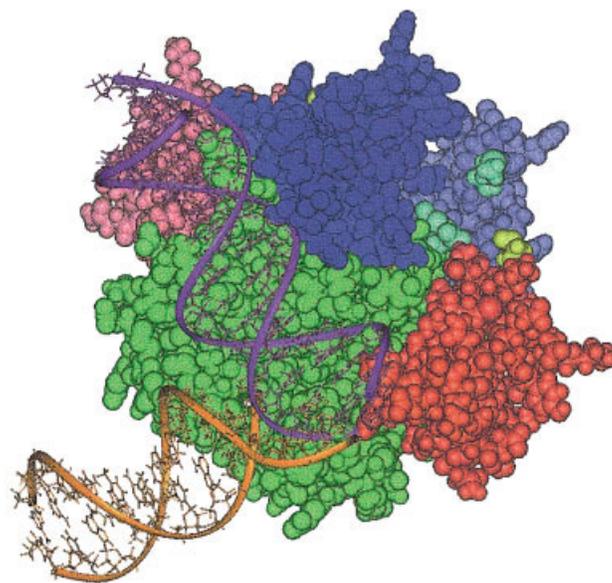


FIGURE 3 A model structure of the integrase dimer. The integrase domains are colored as in FIGURE 2A. The viewer faces one of the monomers in the structure. The other monomer is behind (in pale colors).

appears to be formed consistently along the same interface between the monomers in the crystallographic studies. Therefore, to construct the model integrase dimer, we decided to superimpose the structures of two model monomers with the dimeric crystal structure of catalytic core domain. The resulting dimeric model is presented in Figure 3.

We tested the hydrodynamic properties of the model dimer of HIV-1 integrase. Treating the dimer as a rigid structure, we obtained the rotational correlation times in the range 35–45 ns. This is consistent with the experimental values obtained in the TFA studies of about 35–50 ns.²⁰

We believe that the dimer of HIV-1 integrase formed along the interface between the catalytic cores is not sufficient for performing both joining reactions during integration. The distance between the active sites of such a dimer is about 45 Å, while the sites of integration are separated by exactly 5 base pairs of target DNA¹ with the distance between the sites of about 20 Å on double-stranded B-DNA. Therefore, we think that the joining should involve a tetrameric structure of HIV-1 integrase formed by two dimers. Each of the dimers contributes one active monomer catalyzing reactions and an inactive one serving a structural role.

Model Tetramer Bound to DNA

In the final step of the modeling, we put together a structure of the integrase tetramer based on the fol-

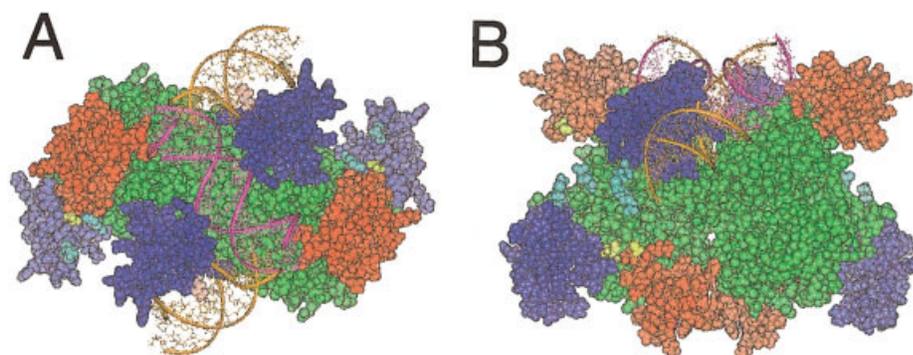


FIGURE 4 A model of HIV-1 integrase tetramer complexed with DNA. The integrase domains are colored as in FIGURE 2. The viral LTRs are in yellow. The host DNA is in purple. The interactions between the DNA and an active monomer are the same as in Figures 2B. Two views of the same complex are presented. (A) The orientation of the dimer on the right is close to the one in FIGURE 3. The axis of twofold symmetry is perpendicular to the plane of the picture. (B) The view of the complex revealing the interface between the dimers. One can see possible interactions between NTDs of inactive monomers at the bottom of the picture.

lowing principles. First, two active sites on the opposite model dimers were placed roughly at 20–25 Å from each other to allow both joining reactions on undisturbed host DNA.³⁴ Second, since both reactions are identical and occur symmetrically on the opposite strands of DNA, the principles of stereochemistry suggest the same (C_2) symmetry of the integrase tetramer. Third, the interface between the dimers should have extensive contacts between them. If possible, CTD or NTD should be involved in this interface, since they were shown to be important for multimerization.¹ The final model tetrameric structure of HIV-1 integrase is shown in Figure 4.

The calculated rotational correlation times for the presented model tetramer are in the range 75–95 ns, which is consistent with the TFA measurements.²⁰ To further verify the final tetrameric structures, we fitted both LTRs of viral DNA and host DNA into the structure and checked it against the cross-linking data.

In the tetrameric model presented here, two CTDs bind the segment of host DNA between the sites of integration. On the other hand, NTDs bind DNA outside the internal segment (Figure 4). This pattern of host DNA binding to the terminal domains of integrase follows from the cross-linking data¹⁶ discussed in the section “Model Monomer” and the common understanding of the integration pathway (Figure 1). We built our model to fit to this pattern of binding to host DNA.

In our model the interface between the dimers involves interactions between two NTDs of inactive monomers (Figure 4B). This observation is consistent with previous observations that the binding of Zn^{2+} to NTD promotes tetramerization.^{20,35}

Strong cross-linking has been detected between the peptide containing amino acids 247–270 of CTD and the fifth base pair of viral LTR.^{15,16} This interaction was also shown to be very specific to the LTR sequence.¹⁵ Later, the protection of the residues E246, K258, and K273 by DNA binding were identified in protein footprinting experiments.¹⁸ Recently, by means of chemical cross-linking, Gao et al.¹³ have shown that E246 is near the major groove at the fifth base pair of LTR. They also demonstrated that this binding occurs in *trans*, i.e., the residue E246 does not belong to the integrase monomer that performs the processing and the joining of this LTR. Our model satisfies these experimental observations. At this stage only a small additional adjustment of the CTD orientation was required to satisfy these cross-linking data. Figure 5 illustrates the proximity between the residue E246 and the position 5 of the viral LTR.

Probing the Model Tetramerization Interface by Site-Directed Mutagenesis

Another important result of our modeling is that it can potentially explain the phenotype of the F185K substitution in integrase. This substitution causes an increase of the integrase solubility.²² In our model structure this residue participates in joining of two integrase dimers into the tetramer, i.e., F185 of one active integrase is in close proximity of the peptide QVRDQ, amino acids 164–168, of the opposite active integrase monomer (Figure 6A).

Integrase containing the mutation F185K has been reported to be near wild type in activity in vitro,³⁶ but we have found that the substitution is actually quite

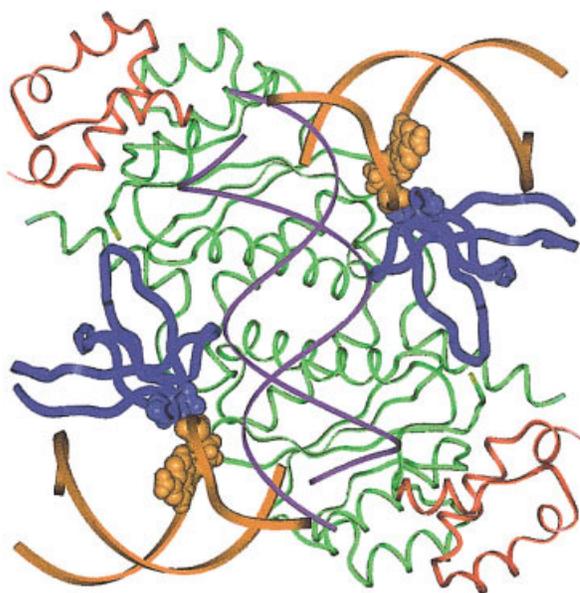


FIGURE 5 The specific interaction between E246 and the fifth base pair of LTR. A fragment of the structure in FIGURE 4 is shown. The fragment includes only active integrase monomers. The colors are the same as in Figure 2B. Adenosines in the fifth position from the recessed 3' ends are shown in bulk orange. These residues were modified for cross-linking purposes.¹³ The neighboring E246 are in bulk blue.

detrimental in some assays (Figure 6B and C). In reactions in which integrase must assemble with an LTR oligonucleotide and carry out integration into a circular plasmid DNA, we find that the F185K substitution is actually severely impaired (Figure 6C, compare lanes 1–4 and 7–10). The viral NC protein stimulates integration in Mg^{2+} ,^{28,29} but even in the presence of NC activity of F185K is not detectable (Figure 6C, lane 6). The F185K mutation also confers a strong defect on viral replication.³⁶ These phenotypes of the F185K substitution *in vitro* and *in vivo* are consistent with the 185 side chain making important protein–protein contacts in the fully assembled integrase multimer.

CONCLUSIONS

In this work we present a model of an integrase tetramer and its complex with viral and host DNA. This model was built from the structures of integrase domains resolved by x-ray crystallography and NMR. We used computer simulations to approximate the relative positions of the IN domains in the full-length IN monomer from the data of the TFA experiments.

The model was also adjusted to satisfy the following constraints imposed by the data on cross-linking the integrase with DNA substrates:

1. The terminus of the viral DNA binds to CCD near the site of integration and has no important contacts with the other integrase domains within 3 base pairs of the recessed 3'-end.^{14–16} Specifically, deoxyadenosine of the conserved 5'-CA-3' of viral DNA is placed in the active site (Figure 2B).
2. More distal in the viral DNA, base pairs 5 and 6 are in contact with peptide 247-270 of CTD of integrase monomer that catalyzes the integration of the opposite viral DNA terminal.^{15,16} Also, E246 is close to the position 5 on the viral DNA, making the predicted *trans* contact¹³ (Figure 5).
3. The host DNA binds CCD near the site of integration and also has extensive contacts with the terminal domains. The internal segment of host DNA between the sites of integration is close to both CTDs of active integrase monomers. Amino acids 1–12 of NTDs bind host DNA outside the internal segment¹⁶ (Figure 2B).

The model is also consistent with additional experimental observations. The interface between the dimers in the model tetramer involves NTDs of inactive monomers. This may serve as an explanation of Zn^{2+} dependence of integrase multimerization.³⁵ Our model structure resembles the structure of Tn5 transposase–DNA complex recently resolved by x-ray crystallography.³¹ Both structures possess twofold symmetry and are similar in orientations of catalytic cores. In addition, the model tetramerization interface contains F185 residue of integrase. This suggests that the F185K substitution may increase the solubility of integrase and disrupt integration *in vivo*³⁶ and *in vitro* (this work) by preventing correct formation of the integrase tetramer.

The presented model is not a unique solution but rather represents our favored class of models. This model can help to define additional experimental tests with the goal of increasing its accuracy. In addition, this model can be improved computationally by drawing upon continuum models to assess the protein oligomers and their interactions with nucleic acids. Further conformational adjustments of the oligomers and their complexes with DNA can be made to satisfy energetic criteria, while maintaining the critical experimental constraints mentioned above. The tetrameric model may be useful

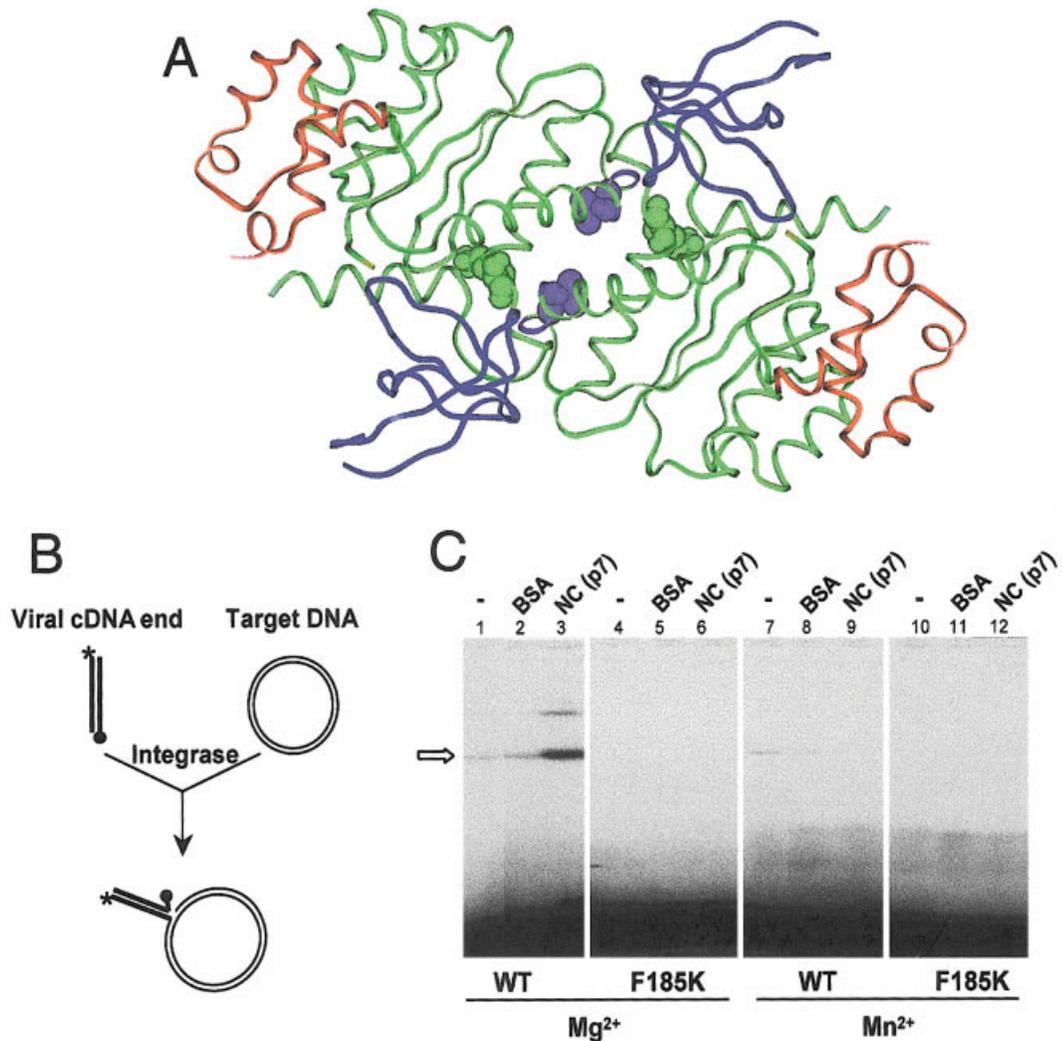


FIGURE 6 Phenotype of F185K substitution. (A) A fragment of the model structure with active monomers only is shown. The color scheme is the same as in previous pictures. F185 that are in the interface between the dimers are shown as a bulk green residues. The close amino acids of the opposite, also active, monomer Q164 and V165 are in purple. (B) Diagram of the integration assay. A synthetic oligonucleotide was used to mimic the viral cDNA end. The oligonucleotide was end labeled (asterisk), so that integration resulted in incorporation of label into the plasmid DNA. Unlabeled DNA 5' ends are shown as balls. (C) Inhibition of integration in vitro by the F185K substitution in integrase. Protein additives are as indicated above the gel. The presence or absence of the F185K substitution is indicated beneath below the gel together with the metal cofactor added. Reaction products (arrow on the autoradiogram) were separated by electrophoresis on an agarose gel and visualized by autoradiography after drying of the gel. Comparison of EtBr staining indicates that the indicated product comigrated with the relaxed circle form of the target DNA. All lanes are from the same exposure of a single experiment.

in defining new drug targets in the interfaces between the monomers.

This work has been supported in part by grants from NIH, the NSF supercomputer centers, NBCR, and the W. M. Keck Foundation.

REFERENCES

1. Varmus, H.; Coffin, J. M.; Hughes, S. H. *Retroviruses*; Cold Spring Harbor Laboratory Press: Plainview, NY; 1997.
2. Miller, M. D.; Farnet, C. M.; Bushman, F. D. *J Virol* 1997, 71, 5382–5390.

3. Patel, P. H.; Preston, B. D. *Proc Natl Acad Sci USA* 1994, 91, 549–553.
4. Hansen, M. S.; Carreau, S.; Hoffmann, C.; Li, L.; Bushman, F. *Genet Eng (NY)* 1998, 20, 41–61.
5. Cai, M.; Zheng, R.; Caffrey, M.; Craigie, R.; Clore, G. M.; Gronenborn, A. M. *Nat Struct Biol* 1997, 4, 567–577.
6. Chen, J. C.; Krucinski, J.; Miercke, L. J.; Finer-Moore, J. S.; Tang, A. H.; Leavitt, A. D.; Stroud, R. M. *Proc Natl Acad Sci USA* 2000, 97, 8233–8238.
7. Eijkelenboom, A. P.; Sprangers, R.; Hård, K.; Puras Lutzke, R. A.; Plasterk, R. H.; Boelens, R.; Kaptein, R. *Proteins* 1999, 36, 556–564.
8. Goldgur, Y.; Dyda, F.; Hickman, A. B.; Jenkins, T. M.; Craigie, R.; Davies, D. R. *Proc Natl Acad Sci USA* 1998, 95, 9150–9154.
9. 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 USA* 1999, 96, 13040–13043.
10. 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.
11. Maignan, S.; Guilloteau, J. P.; Zhou-Liu, Q.; Clément-Mella, C.; Mikol, V. *J Mol Biol* 1998, 282, 359–368.
12. Wang, J. Y.; Ling, H.; Yang, W.; Craigie, R. *EMBO J* 2001, 20, 7333–7343.
13. Gao, K.; Butler, S. L.; Bushman, F. *EMBO J* 2001, 20, 3565–3576.
14. Jenkins, T. M.; Esposito, D.; Engelman, A.; Craigie, R. *EMBO J* 1997, 16, 6849–6859.
15. Esposito, D.; Craigie, R. *EMBO J* 1998, 17, 5832–5843.
16. Heuer, T. S.; Brown, P. O. *Biochemistry* 1997, 36, 10655–10665.
17. Heuer, T. S.; Brown, P. O. *Biochemistry* 1998, 37, 6667–6678.
18. Dirac, A. M.; Kjems, J. *Eur J Biochem* 2001, 268, 743–751.
19. Leh, H.; Brodin, P.; Bischerour, J.; Deprez, E.; Tauc, P.; Brochon, J. C.; LeCam, E.; Coulaud, D.; Auclair, C.; Mouscadet, J. F. *Biochemistry* 2000, 39, 9285–9294.
20. Deprez, E.; Tauc, P.; Leh, H.; Mouscadet, J. F.; Auclair, C.; Brochon, J. C. *Biochemistry* 2000, 39, 9275–9284.
21. de la Torre, J. G.; Huertas, M. L.; Carrasco, B. *Biophys J* 2000, 78, 719–730.
22. Engelman, A.; Liu, Y.; Chen, H.; Farzan, M.; Dyda, F. *J Virol* 1997, 71, 3507–3514.
23. Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*, 2nd ed.; Kluwer Academic/Plenum: New York, 1999.
24. Carrasco, B.; de la Torre, J. G. *Biophys J* 1999, 76, 3044–3057.
25. Garcia de la Torre, J.; Carrasco, B.; Harding, S. E. *Eur Biophys J* 1997, 25, 361–372.
26. Brochon, J. C. *Methods Enzymol* 1994, 240, 262–311.
27. Rotne, J.; Prager, S. *J Chem Phys* 1969, 50, 4831–4837.
28. Carreau, S.; Gorelick, R. J.; Bushman, F. D. *J Virol* 1999, 73, 6670–6679.
29. Carreau, S.; Batson, S. C.; Poljak, L.; Mouscadet, J. F.; de Rocquigny, H.; Darlix, J. L.; Roques, B. P.; Käs, E.; Auclair, C. *J Virol* 1997, 71, 6225–6229.
30. Engelman, A.; Hickman, A. B.; Craigie, R. *J Virol* 1994, 68, 5911–5917.
31. Rice, P. A.; Baker, T. A. *Nat Struct Biol* 2001, 8, 302–307.
32. Yang, Z. N.; Mueser, T. C.; Bushman, F. D.; Hyde, C. C. *J Mol Biol* 2000, 296, 535–548.
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. Dyda, F.; Hickman, A. B.; Jenkins, T. M.; Engelman, A.; Craigie, R.; Davies, D. R. *Science* 1994, 266, 1981–1986.
35. Zheng, R.; Jenkins, T. M.; Craigie, R. *Proc Natl Acad Sci USA* 1996, 93, 13659–13664.
36. Jenkins, T. M.; Engelman, A.; Ghirlando, R.; Craigie, R. *J Biol Chem* 1996, 271, 7712–7718.