Article No. jmbi.1999.3451 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2000) 296, 521–533

JMB

X-ray Structure of Simian Immunodeficiency Virus Integrase Containing the Core and C-terminal Domain (Residues 50-293) - An Initial Glance of the Viral DNA Binding Platform

Zhongguo Chen*, Youwei Yan, Sanjeev Munshi, Ying Li, Joan Zugay-Murphy, Bei Xu, Marc Witmer, Peter Felock, Abigail Wolfe Vinod Sardana, Emilio A. Emini, Daria Hazuda and Lawrence C. Kuo*

Department of Antiviral Research, Merck Research Laboratories, West Point PA 19486-0004, USA

**Corresponding authors*

The crystal structure of simian immunodeficiency virus (SIV) integrase that contains in a single polypeptide the core and the C-terminal deoxyoligonucleotide binding domain has been determined at 3 Å resolution with an *R*-value of 0.203 in the space group $P2_12_12_1$. Four integrase core domains and one C-terminal domain are found to be well defined in the asymmetric unit. The segment extending from residues 114 to 121 assumes the same position as seen in the integrase core domain of avian sarcoma virus as well as human immunodeficiency virus type-1 (HIV-1) crystallized in the absence of sodium cacodylate. The flexible loop in the active site, composed of residues 141-151, remains incompletely defined, but the location of the essential Glu152 residue is unambiguous. The residues from 210-218 that link the core and C-terminal domains can be traced as an extension from the core with a short gap at residues 214-215. The C^{α} folding of the C-terminal domain is similar to the solution structure of this domain from HIV-1 integrase. However, the dimeric form seen in the NMR structure cannot exist as related by the non-crystallographic symmetry in the SIV integrase crystal. The two flexible loops of the C-terminal domain, residues 228-236 and residues 244-249, are much better fixed in the crystal structure than in the NMR structure with the former in the immediate vicinity of the flexible loop of the core domain. The interface between the two domains encompasses a solvent-exclusion area of 1500 $Å^2$. Residues from both domains purportedly involved in DNA binding are narrowly distributed on the same face of the molecule. They include Asp64, Asp116, Glu152 and Lys159 from the core and Arg231, Leu234, Arg262, Arg263 and Lys264 from the C-terminal domain. A model for DNA binding is proposed to bridge the two domains by tethering the 228-236 loop of the C-terminal domain and the flexible loop of the core.

Keywords: SIV; integrase; X-ray crystal structure

Introduction

Integration of viral cDNA into the host genome is a critical step in the life cycle of retroviruses. Integration is catalyzed by the viral enzyme integrase in a two-step process. In the first (3′ end processing), two nucleotides are removed from each 3′ end of the viral DNA leaving behind a 5′ overhang and a new 3′ end beginning with adenosine followed by cytosine. In the second (strand transfer),

Abbreviations used: APS, Advanced Photon Source; ASV, avian sarcoma virus; DBD, DNA-binding domain(s) or the C-terminal domain(s); HIV-1, human immunodeficiency virus type 1; IMCA, Industrial Macromolecular Crystallography Association; rms, root mean square; SIV, simian immunodeficiency virus.

E-mail address of the corresponding author: lawrence_kuo@merck.com

0022-2836/00/020521–13 $35.00/0

the 3′ viral termini are joined to 5′ termini of the host DNA. Subsequent trimming of the 5′ overhangs of viral DNA and repair of the single-stranded gaps of the integrated DNA, catalyzed by host-cell enzymes, completes the proviral integration into the host cell genome. (For recent reviews on retroviral integration, see Katz & Skalka, 1994; Vink & Plasterk, 1993.)

Mutagenesis and proteolytic studies have shown that integrase consists of three functional domains (Engelman & Craigie, 1992). The N-terminal domain contains an HHCC motif, which binds a Zn^{2+}. Binding of Zn^{2+} to the N-terminal domain enhances multimerization of the native enzyme and increases its catalytic activity (Burke *et al*., 1992; Zheng *et al*., 1996). The N-terminal domain of HIV integrase (residues 1 to 50), when isolated from the remainder of the enzyme, exists as a dimer in solution and its NMR structure has been solved (Cai *et al*., 1997).

The central catalytic domain of integrase contains the DD35E triad motif (Asp64, Asp116 and Glu152). As in other DNA-binding proteins containing this motif, these acidic residues coordinate a divalent Mg^{2+} in the resting enzyme. Substituting any one of these residues abolishes both processing and integration activities of integrase (Bushman *et al*., 1993; Drelich *et al*., 1992; Engelman & Craigie, 1992; Kulkosky *et al*., 1992; Leavitt *et al*., 1993; van Gent *et al*., 1992; Vink *et al*., 1993). The crystal structures of integrase core protein from the human immunodeficiency virus (HIV-1) and the avian sarcoma virus (ASV) have been determined in the absence and presence of bound metal ions and/or inhibitors (Bujacz *et al*., 1996a,b; Dyda *et al*., 1994; Goldgur *et al*., 1998; Lubkowski *et al*., 1998a). The observed structures are similar to those of Mu transposase, Ruv C and RNase H (Dyda *et al*., 1994; Yang & Steitz, 1995; Bujacz *et al*., 1996a; Rice *et al*., 1996); all are proposed to catalyze phosphoryl transfer reactions involving at least one metal ion and the DD35E catalytic triad.

The C-terminal domain of integrase possesses a low degree of sequence homology among integrases of retroviruses. This domain has been suggested to bind DNA and is thus also known as the DNA-binding domain (DBD) (Vink *et al*., 1993; Woerner & Marcus-Sekura, 1993; Engelman *et al*., 1994). Its three-dimensional structure in solution has been characterized with NMR to reveal a dimer with an SH3-like fold (Eijkelenboon *et al*., 1995; Lodi *et al*., 1995).

Although the catalytic core of integrase, by itself, is capable of catalyzing disintegration (the apparent reverse, but non-physiological reaction of strand transfer), only the full-length enzyme is capable of catalyzing 3′ end processing and integration. Neither the N- nor the C-terminal domain possesses catalytic activities. Despite the availability of three-dimensional structures for individual domains of integrase, the quaternary juxtaposition of the domains and their interactions are unknown. The mode of binding of DNA to the native enzyme is also unknown. To elucidate the mechanism of action of integrase, a structure of the enzyme containing more than one domain is needed. We describe here the X-ray structure of the simian immunodeficiency virus (SIV) integrase that contains a contiguous core and DBD, to encompass amino acid residues 50-293 and a site-specific F185H mutation (Li *et al*., 1999). The structure reveals an interface between the two domains covering a surface area of 1500 $Å^2$. Residues from both domains purportedly involved in DNA binding are narrowly distributed on the same face of the two-domain molecule. Based on this information, a model for DNA binding is proposed to tether the 228-236 loop of the DBD and the flexible loop (residues 141-151) of the core domain.

Results and Discussion

The asymmetric unit

There are four core domains in the asymmetric unit and they are labeled in Figure 1 as molecules *A*, *B*, *C*, and *D*†. These molecules can be grouped into two pairs, each contains a non-crystallographic dyad axis. Each pair, *A*-*B* or *C*-*D*, displays similar interface as seen for the core protein dimer of the HIV-1 integrase (Dyda *et al*., 1994; Maignan *et al*., 1998). The two pairs in the asymmetric unit are in turn related to each other by a 90° rotation about and a 23 Å translation along the *c*-axis. The only interactions between the two pairs of core molecules are *via* van der Waals contacts at the 190-loop (between α-helices 5 and 6) of *B* and *C*. Since *B* and *C* are not related by a 2-fold axis, the head segment of this loop (residues 189-192) for each interacts with different areas of the other (thus, *B*-Ile191 is close to *C*-Asn202 while *C*-Ile191 is close to the *B*-Gly82 and *B*-Met154 region). In contrast to the core domain of the HIV-1 integrase, residue 185 of each of the four molecules does not interact with neighboring molecules in adjacent asymmetric units. Also, none of the active sites of the four core domains in the asymmetric unit is within 15 Å of each other, a length suggested to be the critical distance for five nucleotide base-pairs to span between two active sites. The strand transfer reaction necessitates the action of at least two integrase molecules.

There is only one defined DNA binding domain in the asymmetric unit, and it is linked to molecule *A via* a tether and a large surface of van der Waals contacts. The other three DNA binding domains cannot be traced. However, traces of electron density are seen in proximity of the core domains, but they cannot be meaningfully modeled. This finding indicates that the linkage between the core and the

† Residues defined unambiguously by the electron density map for the four core domains in the asymmetric unit are *A*, residues 50-140, 152-213; *B*, 55-140, 152-207; *C*, 55-140, 151-207; *D*, 55-140, 152-207.

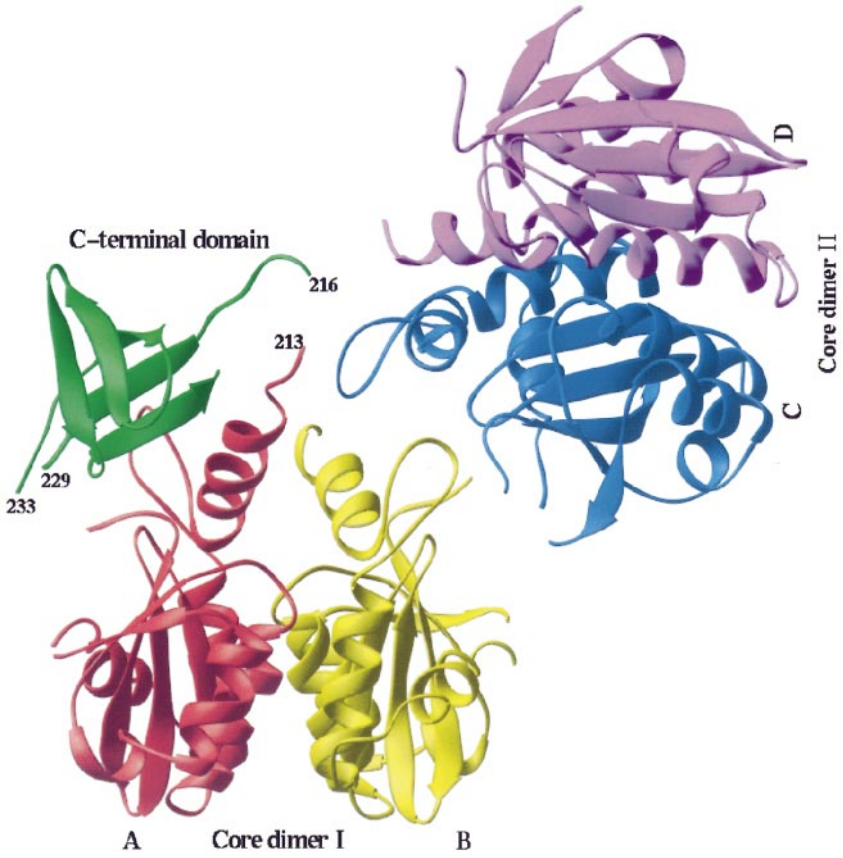

Figure 1. Ribbon plot of the four molecules as seen in the asymmetric unit of the two-domain SIV integrase crystals in the space group of $P2_12_12_1$. The four molecules are labeled as molecule *A* (red), *B* (yellow), *C* (blue), and *D* (purple). The core domain is traced for all four molecules, but only one DBD (green) can be traced and it is part of the polypeptide of molecule *A*. Molecules *A* and *B* are related by a non-crystallographic dyad about the *b*-axis. Molecules *C* and *D* are related by a non-crystallographic dyad about the *a*-axis. A 90° rotation about and 23 Å translation along the *c*-axis relate the dimeric pair *A-B* and *C-D*.

DBD is insufficiently rigid to impose crystal packing to necessitate accommodation of all four in an ordered, symmetry-related manner.

Description of the two-domain structure

The two-domain structure of SIV integrase as depicted by molecule *A* is well defined in the electron density map. The core portion of the structure shows an RNase H fold with six α-helices and five β-strands (Yang *et al.*, 1990) while the DBD displays a SH3-fold with five β-strands (Eijkelenboom *et al.*, 1995). A stereo view of the ribbon model is shown in Figure 2. The helices and strands are labeled α1-α6 and β1-β10, respectively. Other than the segments between residues 50-54 and 141-151, the entire core domain can be traced. Differences among the core domains of the four molecules are restricted mainly to side-chain positions; they are minor and not discussed here.

The two-domain SIV integrase has been crystallized in the absence of sodium cacodylate. As such, its conformation between residues 114-121, which includes the essential active site Asp116, is similar to the HIV-1 integrase core protein crystallized in the absence of cacodylate or the ASV integrase core protein. The previously named ''flexible loop'' (residues 141-151 of the HIV-1 integrase core) remains ill defined in our structure. However, the position and side-chain orientation of the active-site Glu152 can be assigned unambiguously. Despite an effort to improve the map with non-crystallographic symmetry averaging, residues 50-54 and 141-151 were not seen. Continuing towards the DBD, the peptide segment 188-193 (the 190-loop) is very well defined, and it is not a β-strand as seen for the HIV-1 integrase core (Goldgur *et al.*, 1998).

The connection from the core to the DBD is defined by residues 210-220, a stretch that has not been seen in previous structures of either the core or the DBD. This stretch connects the last α-helix, α6, in the core domain and the first β-strand, β6, in the DBD. However, residues 214-215 of this segment are not defined due to poor electron density, further echoing the contention that the tether between the two domains is a flexible one. The first residue of the DBD, Glu213, connects with the last residue of the core, Ala212, to lead an extension of α6. Other than residues 230-232, due to discontinuity in the electron density map, the DBD is well defined, including the two loops encompassing residues 228-236 and residues 244-249. Both are clearly fixed in the crystal structure with the former in the immediate vicinity of the ''flexible loop'' (residues 141-151) of the core domain. The electron density map of the DBD prior to refinement is shown in Figure 3(a) together with its final refined C^{α} model. In Figure 3(b) is shown the final electron density as well as the refined model of the DBD for residues 255-261.

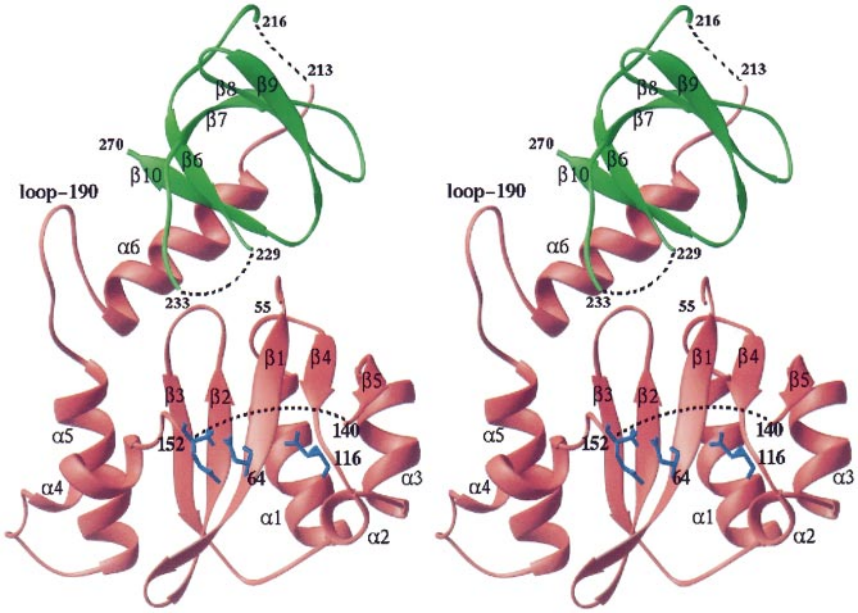

Figure 2. Stereo view of the ribbon representation of molecule *A* in the asymmetric unit of the two-domain SIV integrase crystal. The core domain is colored red and the DBD green. The three active-side residues are colored blue. The flexible loop of the core domain, residues 141-151, is disordered in this structure and is depicted by a dotted line in black between β5 and α4. Residues 214 and 215, as well as 230 to 232, are also not defined by the electron density map and are represented by black dotted lines. The terminal residues 271-293 of the DBD are disordered.

Finally, the N terminus of the two-domain integrase makes a turn at residues 55-57 to point towards the bulk solvent. This orientation of the N terminus is suitable for extension of the core into the N-terminal domain (residues 1-50) of a full length integrase molecule.

The active site triad

The conformation of the side-chains of the DD35E motif of integrase varies significantly in different published structures (Dyda *et al.*, 1994; Bujacz *et al.*, 1996a,b; Goldgur *et al.*, 1998; Lubkowski *et al.*, 1998a, 1999). The spatial arrangement of this triad can be metal, pH, or buffer-dependent. In the case of the SIV integrase, the side-chain orientation of the triad is not identical among the four core molecules within the asymmetric unit. However, all are well defined; the electron density and model for one of the molecules is illustrated in Figure 4(a).

As Glu152 is located at the end of the flexible 141-151 loop, its position is the most varied one of the active-side triad. Depending on the conformation of the 141-151 loop, Glu152 can change from facing in the general direction towards or pointing away from Asp64 and Asp116, as seen previously for two different HIV-1 integrase core structures (Goldgur *et al.*, 1998). When the HIV-1 integrase core domain is crystallized in the presence of sodium cacodylate, Glu152 is completely disordered (Dyda *et al.*, 1994). In the case of the four core domains seen in the SIV integrase crystal, Glu152 is fixed in each but its side-chain either points toward or away from Asp116 and Asp64.

Asp116 is always fixed in all integrase core structures. However, the peptide segment containing residues 114-121 varies significantly among different crystal structures and the C^{α} position of Asp116 can vary by as far as 3 Å (Goldgur *et al.*, 1998). The conformation of the 114-121 segment of the SIV integrase is similar to that of the ASV core protein as well as the HIV-1 core protein crystallized in the absence of sodium cacodylate.

The position of the C^{α} atom of Asp64 is almost always fixed in the active site of the integrase core structures due to a well defined stretch of residues on which it resides, but the orientation of its side-chain differs among various integrase structures. In the ASV integrase core, Asp64 forms water-mediated hydrogen bonds with Asp116 and with Gln153 (Lubkowski *et al.*, 1998b). In the HIV-1 integrase core, Asp64 forms hydrogen bonds with both the main-chain (-NH) of Asp116 and the side-chain (-NH_2) of Gln62 (Goldgur *et al.*, 1998). In the SIV structure, Asp64 is again oriented differently. It points towards the side-chain of Glu152 in molecules *A* and *B* of the *A-B* dimer but not in molecules *C* and *D* in the *C-D* dimer.

The flexibility of the active-site triad could be due to the absence of DNA and/or the low pH of the integrase crystals. There is no strong electron density in the active site of the SIV integrase map attributable to a metal ion. The variability of the side-chains (Asp64, Glu152) as well as main-chain (Asp116) of the catalytic triad, as shown in Figure 4(b), could in part be due to the absence of a firmly coordinated metal ion at low pH, a condition under which the SIV integrase single crystals are cultivated (see Materials and Methods).

Interactions between the core and DBD

The major interactions between the core and DBD are hydrophobic in nature. There are three hydrogen bonds buried within the interface. The extended interactions yield an interface of 1500 $Å^2$

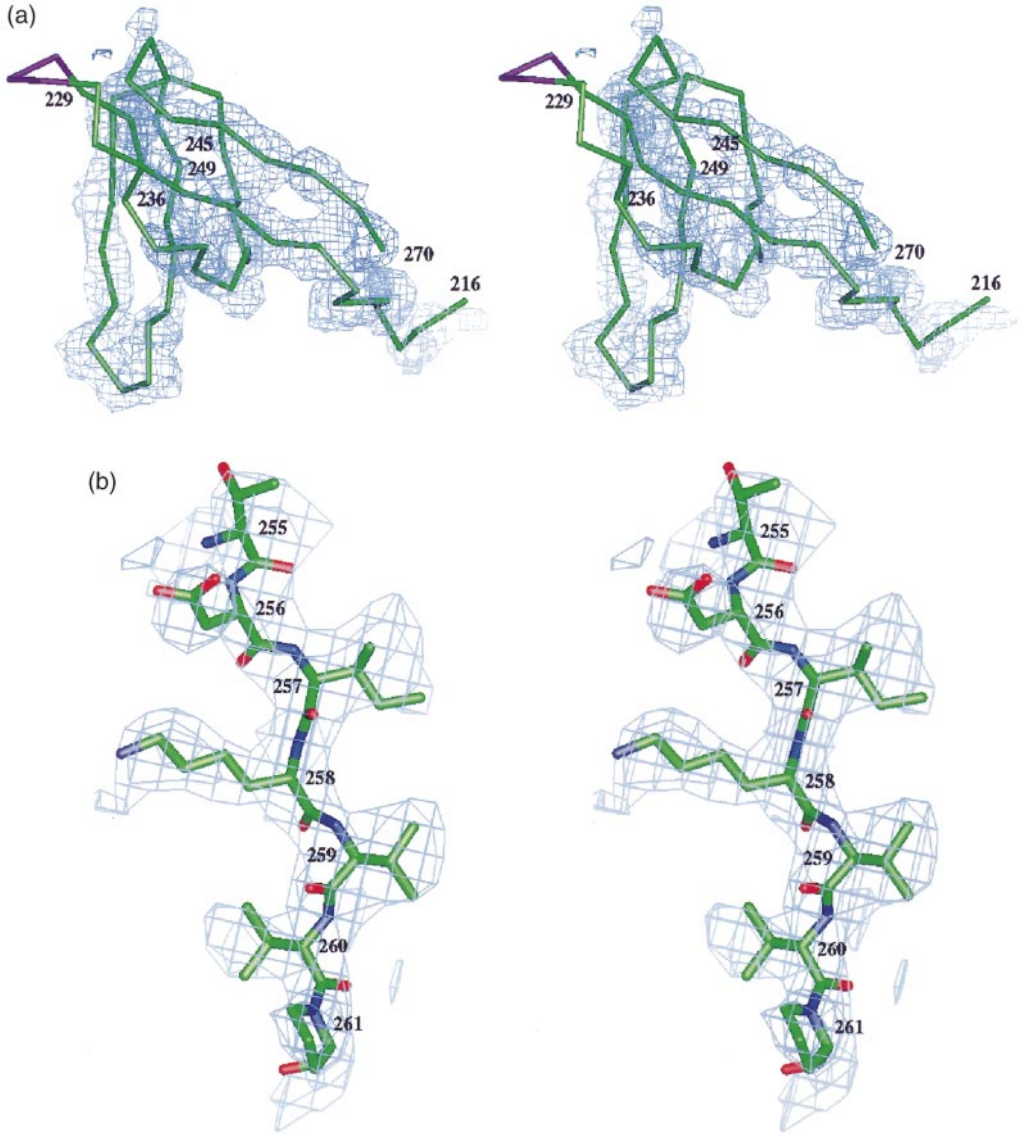

Figure 3. Stereo views of electron density of the DBD superimposed upon its refined coordinates. (a) The electron density map shown is that calculated using the coefficients $2F_o - F_c$ and the combined phases from the model containing four core domains and the samarium derivative. Map contoured at 1.5σ is superimposed upon the C^α trace of the refined model of the DBD. Density for the loop regions 229-236 and 245-249 are poorly defined in the map and not included in the initial refinement. Following a few cycles of refinement and model building, density for all the residues excepting 230-232 is well defined. The observed SH3 fold of the DBD monomer is similar to that revealed by NMR. (b) Map contoured at 1.0σ, calculated using the coefficients $2F_o - F_c$ and the phases from the final refined model. The segment of residues 255-261 of the DBD are superimposed upon the density.

that is devoid of water molecules and encompasses the N-terminal residues 55-57, the hairpin turn between β2 and β3, 190-loop (the peptide segment between α5 and α6), α6, β6, as well as β10. This interface is displayed in a surface contour model in Figure 5, and the residues participating in interface interactions are listed in Table 1.

The last strand in the DBD, β10, is the most involved in interface interactions. It is inserted into a V-shaped region provided by the 190-loop and α6 (see Figure 2) to form a ridge that ''dovetails'' the two domains (see Figure 5). The N-terminal half of β10 (residues 262-265) lies within van der Waals distances from the N terminus (residues 55-58) of the core domain. The C-terminal half of β10 (266-269) is in van der Waals contact with a stretch of side-chains in the core domain (residues 199-209). In addition, β10 also contributes all three hydrogen bonds found in the interface. Flanking β10 and providing additional inter-domain interactions is residue 223 of β6 (see Table 1).

The DBD structure in solution as determined with use of NMR (Eijkelenboon *et al.*, 1995; Lodi *et al.*, 1995) has revealed that it is a well-folded domain. Our finding that only one out of four DBD is observable suggests that its quaternary relationship with the core domain is a flexible one. However, we cannot rule out the possibility that

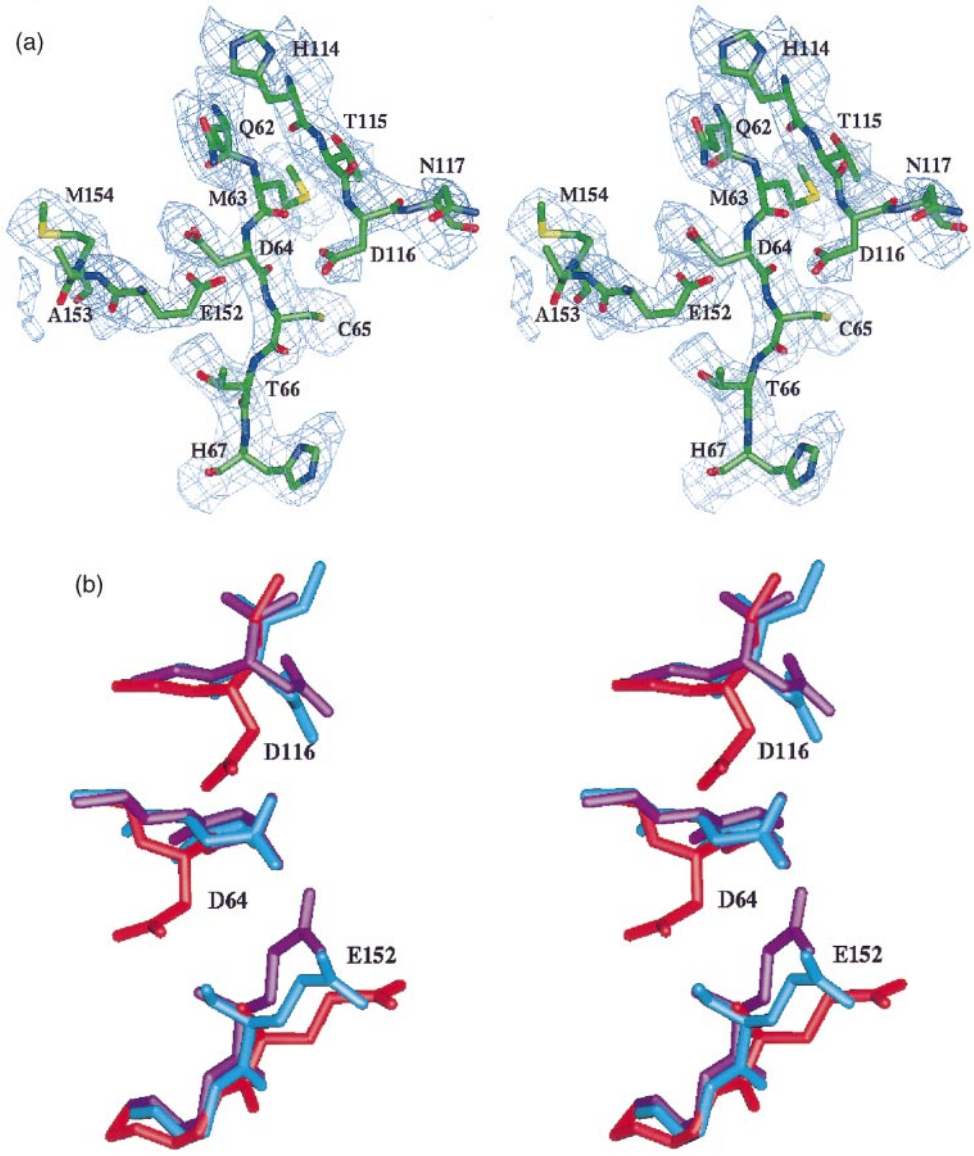

Figure 4. (a) Stereo views of electron density for the catalytic triad (Asp64, Asp116, Glu152) for one of the four core domains. The map used is that contoured at 1.0σ and calculated with coefficients $2F_o - F_c$ and the phases from the final refined model. (b) Superposition of the active-site triad (Asp64, Asp116, Glu152) in the crystal structures of SIV (red), ASV (purple), and HIV (blue) integrase (see the text for discussion and references).

crystal packing might have played a role in the ordering of the DBD. As shown by a crystal packing diagram in Figure 6, some crystal contacts can be observed between one part of the DBD (labeled) and its closest neighbor in an adjacent asymmetric unit. Nonetheless, other information, such as the extensive domain-domain interaction surface (see Figure 5) as well as the DNA binding residues of the DBD appearing on the same face of the two-domain structure as the active-site residues of the core (see below), suggests that the quaternary relationship of the two domains as depicted here represents at least one of the biologically relevant conformations.

Model for viral DNA-integrase interactions

The C-terminal domain of integrase has been shown to bind DNA in a non-specific manner (Puras Lutzke *et al*., 1994; Engleman *et al*., 1994). However, this interaction may be essential in providing an anchor for proper positioning of the 3′ viral DNA end that is close to the active site residues in the core domain. To date, no structural information is available for a DNA bound integrase. In order to provide some understanding of the binding mode of viral DNA to integrase, a model is proposed here. Residues Arg263, Lys264, and Arg 231 in the C-terminal domain have been implicated in viral DNA binding (Puras Lutzke &

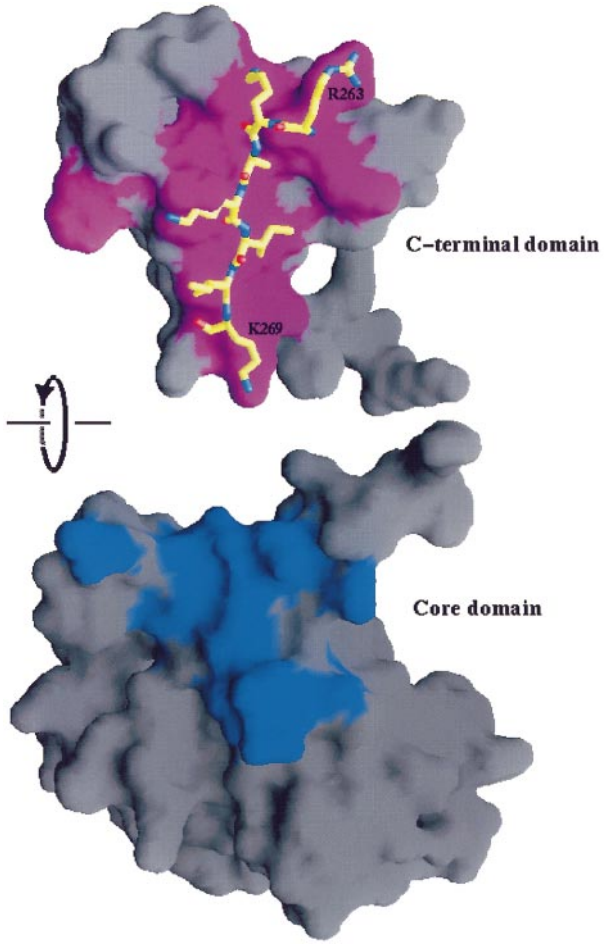

Figure 5. Surface areas that contribute to the interface between the core and the C-terminal domains (DBD) in the crystal structure of SIV integrase. To expose the interface for viewing, the core and the DBD of molecule *A* (see Figures 1 and 2) are graphically pulled away from each other and then rotated by 90°, as indicated by an arrow about the rotation axis. The residues in the interface are colored blue in the core and purple in the DBD. Residues from 263 to 269 are depicted on the DBD moiety to show where β10 is located. With the exception of three hydrogen bonds, interactions in the interface are hydrophobic in nature. The total solvent-exclusion interface is 1500 Å^2.

Table 1. Amino acid residues in the interface of the core and the C-terminal domains (DBD) of SIV integrase

A. *van der Waals contacts (within a 4.2 Å radius)*

Residues of DBD	Residues of core domain
Phe223	Thr206, Glu209
Arg262	Glu207
Arg263	Asn55, Ser56, Asp57
Ala265	Ser56
Lys266	Ala80
Ile267	Met203, Thr206
Ile268	Arg199, Asn202, Met203
Lys269	Asn202, Thr206, Glu209

B. *Hydrogen bonds*

Residues of DBD	Residues of core domain	Distance (Å)
Arg263(O)	Ser56(N)	2.81
Lys269(N)	Asn202($O^{\delta1}$)	3.14
Lys269(O)	Asn202($N^{\delta2}$)	3.46

Plasterk, 1998; Lodi *et al.*, 1995). As shown in Figure 7(a), all of these residues and the active site appear on the same face of the two-domain SIV integrase structure. This interesting finding not only offers an understanding of the role of DBD in integrase function but also renders it feasible to propose a model for viral DNA binding, as shown in Figure 7(b). Coordinates for the model of the dumbbell DNA substrate developed by Heuer & Brown (1998) are employed here for modeling a bound viral DNA. The duplex DNA in which the three base-pairs are frayed at the 3′ end has been docked to the protein such that the conserved adenosine base at the 3′ end is proximal to the active site Glu152. This residue is required for recognition of N7 of the adenine base three nucleotides away from the viral 3′ end (Gerton & Brown, 1997). The viral DNA model has been optimized by maximizing the number of contacts between the DNA and the residues of core and DBD, implicated in DNA binding. The model brings the flexible loop 141-151 close to the 3′ end of the DNA. Residues Asn55, Gln62, and His114 of the core also appear in proximity of the DNA. However, the side-chain of Lys159, while implicated in DNA binding (Jenkins *et al.*, 1996) and in the vicinity of 3′ end, does not directly interact with the DNA in our model because its side-chain points away from the modeled DNA.

Residues Arg263 and Lys264 of the DBD interact with the phosphodiester backbone of the modeled DNA at five to six base-pairs from the 3′ end adenosine. Similarly, the loop structure composed of the residues 230-234 in the DBD is in proximity of the DNA backbone. Arg231 is conserved among retroviruses and has been implicated to bind to DNA (Puras Lutzke & Plasterk, 1998). Similarly, Lys234 implicated in DNA binding is also in close proximity of the modeled DNA. However, Ser230 in the HIV-1 integrase, suggested to bind DNA (Heuer & Brown, 1998), is Gly230 in the SIV integrase and does not interact with the modeled DNA.

In HIV-1 integrase, Arg262 has been suggested to bind DNA because replacing this residue with a Gly leads to a 50% decrease in DNA binding and in 3′ end processing (Puras Lutzke & Plasterk, 1998). However, in the crystal structure the side-chain of this residue points away from the proposed DNA binding site. Similarly, Val260 in the HIV-1 integrase has been implicated in DNA binding but does not interact directly with the DNA in our model. It is conceivable that upon DNA binding in solution, both the protein and the DNA

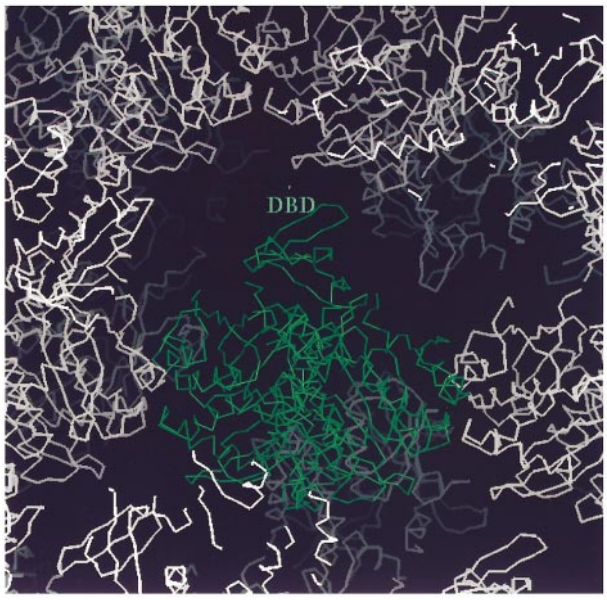

Figure 6. A crystal packing diagram showing the spatial relationship between asymmetric units. Each asymmetric unit contains four core domains plus one DBD domain (labeled). The C^{α}-C^{α} distance between the DBD domain and its nearest neighbor is 8 Å.

undergo conformational adjustments, bringing these amino acids closer to the DNA.

Modeling viral DNA-integrase interactions in the SIV integrase structure is feasible due to the proximity of the core and the DBD. The majority of the amino acid residues in the core and the DBD in contact with DNA spans about five to six base-pairs of the duplex DNA in this model. The model is consistent with the photo cross-linking studies identifying the proximity of different integrase residues to the viral DNA (Heuer & Brown, 1998). The precise mode of interactions between the viral DNA and the functional form of integrase would depend upon the oligomeric state of integrase. However, it is possible that the two ends of the viral DNA interact with the protein in a similar fashion, bridging two integrase dimers. This minimal model for interactions between the viral DNA and the SIV integrase is different from that proposed by Heuer & Brown (1998). In their model, the relative orientation of the DBD with respect to the core domain is different from that observed in the SIV integrase crystal structure. As a result, the length of DNA that spans the protein-binding region is different.

A dimeric SIV integrase

The isolated core domain appears as a dimer in all the crystal structures determined to date for the HIV or ASV integrase. In solution, both the isolated N and C-terminal domains are also dimers, as revealed by NMR spectroscopy (Eijkelenboom *et al.*, 1995; Lodi *et al.*, 1995; Cai *et al.*, 1997). Together, these results suggest that the dyad symmetry may exist throughout the 32 kDa, full-length integrase protein.

In the two-domain integrase crystal structure determined here, only one DBD can be traced for molecule *A* of the *AB* dimer. The symmetry-related DBD to be associated with molecule *B*, when generated with the dyad axis of the *AB* core, results in two DBD units that *do not* interact with each other, as shown in Figure 8. As a result, the DBD as defined in our structure cannot exist as a dimer. In addition, the 2-fold generated DBD of molecule *B* would intrude into the space occupied by the core of molecule *C* and is therefore incompatible with the molecular packing in the asymmetric unit of the crystal. The same would be true for the *CD* dimer, in which case the DBD of molecule C would intrude into the space occupied by the core of molecule *B*. If one were to superposition one monomeric unit of the NMR DBD dimer onto the DBD of molecule *A* in the SIV integrase asymmetric unit, the second monomer of the NMR DBD dimer would be far away from, and would not interact with, the core domain of molecule *B* (not shown). This is inconsistent with the fact that the DBD is covalently linked to the core domain. Thus, the dimeric structure of the single-domain DBD in solution is incompatible with the crystal structure of DBD in the two-domain SIV integrase. The DBD structure as observed with NMR spectroscopy may exist only in the absence of a linked core domain.

Summary

We have reported here the two-domain structure of the SIV integrase that encompasses residues 50-293 in a single polypeptide chain. This structure reveals for the first time the spatial relationship between the core and the C-terminal domains of retroviral integrase to allow modeling of a bound viral DNA. The quaternary hierarchy shown by this structure should afford new experimental designs to not only verify its validity in solution but also probe the inter-molecular relationship between subunits of the integrase oligomer involved in the catalytic reactions of viral-host DNA integration.

Materials and Methods

Protein purification and crystallization

Details of purification and crystallization of the two domain SIV integrase have been documented in another paper (Li *et al.*, 1999). Briefly, the C-terminal two-thirds segment of integrase derived from SIV was cloned, expressed in *Escherichia coli*, and purified to homogeneity from the soluble fraction with significant modifications of the published protocols for the HIV-1 and SIV integrase (Du *et al.*, 1997; Jenkins *et al.*, 1996). A three-step His-trap, benamidine, and SP Sepharose chromatography was employed. The resultant enzyme was >95%

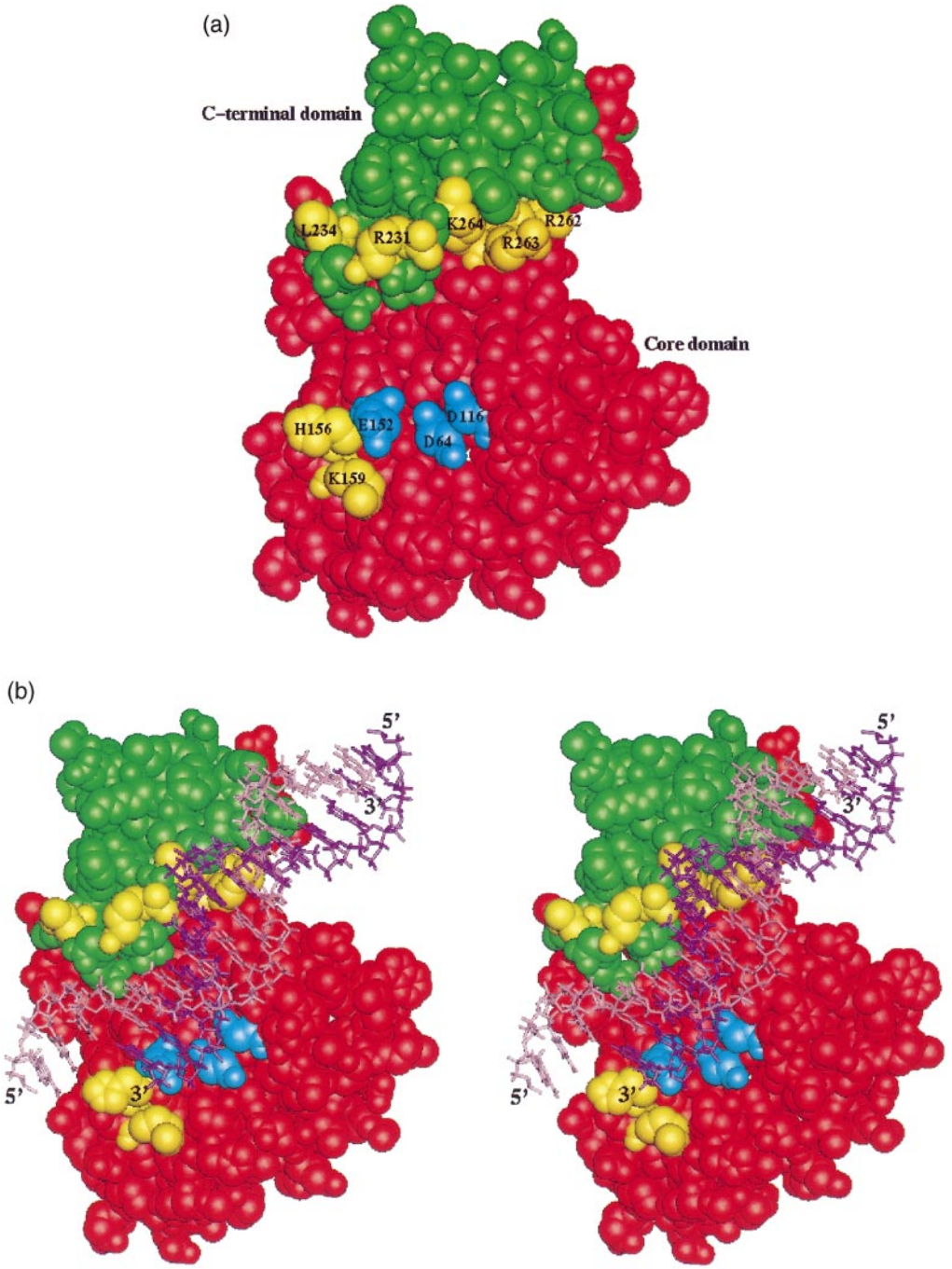

Figure 7. (a) Space-filling diagram of the two-domain SIV integrase showing that residues suggested to be involved in DNA binding are found on the same face of the enzyme molecule. Residues Arg231, Arg262, Arg263 and Lys264 (gold) that form a positively charged patch on the DBD is 15 Å away from the active-site triad (blue) and intersected by the flexible loop of the core domain (residues 141-150); (b) A molecular graphics DNA binding model. In this model, a viral DNA is built onto the crystal structure of the SIV integrase to cover binding regions of both the core domain and DBD that have been suggested to be involved in polynucleotide binding. The core domain and DBD are colored in red and green, respectively. The two strands of the DNA molecule are colored in pink and purple. The 3′ end of the DNA strand (purple) that contains the terminal CA dinucleotide is positioned adjacent to the active site. As depicted, the carboxylate oxygen of Glu152 is allowed to interact directly with N7 of adenosine. See the text for details.

pure and gave only a single N-terminal sequence beginning with GSHMIHGQVN.... The enzyme was crystallized in 0.1 M Mes (pH 5.7), 8% (v/v) PEG 6000 and 1.5% (v/v) dioxane. A quantity of 0.1 M $MgCl_2$ was optional, but the presence of microscopic crystalline seeds was a must to bring about growth of crystals usable for data collection.

The SIV integrase crystals contain multiple lattices in most instances and are very sensitive to cryo protective solvents. Therefore, data collection with the SIV crystals

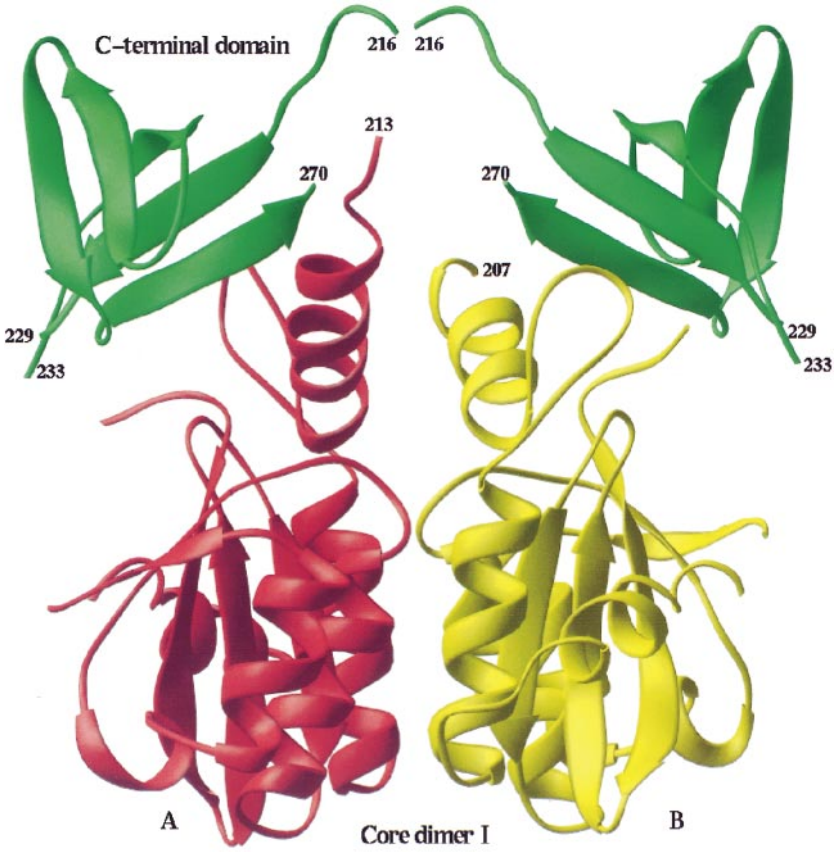

Figure 8. A dimeric model of the two-domain SIV integrase. The DBD of molecule *A* together with the core domain as traced from the electron density map is given on the left-hand side of this Figure. The DBD, generated with the non-crystallographic dyad of dimer *A-B*, is shown on the right-hand side with the core of molecule *B*. This ribbon diagram illustrates that if the second DBD in dimer *A-B* is related by the same dyad as governed by their core domains, there should be no interactions between the two DBD.

has been difficult. With single crystals that show low mosaicity, diffraction can reach 2.5 Å resolution. The cell constants are $a = 79.57$ Å, $b = 100.0$ Å, $c = 150.5$ Å in the space group $P2_12_12_1$. The solvent content is 54% and there are four molecules in an asymmetric unit.

Diffraction data collection

A native data set, at 3.2 Å resolution, and a heavy-atom derivative data set, at 3.5 Å resolution, were collected at −166 °C on an Raxis II imaging plate system equipped with a Rigaku Ru200 rotating anode generator operated at 50 kV and 100 mA. The X-ray beam was monochromatized with a Yale double mirror focusing system. A 3.0 Å resolution native data set was collected at the 0.9871 Å wavelength at −170 °C on the Industrial Macromolecular Crystallography Association (IMCA) beam line at the Advanced Photon Source (APS) of Argonne National Laboratories. This data set was used as the native set for refinement. All diffraction data were integrated with DENZO and scaled with the SCALEPACK program package (Otwinowski, 1993). The data collection statistics are given in Table 2.

Structure determination

Both monomer and dimer coordinates of the HIV-1 integrase core protein were used as the initial search models. The AMoRe program of CCP4 package (Navaza, 1994; CCP4, 1994) was used in the molecular replacement search. A self-rotation function revealed that there were maxima at the 90° and 180° about an axis parallel to the *c*-axis, suggesting that the four molecules could be arranged as two dimers related by a 90° rotation. Cross-rotation function calculations were successful only using the dimer model of HIV-1 integrase core protein (Dyda *et al.*, 1994). There were two peaks in the cross-rotation function with correlation coefficients of 0.372 and 0.358. These two peaks were related with a 90° rotation about the *c*-axis that is consistent with the result obtained from the self-rotation function. The translation function search based on the cross-rotation peaks resulted in the location of two peaks with correlation coefficients and *R*-factors of 0.278 and 0.562 for the first peak and of 0.365 and 0.535 for the second, respectively. Applying the rotation and translation matrices, the model based on molecular replacement solution was further refined *via* rigid body refinement to give a correlation coefficient of 0.425 and an *R*-factor of 0.484. Cross-rotational searches using dimer or monomer coordinates from the DBD model alone, or in conjunction with the coordinates of the core model, did not reveal any meaningful peaks.

Samarium acetate that was reported to give an isomorphous heavy-atom derivative for the HIV-1 integrase core domain (Dyde *et al.*, 1994) was found to be suitable as a heavy-atom to yield an isomorphous derivative for the two-domain SIV integrase crystals. The cell constants of the heavy-atom derivative were $a = 79.29$ Å, $b = 97.03$ Å, and $c = 150.7$ Å. Although there was a 3% difference in the *b*-axis length, the data set could be scaled with the native data set with reasonable statistics, indicating that there was no significant lack of isomorphism. Using the difference Fourier method based on the phases of the four core domains, four samarium sites were found in the asymmetric unit, each close to Asp116 in the four core domains. Electron density map was cal-

Table 2. Data collection and refinement statistics

A. *Data collection* Data set (wavelength)	Resolution (Å)	Detector source	Total/ unique	Completeness (%)	R_{sym}	R_{cullis}	Phasing power
Native 1 (1.5418 Å)	3.2	*R*-axisII RU200	103,305/ 16,892	78.1	0.097	Not applicable	Not applicable
Native 2 (0.9871 Å)	3.0	*Mar*-detector IMCA	309,679/ 22,127	92.5	0.112	Not applicable	Not applicable
Samarium acetate (1.5418 Å)	3.5	*R*-axisII RU200	147,339/ 12,399	81.5	0.112	0.66	0.79
B. *Refinement*							
Resolution (Å)	6.0-3.0						
Non-hydrogen atoms	4960						
H_2O molecules	73						
R-value	0.203						
R_{free}	0.362						
rms bond length (Å)	0.009						
rms bond angle (deg.)	2.1						

$R_{sym} = \Sigma|I - \langle I\rangle|/\Sigma\langle I\rangle$; $R_{Cullis} = \Sigma||FPH_o| + |FP_o|| - |FH_o||/\Sigma||FPH_o| + |FP_o||$ for centric reflections; $R\text{-value} = |FP_o - FP_c|/\Sigma FP_o$; R_{free} is computed on a randomly chosen 10% of the data excluded from the refinement; FP, protein structure factor; FPH, derivative structure factor; FH, heavy-atom structure factor. The isomorphous phasing power is defined as FH_c/E, where E denotes the rms lack of closure.

culated by first combining the phases obtained separately from molecular replacement and the single heavy-atom derivative, followed with solvent flattening (CCP4, 1994). In addition to the four core domains the calculated map revealed additional electron density which was assigned to the DBD. Molecular averaging procedure was applied to the four core domains.

Model building and refinement

Model building was performed on a Silicon Graphics system with use of the program CHAIN (Sack, 1988). HIV-1 integrase side-chains in the core domains were replaced with the SIV side-chains. An initial model with four core domains and one DBD domain was then refined. A single rigid-body refinement cycle and an energy minimization refinement cycle with use of X-PLOR reduced the *R*-value from 0.491 to 0.325 (Brünger, 1992). Several cycles of X-PLOR refinement and model building employing CHAIN were performed. Seventy three water molecules were introduced in a judicious manner into density peaks greater than 2.5σ in the difference Fourier map to yield reasonable hydrogen bonding interactions. Two additional cycles of refinement were conducted, and no significant electron density was seen to account for the three absent DBD. The final *R*-value at 3.0 Å is 0.203 and the R_{free} value is 0.362 (calculated with 10% of the data excluded from the refinement). Given a final *R*-value of 0.203, the value for R_{free} is higher than anticipated. One possible reason for this result is the lack of atom assignments to the electron densities contributed by the three non-traceable DBD of molecules *B*, *C*, and *D*. These three DNA binding domains could exist with low occupancy in multiple positions, thus leading to a high R_{free} as well as the large difference between R_{free} and the *R*-value. The rms deviation from the ideal bond length and angle were 0.009 Å and 2.1°, respectively. Anisotropic temperature refinement of diffraction data set Native 2 (Table 2) did not improve R_{free}. The model included 4960 non-hydrogen atoms and 73 water molecules. Refinement statistics are given in Table 2. It should be made clear that a strong attempt has been made to improve the current phases as much as is technically possible *via* the available phase refinement programs for several data sets without success. Better refined phases must await a higher resolution data set and/or with the availability of better single crystals, both are currently being pursued. To ensure that the crystalline protein has not been degraded during storage or crystallization and that we are dealing with a single protein species, we have re-dissolved a single crystal of the two-domain SIV integrase that has undergone the same manipulation and storage processes, to show with SDS-polyacrylamide gel electrophoresis and protein sequencing a single, expected N terminus (see Li *et al*. (1999) for discussion of SIV integrase N terminus sequence). The programs GRASP (Nicholls *et al*., 1991), O (Jones *et al*., 1991), QUANTA (MSI), and Ribbons (Carson, 1991) were used in preparing the Figures.

Data Bank accession numbers

The X-ray coordinates of the SIV integrase can be found in the RCSB Protein Data Bank under the code: 1C6V (to be added subsequent to assignment).

Acknowledgments

The diffraction data set at 0.9871 Å wavelength was collected at beam line 17-ID in the facilities of the Industrial Macromolecular Crystallography Association Collaborative Access Team at the Advanced Photon Source. These facilities are supported by the companies of the Industrial Macromolecular Crystallography Association through a contract with Illinois Institute of Technology, executed through its Center for Synchrotron Radiation Research and Instrumentation. Use of the Advanced Photon Source was supported by the US Department of Energy, Basic Energy Sciences, Office of Energy Research, under Contract number W-31-109-Eng-38. We thank Dr John Chrzas and the staff at beam line 17-ID for their help during our data collection and data processing at the APS. The results reported here have been presented at the American Crystallographic Association Annual Meeting in Buffalo, New York, May 22-24, 1999 and at the Cold Spring Harbor Meeting on Retroviruses, New York, May 25-30, 1999.

References

Brünger, A. T. (1992). *X-PLOR Version 3.1 System for Crystallography and NMR*, Yale University Press, New Haven, CT.

Bujacz, G., Alexandratos, J., Qing, Z. L., Clement-Mella, C. & Wlodawer, A. (1996a). The catalytic domain of human immunodeficiency virus integrase: ordered active site in the F185H mutant. *FEBS Letters,* **398**, 175-178.

Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A. & Skalka, A. M. (1996b). The catalytic domain of avian sarcoma virus integrase: conformation of the active site residues in the presence of divalent cations. *Structure,* **4**, 89-96.

Burke, C. J., Sanyal, G., Bruner, M. W., Ryan, J. A., LaFemina, R. L., Robbins, H. L., Zeft, A. S., Middaugh, C. R. & Cordingley, M. G. (1992). Structural implications of spectroscopic characterization of a putative zinc finger peptide from HIV-1 integrase. *J. Biol. Chem.* **267**, 9639-9644.

Bushman, F. D., Engelman, A., Palmer, I., Wingfield, P. & Craigie, R. (1993). Domains of the integrase protein of human immunodeficiency virus type 1 responsible for polynucleotidyl transfer and zinc binding. *Proc. Natl Acad. Sci. USA,* **90**, 3428-3432.

Cai, M., Zheng, R., Caffrey, M., Craigie, R., Clore, G. M. & Gronenborn, A. M. (1997). Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nature Struct. Biol.* **4**, 567-577.

Carson, M. (1991). Ribbons 2.0. *J. Appl. Crystallog.* **24**, 958-961.

CCP4 (1994). Collaborative Computational Project No. 4. *Acta. Crystallog. sect. D,* **50**, 760-763.

Drelich, M., Wilhelm, R. & Mous, J. (1992). Identification of amino acid residues critical for endonuclease and integration activities if HIV-1 IN protein *in vitro*. *Virology,* **188**, 459-468.

Du, Z., Ilyinskii, P. O., Lally, K., Desrosiers, R. C. & Engelman, A. (1997). A mutation in integrase can compensate for mutations in the simian immunodeficiency virus attachment site. *J. Virol.* **71**, 8124-8132.

Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R. & Davies, D. R. (1994). Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science,* **266**, 1981-1986.

Eijkelenboom, A. P., Lutzke, R. A., Boelens, R., Plasterk, R. H., Kaptein, R. & Hard, K. (1995). The DNA-binding domain of HIV-1 integrase has an SH3-like fold. *Nature Struct. Biol.* **2**, 807-810.

Engelman, A. & Craigie, R. (1992). Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function *in vitro*. *J. Virol.* **66**, 6361-6369.

Engelman, A., Hickman, A. B. & Craigie, R. (1994). The core and carboxyl-terminal domains of the integrase protein of human immunodeficiency virus type 1 each contribute to non-specific DNA binding. *J. Virol.* **68**, 5911-5917.

Gerton, J. & Brown, P. O. (1997). The core domain of HIV-1 integrase recognizes key features of its DNA substrate. *J. Biol. Chem.* **272**, 25809-25815.

Goldgur, Y., Dyda, F., Hickman, A. B., Jenkins, T. M., Craigie, R. & Davies, D. R. (1998). Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium. *Proc. Natl Acad. Sci. USA,* **95**, 9150-9154.

Heuer, T. S. & Brown, P. O. (1998). Photo-cross-linking studies suggest a model for the architecture of an active human immunodeficiency virus type 1 integrase-DNA complex. *Biochemistry,* **37**, 6667-6678.

Jenkins, T. M., Engelman, A., Ghirlando, R. & Craigie, R. (1996). A soluble active mutant of HIV-1 integrase: involvement of both the core and carboxyl-terminal domains in multimerization. *J. Biol. Chem.* **271**, 7712-7718.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and location of errors in these models. *Acta Crystallog. sect. A,* **97**, 110-119.

Katz, R. & Skalka, A. (1994). The retroviral enzymes. *Annu. Rev. Biochem.* **63**, 133-163.

Kulkosky, J., Jones, K. S., Katz, R. A., Mack, J. P. & Skalka, A. M. (1992). Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Mol. Cell Biol.* **12**, 2331-2338.

Leavitt, A. D., Shiue, L. & Varmus, H. E. (1993). Site-directed mutagenisis of HIV-1 integrase demonstrates differential effects on integrase functions *in vitro*. *J. Biol. Chem.* **268**, 2113-2119.

Li, Y., Yan, Y., Zugay-Murphy, J., Xu, B., Cole, J. L., Witmer, M., Felock, P., Wolfe, A., Hazuda, D., Sardana, M. K., Chen, Z., Kuo, L. C. & Sardana, V. (1999). Purification, solution properties, and crystallization of SIV integrase containing a continuous core and C terminus domain. *Acta. Crystallog. sect. D,* **55**, 1906-1910.

Lodi, P. J., Ernst, J. A., Kuszewski, J., Hickman, A. B., Engelman, A., Craigie, R., Clore, G. M. & Gronenborn, A. M. (1995). Solution structure of the DNA binding domain of HIV-1 integrase. *Biochemistry,* **34**, 9826-9833.

Lubkowski, J., Yang, F., Alexandratos, J., Wlodawer, A., Zhao, H., Burke, T. R., Jr, Neamati, N., Pommier, Y., Merkel, G. & Skalka, A. M. (1998a). Structure of the catalytic domain of avian sarcoma virus integrase with a bound HIV-1 integrase inhibitor. *Proc. Natl Acad. Sci. USA,* **95**, 4831-4836.

Lubkowski, J., Yang, F., Alexandratos, J., Merkel, G., Katz, R. A., Gravuer, K., Skalka, A. M. & Wlodawer, A. (1998b). Structural basis for inactivating mutations and pH-dependent activity of ASV integrase. *J. Biol. Chem.* **273**, 32685-32689.

Lubkowski, J., Dauter, Z., Yang, F., Alexandratos, J., Merkel, G., Skalka, A. M. & Wlodawer, A. (1999). Atomic resolution structures of the core domain of avian sarcoma virus integrase and its D64N mutant. *Biochemistry,* **38**, 13512-13522.

Maignan, S., Guilloteau, J., Zhou-Liu, Q., Clément-Mella, C. & Mikol, V. (1998). Crystal structures of the catalytic domain of HIV-1 Integrase free and complexed with its metal cofactor: high level of similarity of the active site with other viral integrases. *J. Mol. Biol.* **282**, 359-368.

Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallog. sect. A,* **50**, 157-163.

Nicholls, A., Sharp, K. A. & Honig, B. H. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281-286.

Otwinowski, Z. (1993). In *Data Collection and Processing* (Sawyer, L. N., Issacs, N. W. & Bailey, S., eds), pp.

55-62 SERC Daresbury Laboratory, Warrington, UK.

Puras Lutzke, R. A. & Plasterk, R. H. A. (1998). Structure-based mutational analysis of the C-terminal DNA-binding domain of human immunodeficiency virus type 1 integrase: critical residues for protein oligomerization and DNA binding. *J. Virol.* **72**, 4841-4848.

Puras Lutzke, R. A., Vink, C. & Plasterk, R. H. A. (1994). Characterization of the minimal DNA binding domain of the HIV integrase protein. *Nucl. Acid Res.* **22**, 4125-4131.

Rice, P., Craigie, R. & Davies, D. R. (1996). Retroviral integrases and their cousins. *Curr. Opin. Struct. Biol.* **6**, 76-83.

Sack, J. S. (1988). CHAIN: a crystallographic modeling program. *Mol. Graph.* **6**, 224-225.

van Gent, D. C., Oude, Groeneger A. A. M. & Plasterk, R. H. A. (1992). Mutational analysis of the integrase protein of human immunodeficiency virus type 2. *Proc. Natl Acad. Sci. USA,* **89**, 9598-9602.

Vink, C. & Plasterk, R. H. (1993). The human immunodeficiency virus integrase protein. *Trends Genet.* **9**, 433-438.

Vink, C., Oude, Groeneger A. M. & Plasterk, R. H. (1993). Identification of the catalytic and DNA-binding region of the human immunodeficiency virus type I integrase protein. *Nucl. Acids Res.* **21**, 1419-1425.

Woerner, A. M. & Marcus-Sekura, C. J. (1993). Characterization of a DNA binding domain in the C terminus of HIV-1 integrase by deletion mutagenesis. *Nucl. Acids Res.* **21**, 3507-3511.

Yang, W. & Steitz, T. A. (1995). Recombining the structures of HIV integrase, RuvC and RNase H. *Structure,* **3**, 131-134.

Yang, W., Hendrickson, W. A., Crouch, R. J. & SatowY., (1990). Structure of ribonuclease H phased at 2 Å resolution by MAD analysis of the seleno methionyl protein. *Science,* **249**, 1398-1405.

Zheng, R., Jenkins, T. M. & Craigie, R. (1996). Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization and enhances catalytic activity. *Proc. Natl Acad. Sci. USA,* **93**, 13659-13664.

Edited by I. A. Wilson

(Received 5 August 1999; received in revised form 6 December 1999; accepted 6 December 1999)