4

STRUCTURAL STUDIES OF RETROVIRAL INTEGRASES

MARIUSZ JASKOLSKI

Department of Crystallography, Faculty of Chemistry, A. Mickiewicz University, Poznan, Poland and
Center for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland

JERRY N. ALEXANDRATOS

Macromolecular Crystallography Laboratory, National Cancer Institute at Frederick, Frederick, Maryland, USA

GRZEGORZ BUJACZ

Institute of Technical Biochemistry, Technical University of Lodz, Poland and
Center for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland

ALEXANDER WLODAWER

Macromolecular Crystallography Laboratory, National Cancer Institute at Frederick, Frederick, Maryland, USA

4.1 Introduction
4.2 Amino Acid Sequence and Domain Structure of Retroviral Integrases
4.3 Crystallization of Integrase
4.4 Catalytic Domain of Integrase
4.5 N-Terminal Domain of Integrase
4.6 C-Terminal Domain of Integrase
4.7 Two-Domain Constructs Consisting of N and Cat Domains
4.8 Two-Domain Constructs Consisting of Cat and C Domains
4.9 Oligomeric States of Full-Length Integrase and Modeling of Its Structure
4.10 Structural Studies of Inhibitor Complexes of Integrase
4.11 Structural Basis of Enzymatic Activity of Integrase
4.12 Concluding Remarks
Acknowledgment
References

4.1 INTRODUCTION

Integrase (IN) is one of three enzymes encoded by all retroviral genomes[37] and the one least well characterized in structural terms. The structures of the other two enzymes, protease (PR)[60] and reverse transcriptase (RT),[52,54] have been investigated in detail during the last 20 years, using crystallography and nuclear magnetic resonance (NMR) spectroscopy. A very large number of such structures, solved for both full-length apoenzymes and complexes with substrates, products, effectors, and inhibitors, have been published.[57] The detailed structural knowledge, based on low- to medium-resolution structures of RT and medium- to atomic-resolution structures of PR, has been of considerable use in the design of clinically relevant inhibitors of these enzymes. At this time, 15 nucleoside and nonnucleoside inhibitors of RT as well as 10 inhibitors of PR have been approved by the U.S. Food and Drug Administration (FDA) for treatment of AIDS. By contrast, fewer inhibitors of IN have been discovered so far, and only one has gained FDA approval as an AIDS drug, although acceptance of the other members of this family is likely in the near future.

Although many anti-HIV drugs are already available, serious side effects and the emergence of resistance necessitate development of novel compounds, especially utilizing different targets. Targeting IN seems to be particularly promising[44,47] since, unlike PR and RT, this enzyme does not appear to have its human homologues. Drugs against IN might be given in higher, more effective doses with better tolerated side effects. The inhibitors/drugs currently in animal

HIV-1 Integrase: Mechanism and Inhibitor Design, First Edition. Edited by Nouri Neamati.

experimental or human clinical trials seem to be keeping this promise, having in the short term fewer side effects as compared to FDA-approved anti-PR or anti-RT drugs. In consequence, drugs targeted against IN may be given in sufficiently high doses to completely block the enzyme from integrating viral DNA into the host cell genome, thus allowing the host immune system to fight off the infection completely.

4.2 AMINO ACID SEQUENCE AND DOMAIN STRUCTURE OF RETROVIRAL INTEGRASES

Whereas human immunodeficiency virus type 1 (HIV-1) IN is clearly the most medically relevant IN, extensively investigated for over 20 years, the enzyme encoded by avian sarcoma virus (ASV) was studied even earlier.[29] In addition, enzymes from other retroviruses, including HIV-2, simian immunodeficiency virus (SIV), and feline immunodeficiency virus (FIV), have been investigated as well. Although significant amount of work was done with the latter enzyme,[53] it will not be further discussed here since no crystals could be obtained (Alexandratos, unpublished results). The sequence identity/similarity percentages for full-length HIV-1 IN are 58/74% (in comparison with SIV IN) and 23/37% (ASV IN), respectively (Fig. 4.1).

A single polypeptide chain of retroviral IN comprises ~290 amino acids and consists of three clearly identifiable domains,[8] although definition of the domain boundaries is somehow arbitrary due to the presence of linking sequences of variable lengths. As shown in Figure 4.1, the N-terminal (N) domain of HIV-1 IN contains residues 1−54, the catalytic (CAT) core domain contains residues 55−209 in HIV-1 IN (54−216 in ASV IN), and the C-terminal (C) domain of HIV-1 IN contains residues 210−288 (217−286 in ASV IN). Slightly different starting and ending sequences have been utilized for cloning of individual domains and/or two-domain constructs. For individual domains, the identity/similarity percentages for the N-terminal domain are 55/76% comparing HIV-1 to SIV IN and 26/46% comparing it to ASV IN; for the CAT domain they are 61/77% and 27/46% and for the C

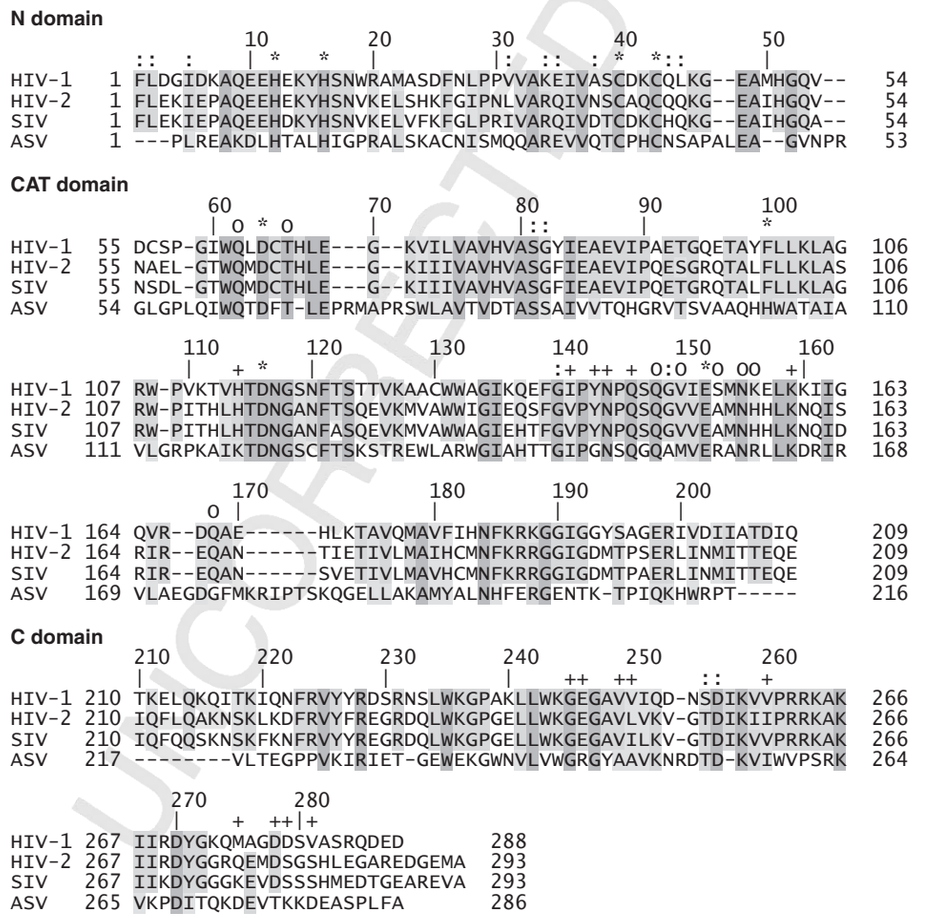

Figure 4.1 Amino acid sequence alignment of retroviral integrases: (green) all residues identical; (blue) 3 of 4 residues identical; (yellow) similar residues; (*) metal cation binding; (:) structurally important; (+) DNA binding; (○) inhibitor binding.

domain they are 53/68% and 14/25%, respectively. Clearly, sequence conservation is the lowest for the C domain.

4.3 CRYSTALLIZATION OF INTEGRASE

No well-diffracting crystals of the full-length HIV or ASV IN have been reported so far, although crystals diffracting to 8 Å resolution have apparently been obtained.[12] The presence in IN of three domains with flexible linkages as well as limited solubility and stability of the protein are likely the main reason for this failure. The success of crystallization of individual domains was also mixed, with no reports of crystals of the smaller N and C domains. Crystals of the CAT domain of HIV-1 IN were only obtained as a result of an extensive mutagenesis study which identified a mutant, F185K, with enhanced solubility.[20] To maintain solubility, 500 mM NaCl, 10% glycerol, and 5 mM dithiothreitol were used as protein buffers. The crystals grew from 0.4 M ammonium sulfate, 3−8% polyethylene glycol (PEG) 8000 and 0.1 M cacodylate buffer, pH 6.5. A protein with a substitution F185H, corresponding to the residue present in ASV-IN, was also crystallized.[5]

Although the CAT domain of ASV IN could be crystallized without mutations, it required special precautions in protein handling. The protein was stored in 20% glycerol and dialyzed into a buffer containing 150 mM NaCl and 5% glycerol directly prior to crystallization. The protein maintained stability for a few days at 4°C. Crystals of the ASV CAT domain were grown using two different precipitants, namely *N*-(2-hydroxyethyl)piperazine-*N*′-2-ethanesulfonic acid (Hepes), pH 7.5, 1.67 M ammonium sulfate and 2% PEG400, or 20% PEG4K, 20% isopropanol with a variety of buffers in the pH range 5−8. The latter conditions were useful in investigating conformational changes of the active site under different pH values and for investigating complexes with divalent cations.

The N/CAT construct of HIV-1 IN was crystallized using a soluble variant of the protein with three mutations: F185K, W131D, and F139D.[58] The protein buffer contained 0.5 M NaCl, 5% glycerol, 0.1 mM $ZnCl_2$, 10 mM 1,1,1-trichloro-2,2-bis(*p*-chlorophenyl)ethane (DTT), and 20 mM Hepes (pH 7.5). The combination of mutations and the specific buffer allowed to increase the protein concentration up to 10 mg/mL. The protein was crystallized in relatively simple conditions with the well solution containing 0.7 M NaH_2PO_4, 1.0 M K_2HPO_4, and 0.1 M acetate buffer (pH 4.6).

For crystallization of the CAT/C construct of HIV-1 IN, two types of mutations were introduced to the protein sequence.[12] Mutations W131D, F139D, and F185K were responsible for increasing solubility, whereas C56S and C286S prevented nonspecific aggregation. This stable and soluble construct was crystallized from 2.2 M sodium formate with 150 mM sodium citrate, 3 mM DTT, and 3 mM Chaps buffer (pH 5.6). An analogous two-domain construct of SIV IN was crystallized from 0.1 M 2-morpholinoethanesulfonic acid (MES) (pH 5.7), 8% PEG6K, 1.5% dioxane, and additionally 100 mM $MgCl_2$ to improve crystal quality.[13] Only a single mutation, F185H, was implemented to improve protein solubility.

4.4 CATALYTIC DOMAIN OF INTEGRASE

The structure of the isolated CAT domain (Fig. 4.2*b*) has been determined in about three dozen crystallographic studies of HIV-1 and ASV INs and in five medium- to low-resolution studies in fusion with one of the terminal domains. Since crystals of the ASV protein were easier to grow, they were studied more frequently, yielding excellent structural data, such as the atomic resolution structure 1CXQ.[40] The CAT domain has been studied in its apo form and in various metal-complexed forms, including the catalytically competent divalent cations Mg^{2+} and Mn^{2+}. Again, ASV IN has provided a more exhaustive picture of metal coordination by the catalytic core domain, including occupancy of multiple metal sites, or the presence of cations such as Zn^{2+} that can also act as inhibitors of IN activity. In addition, a low-resolution structure of the CAT domain in a fusion construct with the C domain confirmed the same structural principles for the SIV IN (1C6V). While eight structures of small-molecule inhibitor complexes of the catalytic domain have been published, it has not been possible to elucidate any structure of a CAT−DNA complex. In variance with the situation concerning the structure of the peripheral IN domains, no solution structure of the CAT domain is available.

The CAT domain is built around a five-stranded mixed β sheet flanked by α helices (Fig. 4.2*b*). The antiparallel $\beta_1-\beta_2-\beta_3$ hairpin-type arrangement is extended by two parallel strands β_4, β_5, which are part of two $\beta-\alpha-\beta$ crossovers, with the intervening helices α_1, α_3 plus a helical turn α_2, all located on one side of the β sheet. Two long, nearly perpendicular helices (α_4, α_5) formed within the C-terminal sequence cover the other side of the sheet. The catalytic residues of the D,D(35)E sequence signature found in all INs are presented by the middle of chain β_1 (D64), the loop connecting $\beta_4-\alpha_2$ (the second aspartate), and the N-terminal segment of α_4 (the glutamate). They are juxtaposed in a row within a patch of negative charge on the surface of the rather flat, slablike molecule. The active-site face of the slab is opposite the CAT domain dimerization face and, therefore, the two active sites of the dimeric enzyme are far apart, nearly as far as the architecture of the dimer allows. Dimerization of the CAT domain involves a tandem of predominantly hydrophobic $\alpha_1,\ldots,\alpha_5$ interactions plus hydrophilic contacts in the middle of the dimer and leads to a twofold symmetric molecule. The latter

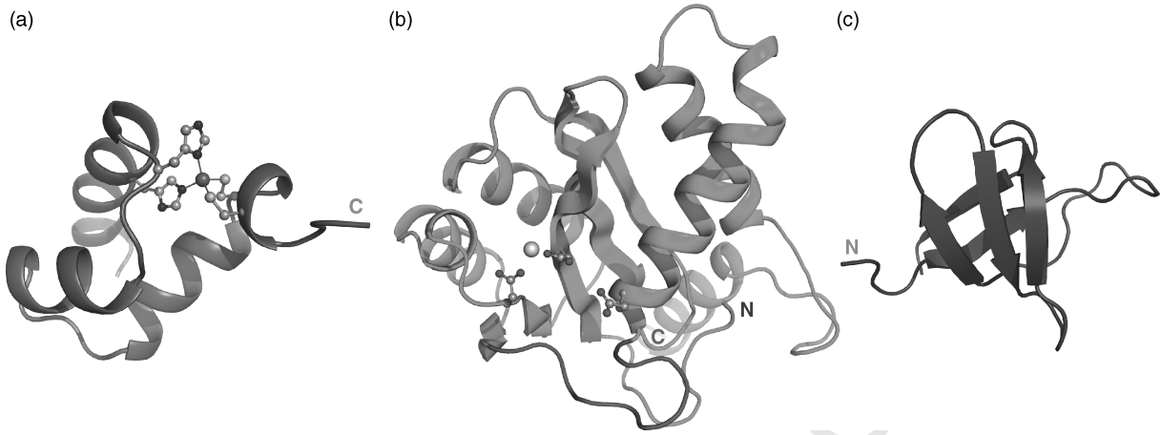

Figure 4.2 Structures of monomers of individual domains of HIV-1 IN. (*a*) The N domain (left, in blue) with a Zn^{2+} cation (large sphere) coordinated (thin lines) by a zinc-finger motif (ball and stick) is represented by the NMR structure 1WJC.[10] (*b*) The CAT domain (center, in green), shown with the D,D (35)E catalytic residues (ball and stick), a magnesium cation (large sphere) coordinated in site I and the flexible active-site loop highlighted, is represented by the crystal structure 1BL3.[46] (*c*) The C domain (right, in red) is represented by the NMR structure 1IHV.[39] This and all subsequent figures were prepared with PyMOL.[18]

interactions are interesting because they are connected with the formation of a hydrophilic cavity in the center of the dimer filled by a few water molecules.

While the Cα traces of the ASV and HIV-1 CAT domains superpose quite well, the agreement of their dimers is less optimal and reflects a slight but evident difference in the dimer architecture. As a consequence of this difference, the two active sites of the HIV-1 IN CAT dimer are less distant (38.5 vs. 42.5 Å, as measured by the separation of the catalytic Mg^{2+} ions), in agreement with the different length of the stagger cut (5 vs. 6 bp) introduced by IN in the target DNA. The absolute distance between the two active sites is incommensurate with a 5−6-bp segment of double-helical B-DNA and suggests that the host DNA must be unwound for coordinated processing of the two strands. Until the structure of the complete IN enzyme is solved, it can only be assumed that dimerization of the core domains of the full-length proteins is not different from what has been observed for the isolated CAT domains. This assumption is supported by the consistent picture of CAT dimerization revealed by two-domain IN constructs.

The CAT domain of HIV-1 IN used in the first structure determination [Protein Data Bank (PDB) code 1ITG] contained the F185K mutation introduced in order to enhance solubility. The cacodylate from the crystallization buffer was found attached to the cysteine side chains of the protein, including C65 located in the active-site area.[20] The constellation of the catalytic acids D64, D116, and E152 was found in the 1ITG structure to be in an "inactive," nonnative configuration (Fig. 4.3*a*). The distortion of the catalytic apparatus became apparent only later by comparison with other unperturbed structures, notably the ASV IN CAT domain.[6,7] The nonnative character of the active site is manifested by the altered conformations of the two aspartic acids, including a major reorientation of the loop carrying the D116 residue, and in complete disorder of the helix fragment with the E152 residue and the entire flexible active-site loop in front of it (in total, residues 141−153). It is unlikely that the distortion of the active site was affected by the presence of the unnatural arsenic substituent, as in a related arsenic-free HIV-1 IN structure the catalytic aspartic acids are found in exactly the same inactive conformation.[5] Although the structure 1ITG failed to map the functional state of the protein, it provided the first chain tracing and was important in revealing the plasticity of the IN active site and its ability to adopt different conformations.

Perhaps the most significant consequence of the inactive conformation of the catalytic residues is the inability of the two aspartate side chains to bind a catalytic divalent metal cation in a coordinated fashion. Such a cation, revealed by Mg and Mn complexes of ASV IN[4,7] and later by Mg complexes of HIV-1 IN,[28,46] has an octahedral coordination completed by four water molecules (Fig. 4.3*b*). The triad of the catalytic acids can remain in the active conformation even in the absence of metal cations, but then the carboxylate groups are held in place by water-mediated hydrogen bond bridges (D−Wat−D64−Wat−E). However, as revealed by the atomic-resolution structures of ASV IN, and in agreement with the requirement for basic conditions for IN activity (peak endonuclease activity at pH 8.5[41]), conformational changes in the active site take place at pH below 6 and consist of protonation and a concomitant swing of the D64 carboxylate group out of its metal-coordinating position and into a dual-hydrogen-bond lock with a neighboring asparagine residue. In

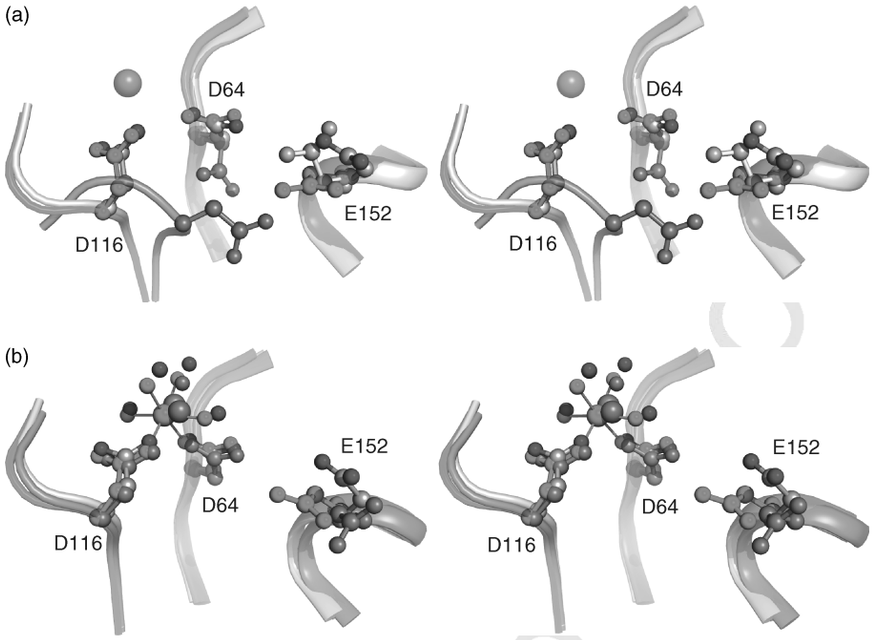

Figure 4.3 Active site of retroviral integrases. The figures show, in stereoview, the three essential acids of the D,D(35)E motif in selected, least-squares-superposed crystallographic structures of the catalytic core domain in (*a*) unliganded or (*b*) Mg-complexed form. The catalytic residues are shown in the context of the protein's secondary structure by which they are contributed, namely an extended β ribbon (the first aspartate, middle of figure), a loop (the second aspartate, left), and an α helix (the glutamate, right). The residue numbering D64/D116/E152 is for the HIV-1 IN sequence and corresponds to D64/D121/E157 in ASV IN. The three divalent-metal-cation-free active sites shown in (*a*) correspond to the first HIV-1 IN structure (1ITG, orange)[20] solved in the presence of arsenic (part of cacodylate buffer) bound to cysteine residues, including one within the active-site area (orange sphere), to another medium-resolution structure of HIV-1 IN (1BI4, molecule C, gray with red O atoms),[46] and to atomic-resolution structure of ASV IN (1CXQ, green).[40] Note that the aspartates in 1ITG have a completely different orientation than in the remaining structures and the entire D116 loop has a different, nonnative conformation. Another symptom of active-site disruption in the 1ITG structure is the absence in the model of the E152 residue, a consequence of disorder in this helical segment. The active sites complexed with the catalytic cofactor Mg^{2+} (large sphere) are shown (*b*) for two crystal forms of HIV-1 IN, 1BL3 (molecule C, gray with red O atoms)[46] and 1BIU (molecule C, orange),[28] and for ASV IN, 1VSD (green).[7] The structure of the ASV IN has the highest resolution, and its quality is reflected in the nearly ideal octahedral geometry (thin green lines) of the Mg^{2+} coordination sphere, which in addition to interactions with the carboxylate groups of both active-site aspartates includes four precisely defined water molecules. The coordination geometry of the HIV-1 IN complex 1BL3 is significantly distorted and in 1BIU only one water molecule was modeled in the coordination sphere. The view direction in both figures is similar, with a small rotation around the horizontal axis.

addition, changes of pH influence the flexible active-site loop, which in HIV-1 IN comprises the residues 141–147, adjacent to the glutamate-bearing N terminus of helix α_4, and which in all the crystal structures shows a variable degree of disorder. The flexible active-site loop contains highly conserved residues and appears to be involved directly in substrate contacts.[34]

There is little doubt that the metal coordination site formed between the two aspartate side chains (site I) corresponds to a cation essential for catalysis. The perfect octahedral geometry of this site explains why mutations of the catalytic aspartates cannot be tolerated. However, increasingly larger cations can still be accommodated, from Mg^{2+} (mean metal–O distance 2.11 Å), to Mn^{2+} (2.23 Å), and even Cd^{2+} (2.43 Å) and Ca^{2+} (2.46 Å for incomplete coordination sphere). Estimation of metal binding geometry is more reliable from the ASV IN structures, which are in excellent agreement with expected coordination stereochemistry, for instance with valence parameters[3] of the central ion, which for the structures listed in Table 4.1 are calculated as 1.95 (1VSD), 1.92 (1A5V), or 1.79 (1VSJ), the ideal target being 2.00. The corresponding values for the

TABLE 4.1 Experimental Atomic Coordinate Sets for IN Determined by X-Ray Crystallography or NMR Spectroscopy in Protein Data Bank

Protein	Virus	PDB code	Method	Resolution (Å)	Ligand	Year	Reference	Remark
N	HIV-1	1WJA	NMR	n/a	Zn	1997	10	
N	HIV-1	1WJC	NMR	n/a	Zn	1997	10	
N	HIV-1	1WJE	NMR	n/a	Cd	1998	9	
N	HIV-2	1E0E	NMR	n/a	Zn	2000	24	
CAT	HIV-1	1ITG	X ray	2.30	Cacodylate	1994	20	*a*
CAT	ASV	1ASU	X ray	1.70		1995	6	
CAT	ASV	1ASV	X ray	2.20		1995	6	
CAT	ASV	1ASW	X ray	1.80		1995	6	
CAT	ASV	1VSD	X ray	1.70	Mg	1996	7	
CAT	ASV	1VSE	X ray	2.20		1996	7	
CAT	ASV	1VSF	X ray	2.05	Mn	1996	7	
CAT	HIV-1	2ITG	X ray	2.60		1996	5	
CAT	ASV	1VSH	X ray	1.95	2 × Zn	1997	4	*b*
CAT	ASV	1VSI	X ray	2.20	Ca	1997	4	
CAT	ASV	1VSJ	X ray	2.10	2 × Cd	1997	4	*c*
CAT	ASV	1A5V	X ray	1.90	Mn + Y-3	1998	42	*d*
CAT	ASV	1A5W	X ray	2.00	Y-3	1998	42	*d*
CAT	ASV	1A5X	X ray	1.90	Y-3	1998	42	*d*
CAT	HIV-1	1BHL	X ray	2.20		1998	46	
CAT	HIV-1	1BI4	X ray	2.50		1998	46	
CAT	HIV-1	1BL3	X ray	2.00	Mg	1998	46	
CAT	HIV-1	1BIS	X ray	1.95		1998	28	
CAT	HIV-1	1BIU	X ray	2.50	Mg	1998	28	
CAT	HIV-1	1BIZ	X ray	1.95	Cacodylate	1998	28	*a*
CAT	ASV	1VSK	X ray	2.20		1998	41	
CAT	ASV	1VSL	X ray	2.20	Zn	1998	41	
CAT	ASV	1VSM	X ray	2.15		1998	41	
CAT	HIV-1	1B92	X ray	2.02	Cacodylate	1999	30	*a*
CAT	HIV-1	1B9D	X ray	1.70	Cacodylate	1999	30	*a*
CAT	HIV-1	1B9F	X ray	1.70	Cacodylate	1999	30	*a*
CAT	ASV	1CXQ	X ray	1.02		1999	40	
CAT	ASV	1CXU	X ray	1.42		1999	40	
CAT	ASV	1CZ9	X ray	1.20		1999	40	
CAT	ASV	1CZB	X ray	1.06		1999	40	
CAT	HIV-1	1QS4	X ray	2.10	Mg + 5ClTEP	1999	27	*d*
CAT	HIV-1	1EXQ	X ray	1.60	2 × Cd	2000	12	*e*
CAT	HIV-1	1HYV	X ray	1.70	TTA	2001	49	*d*
CAT	HIV-1	1HYZ	X ray	2.30	TTO	2001	49	*d*
CAT	HIV-1	3L3U	X ray	1.40	Glycerol	2010	59	
CAT	HIV-1	3L3V	X ray	2.00	Sucrose	2010	59	
CAT	HIV-1	3LPT	X ray	2.00	3	2010	15	*d*
CAT	HIV-1	3LPU	X ray	1.95	6	2010	15	*d*
N-CAT	HIV-1	1K6Y	X ray	2.40	Zn	2001	58	
CAT-C	ASV	1C0M	X ray	2.53		2000	62	
CAT-C	ASV	1C1A	X ray	3.10		2000	62	
CAT-C	SIV	1C6V	X ray	3.00		2000	13	*f*
CCD+LEDGF	HIV-1	2B4J	X ray	2.02		2005	14	
CTD-CCD+LEDGF	HIV-2	3F9K	X ray	3.20		2009	32	
IN+DNA	PFV	3L2R	X ray	2.88	Mg	2010	31	*g*

Note: From Ref. 2. In cases where the results of the NMR experiments were published as both ensembles of structures and regularized mean structures, only the latter is referenced.

[a] Cacodylate residue attached to a cysteine side chain found in the active site, which is in an inactive conformation.

[b] Two tetrahedral Zn^{2+} ions coordinated in the active site, bridged by D64 and a water molecule.

[c] Two octahedral Cd^{2+} ions coordinated in the active site, bridged by D64 and a water molecule.

[d] Inhibitor abbreviations: Y-3, 4-acetylamino-5-hydroxynaphthalene-2,7-disulfonic acid; 5ClTEP, 1-(5-chloroindol-3-yl)-3-hydroxy-3-(2*H*-tetrazol-5-yl)-propenone; TTA, tetraphenyl-arsonium; TTO, (3,4-dihydroxyphenyl)triphenylarsonium; 3, 2-(6-chloro-2-oxo-4-phenyl-1,2-dihydroquinolin-3-yl)acetic acid; 6, 2-(6-chloro-2-methyl-4-phenylquinolin-3-yl)pentanoic acid.

[e] Two Cd^{2+} ions coordinated in the active site, bridged by C65 and with D64 swung away.

[f] Model includes four CAT domains in the asymmetric unit, but only one C domain.

[g] This is the highest resolution structure among seven related data sets described in this chapter.

HIV-1 IN data indicate a high level of error, for example, 1.23/0.91 (1BL3) or even 1.08/0.80/0.79 (1QS4), presumably as a consequence of poor data quality or refinement protocols. There is an important difference between ASV and HIV-1 IN in coordinating high-electron metals in site I, connected with the presence of a cysteine residue at position 65 in the latter enzyme. The thiol group of this residue is found in the coordination sphere of the cadmium cations in 1EXQ.[12] Since no such possibility exists in ASV IN, which has a phenylalanine residue immediately following the first catalytic aspartate, high-electron metals may have different impact on the catalytic properties of integrases from these two viruses. With light metals, such as Mg^{2+}, the thiol group of C65 in HIV-1 IN assumes a totally different orientation and, consequently, there is no difference in the coordination chemistry between ASV and HIV-1 IN.

Recent high-resolution structures of the CAT domain of HIV-1 IN in the presence of bound sucrose and glycerol[59] provided a more detailed description of this fragment of the enzyme as well as identified an allosteric inhibitory binding site. Extension of the resolution to 1.4 Å allowed a large increase in the number of identified water molecules as well as assigning alternate conformation to as many as 27 side chains.

4.5 N-TERMINAL DOMAIN OF INTEGRASE

NMR structures of the isolated N domains of retroviral INs were solved for enzymes from HIV-1[10] and HIV-2.[24] Four independent views of the N-terminal domain are also available in a medium-resolution crystal structure of a two-domain construct of HIV-1 IN that contains the N and CAT domains.[58] The solution structure of the HIV-1 IN N domain showed the existence of dimers consisting of two interconverting protein forms.[10] The two forms, denoted D (PDB code 1WJA) and E (1WJC), were observed together in the NMR spectrum, with the D form seen mostly above ~300 K and the E form below that temperature. A form intermediate between these two was reported for an H12C mutant of the N-terminal domain.[9]

The structure of a monomer of the N domain consists principally of four helices (Fig. 4.2*a*). Helix 1 comprises residues 2−14 in the E form and 2−8 in the D form, helix 2 residues 19−25, helix 3 residues 30−39, and helix 4 residues 41−45. The segment beyond residue 46 is disordered. A zinc atom is tetrahedrally coordinated by H12, H16, C40, and C43, although the details of the interactions with the histidine residues differ between forms D and E.

The E form of the N domain is very similar to its counterpart seen in the crystal structure of the two-domain construct 1K6Y,[58] with a root-mean-square (rms) deviation of 1.05 Å between molecules A of the models. By comparison, the rms deviations between molecule A and the other three molecules are seen in the crystal range from 0.28 to 0.63 Å. Form D of the N domain deviates by almost 2 Å from its crystallographic counterpart. As expected, the interactions of the Zn^{2+} cation with its ligands in the crystal structure correspond to the structurally closer E form.

The structure of the N domain of HIV-2 IN[21,23] is very similar to its HIV-1 counterpart. A comparison between molecule A of the first model in the assembly in 1E0E (no average structure available) and molecule A of 1K6Y[58] shows an rms deviation of 0.86 Å, although sequence identity between the two proteins is only 55%. The details of the interactions with Zn^{2+} are also almost identical between the integrase N domains of HIV-1 (E form) and HIV-2.

4.6 C-TERMINAL DOMAIN OF INTEGRASE

The structure of the isolated C domain of HIV-1 IN (residues 220−270, carboxyl terminus truncated) was solved using NMR by two independent groups (1IHV[39]; 1QMC[21,22]). In addition, two-domain structures of the CAT and C domains were determined by X-ray crystallography for ASV IN (1C0M, 1C1A[62]), SIV IN (1C6V[13]), and HIV-1 IN (1EX4[12]). The structures of the C domains showed the presence of dimeric molecules, which were modeled as identical in 1IHV and were very similar in 1QMC (rms deviation 0.34 Å calculated for model 1, since no average structure is available). The rms deviation between these two structures is 1.2 Å. The deviations between the NMR structures of the isolated C domain and the crystallographic models of the two-domain constructs are larger, 1.65 Å between 1IHV and 1EX4 (both HIV-1 IN), 1.87 Å for 1C6V (SIV IN), and 2.05 Å for 1C0M (ASV IN). The four C domains present in the crystal structure of ASV IN consist of two very similar pairs (A−B and C−D, rms deviation ~0.15 Å), whereas the rms deviation between molecules A and C is 0.77 Å.

A monomer of the C domain of HIV-1 IN consists of five β strands (residues 222−229, 232−245, 248−253, 256−262, 266−270) arranged in an antiparallel manner in a β barrel (Fig. 4.2*c*). Eighteen residues that were not included in the constructs used in the NMR experiments are also not seen in the X-ray structures of HIV-1 and SIV IN and are presumed to be disordered. The topology of the C domain is reminiscent of SH3 domains, found in many proteins that interact with either other proteins or nucleic acids, although no sequence similarity to SH3 proteins could be detected.

4.7 TWO-DOMAIN CONSTRUCTS CONSISTING OF N AND CAT DOMAINS

Only a single 2.4-Å-resolution crystal structure of a construct containing the N and CAT domains of HIV-1 IN is

available (1K6Y[58]), but with four molecules present in the asymmetric unit, it offers multiple views. The interdomain linker region (residues 47–55) is disordered in all molecules, but the postulated domain connectivity is unambiguous (Fig. 4.4*a*).

The structure consists of two dimers, A/B and C/D. The twofold relationship between the catalytic domains resembles that of the isolated CAT domains, and the same twofold axis also relates the N domains of each dimer. Molecules A and D are very similar (rms deviation 0.43 Å), whereas molecules B and C are more distant (rms deviation 1.85 Å) mostly due to small changes in the interdomain angles. The two domains of a single molecule make only limited contacts through the tip of loop 190–192 of the CAT domain and one side of helix 20–24 in the N domain.

4.8 TWO-DOMAIN CONSTRUCTS CONSISTING OF CAT AND C DOMAINS

The structures of two-domain constructs comprising the CAT and C domains were solved independently for HIV-1 IN at 2.8 Å resolution (PDB code 1EX4),[12] for SIV IN at 3.0 Å resolution (1C6V),[13] and for two crystal forms of ASV IN at 2.5 (1C0M) and 3.1 Å (1C1A).[62] The crystals of the HIV-1 protein contain two molecules forming a dimer, although the twofold axis relating the CAT domains differs from the operation connecting the C domains. In each molecule, the two domains are connected by a long, well-defined helix comprising residues 195–222. The helix separates the two C domains by as much as 30 Å (Fig. 4.4*b*).

The two crystal forms of the ASV protein contain a single dimer or a pair of dimers. Similarly to what was observed in HIV-1 IN, the symmetry operations between the two domains of each dimer differ for the CAT and C domains. The linker between the CAT and C domains comprises residues 206–223, which assume an extended conformation, and not the helical form observed for HIV-1 IN. The closest distance between the two domains is only ~7 Å, since the N-terminal part of the CAT domain is comparatively close to parts of the C domain.

Whereas the crystals of SIV IN also contain two dimers in the asymmetric unit, only a single C domain (denoted X) could be traced unambiguously. The chain connecting it to the CAT domain could not be traced and the authors postulated a connection with chain A of the catalytic

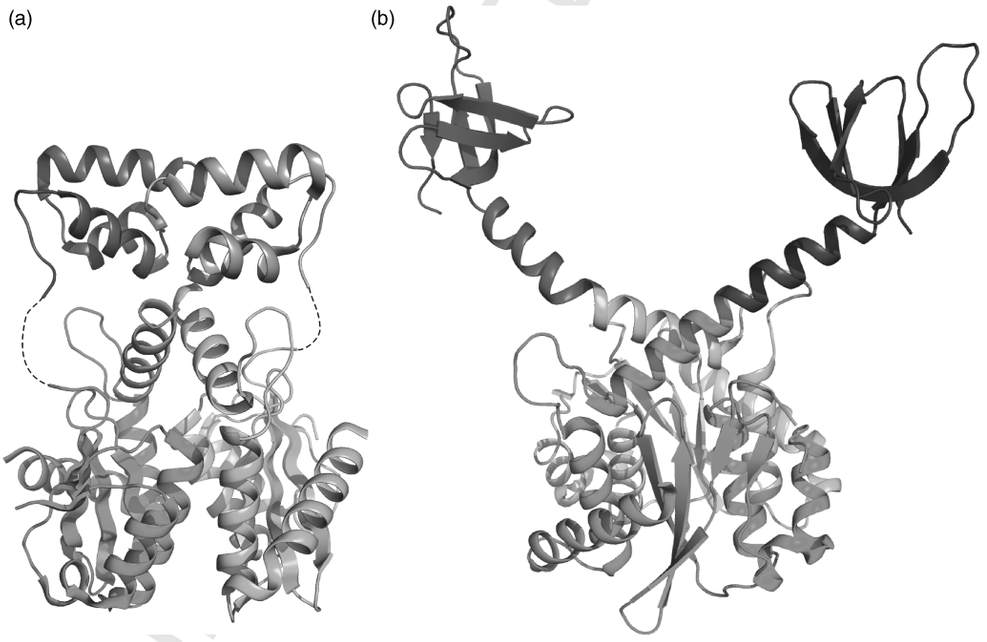

Figure 4.4 Three-dimensional structures of dimeric two-domain constructs of HIV-1 integrase determined by X-ray crystallography. (*a*) In the N/CAT structure (1K6Y[58]), the linker between the domains is disordered (dashed line). (*b*) In the CAT/C dimer (1EX4[12]), the interdomain linker forms a long helix. Because of different degree of deformation of this helix, the relative CAT/C orientation in the two monomers is different. The N domain is shown in shades of blue, the CAT domain in shades of green, and the C domain in shades of red. The two fold axes of the dimeric molecules have a roughly vertical orientation in both figures, although the axes relating the CAT and C domains in (*b*) are not identical.

domain.[13] If that were the case, the two domains would form a fairly compact molecule with multiple interdomain contacts. However, an alternative assignment of the visible C domain to the D chain of CAT[58] would create an extended two-domain molecule not unlike in the other two enzymes, although the interdomain angles would differ in each of the structures. In any case, a comparison of the three enzymes makes it clear that the arrangement of the domains shows considerable variability and may be influenced by other parts of the molecular complex.

4.9 OLIGOMERIC STATES OF FULL-LENGTH INTEGRASE AND MODELING OF ITS STRUCTURE

The oligomerization state of IN in vivo is still not certain, but much in vitro work has shed light on this matter. The isolated N, CAT, and C domains all remain in solution as dimers, a conclusion uniformly supported by solution chemistry and structural biology studies.[16] However, experiments that found IN−DNA interaction sites by photo-cross-linking also suggested that IN acts as an octamer.[35] Comparison of simulation analysis against time-resolved fluorescence anisotropy measurements of rotation correlation times could distinguish monomers, dimers, and tetramers, while octamers could not be resolved from higher order species.[19] At micromolar concentration IN exists as tetramers, octamers, and higher order aggregates, but such concentration is much higher than cellular. At catalytic (submicromolar) concentration, these experiments showed that IN could exist as a monomer, while Zn^{2+} stimulated dimer formation. However, the authors noted that the standard buffer conditions include detergents, which dissociate IN oligomers.[19] With detergent-free conditions during purification and assays, IN exhibits different assembly and catalytic properties. However, all of these experiments use indirect measurements of the size of IN oligomers. Atomic force microscopy images of intact INs in complex with DNA substrate have shown visually that the size of these complexes is consistent with a tetramer of IN molecules.[1] This agrees with several IN−DNA models, with analysis of IN isolated from nuclear extracts and its complex with lens epithelial−derived growth factor (LEDGF),[14,32] and with dynamic light scattering experiments. Models of IN complexes based upon structures of recombinases (which bind DNA molecules forming Holliday junctions) have been proposed.[36] Testing hypotheses derived from purely computational work will require specific predictions of residues designated for mutation combined with appropriate in vitro and in vivo assays.

The structure of Tn5 transposase as a synaptic complex transition state intermediate came as a breakthrough for integrase modelers.[17] The prokaryotic Tn5 transposase performs a series of catalytic steps, with distinct processing (endonucleolytic cleavage) and joining reactions, which are very similar to those catalyzed by retroviral INs. Also, its catalytic core domain is structurally very similar to those of retroviral INs. Tn5 functions as a dimer and its DNA binding sites provide a clear template for modeling IN−DNA interactions. These models can be used to predict IN amino acid residues important for DNA binding.

The assembly of HIV-1 IN into oligomers is different when in complex with Mn^{2+} versus Mg^{2+} under various in vitro conditions.[38] These experiments did not clarify which cation is preferred, but they did show that HIV-1 IN had no active-site cation preference when already in complex with a structural (noncatalytic) Zn^{2+} cation. The authors concluded that binding of the catalytic cation and DNA requires a specific preexisting IN conformation.

While there are no three-dimensional structures of HIV or ASV IN in complex with nucleic acids, DNA cross-linking studies implicate certain positively charged or hydrophobic residues for involvement in IN−DNA interactions, for example, H114, Y143, and K159 in HIV-1 IN.[35] The DNA binding C domain contains less well-conserved residues which have been identified as important for DNA binding, namely, HIV-1 E246, K258, P261, R262, K264, with some weaker involvement of S230 and R231.[25] The somewhat lower degree of sequence conservation in this region may reflect differences in specificity. Finally, the C domain also plays an important role in IN dimerization. When amino acid residues L241 and L242 along the C-terminal dimer interface are mutated to alanine, they disrupt IN dimerization and strongly reduce catalysis.[45]

As no structure of an intact HIV or ASV IN molecule has been reported to date, the two-domain IN constructs, namely N/CAT and CAT/C, are being used as starting points for building models of the complete HIV-1 IN protein and IN−DNA complexes.[58] These structures will be informative since they complement each other and physically fit well together. However, the IN domains are connected by flexible linkers allowing significant interdomain variability, and a three-domain model may not reflect the actual conformation(s) of the intact protein alone or in complex with DNA (Fig. 4.5).

A major advance in our understanding of the oligomeric state of IN and the interactions with DNA was provided by the structures of IN from prototype foamy virus (PFV) complexed with DNA.[31] Unlike the previously studied cross-linked IN−DNA complexes, which were notoriously refractory to crystallization, stable complexes of PFV IN with a 19-bp donor DNA could be prepared and crystallized. Structures of the full-length PFV IN which also contained divalent cations and strand transfer inhibitors have clarified the nature of IN tetramers and have shown that some aspects of the previous models of HIV-1 IN−DNA complexes were incorrect. These structures have also shown that only one of

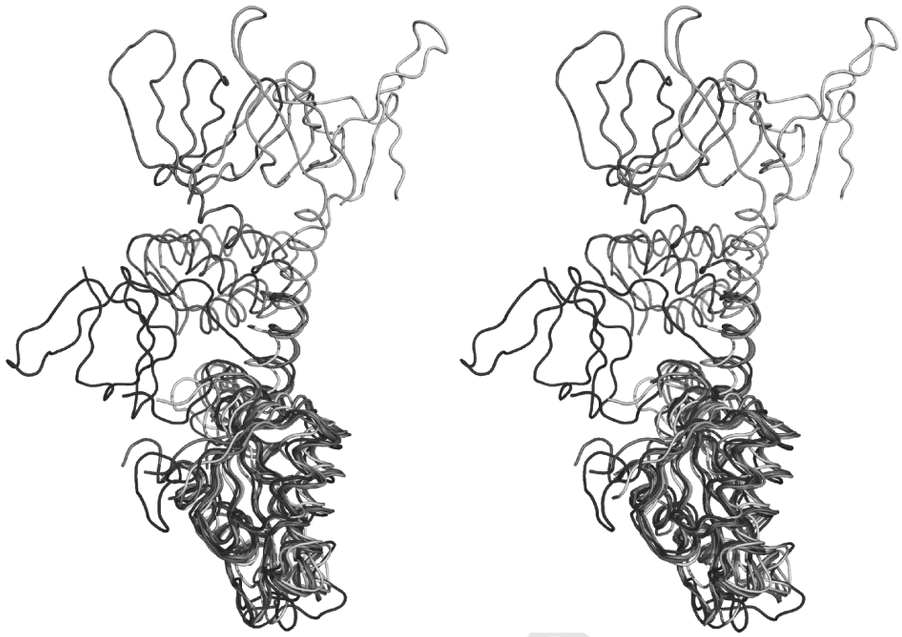

Figure 4.5 Stereoview of structural superposition of several two-domain constructs of retroviral integrases. The superpositions were calculated using only the Cα atoms of the CAT domain (bottom) to show a possible mutual orientation of all three domains. Until the structure of intact IN is determined experimentally, this is the best approximation of the three-dimensional model of the enzyme, here shown only for the monomeric molecule. According to available data on dimeric structure of IN domains, a homodimer of IN could be created by rotating the above model by 180° around the vertical line and placing it face to face with the original copy, so as to re-create the dimeric interface at the flat face (right-hand side) of the CAT domain. The figure uses the following color code: red and orange, molecules A and B of the HIV-1 N-CAT protein 1K6Y[58]; blue, ASV CAT-C protein 1C0M[62]; dark/light green, molecules A and B of the HIV-1 CAT-C protein 1EX4[12]; yellow, SIV CAT-C protein 1C6V.[13] In the latter structure, the domains that are displayed (D and X) were not interpreted as a single molecule in the original publication. Note that the N domain (in the red and orange models) seems to have a rather stable orientation relative to the CAT domain. The C-terminal domain (all other models) covers a wide angular range of its disposition relative to CAT.

the active sites of each dimer of the CAT domain participates in enzymatic activity.

4.10 STRUCTURAL STUDIES OF INHIBITOR COMPLEXES OF INTEGRASE

Structural data on inhibitor complexes of HIV-1 and ASV IN are limited to a few structures of the CAT domain (Table 4.1). In two studies, arsenic derivatives were cocrystallized with the HIV-1 IN,[49] but the inhibitor molecules were found attached to cysteine residues in the same fashion as the cacodylate ion in the 1ITG structure. The structure of an indole derivative in complex with the Mg^{2+}–HIV-1 IN CAT domain[27] showed that indeed the molecule is capable of binding within the active-site area of the enzyme, between the coordinated Mg^{2+} cation and the catalytic glutamate, with which it forms hydrogen bonds. However, this inhibitor was observed only in one of the three independent copies of the enzyme molecule present in the crystal. 4-Acetylamino-5-hydroxynaphthalene-2,7-disulfonic acid (Y-3) was cocrystallized with the ASV IN CAT domain in the absence and presence of Mn^{2+}.[42] This aromatic molecule with several hydrophilic substituents binds not in the active site of the enzyme but rather on its surface, where it participates in crystallographic contacts, although there is no interference with CAT dimerization. Its presence in the crystals is, however, not a crystallographic artifact since it is observed in the same context at different pH conditions and regardless of metal coordination. Although it forms no direct interactions with the catalytic residues, Y-3 does seem to influence the conformation of the flexible active-site loop by binding to Y143 and K159 (ASV numbering). Y-3 very likely directly interferes with DNA

binding by hydrogen bonding to K119 (a residue corresponding to H114 in HIV-1 IN, shown to be capable of cross-linking to DNA). It is quite possible that these interactions are the basis of its inhibitory capacity.

Since IN must form at least a dimer to be catalytically active, prevention of dimerization offers an interesting option for its inhibition.[11] Several studies have reported inhibition of IN activity through the use of peptides derived from amino acid sequences responsible for the dimerization of the CAT domain,[43,55] although no structural data are available. In some cases, it was possible to confirm that such peptides disrupted the association−dissociation equilibrium[48] or the cross-linking of the IN dimer.[63] On the other hand, Hayouka et al.[33] have demonstrated that the opposite concept, namely forcing IN to form higher order oligomers, may be a useful approach to rendering the IN inactive. Specifically, they used peptides (called "shiftides") derived from a cellular IN binding protein, LEDGF/p75, to inhibit the DNA binding of IN by shifting the enzyme's oligomerization equilibrium from the active dimer toward the inactive tetramer, which is incapable of catalyzing the first step of integration, that is, the 3'-end processing.

Crystal structures of the complexes of HIV-1 CAT with inhibitors that bind into the LEDGF binding pocket have been recently published.[15] The compounds used in these studies were derivatives of 2-(quinolin-3-yl)acetic acids that inhibit LEDGF−IN interactions at concentrations in the submicromolar range as well as show significant anti-HIV activity.

4.11 STRUCTURAL BASIS OF ENZYMATIC ACTIVITY OF INTEGRASE

The CAT domain of IN is responsible for the enzymatic activity of the enzyme. The integration reaction is catalyzed by IN in two steps, called processing and joining. First, the viral DNA is nicked near the 3' ends of both strands, and the exposed ends are then inserted into host target DNA, with a characteristic stagger between the two insertion points. Both reactions are chemically similar, proceeding via a nucleophilic attack on a phosphorous atom in the DNA backbone by a donor hydroxyl group (water or the newly formed 3'-OH), activated by the catalytic center of the enzyme. In vitro, these reactions require Mg^{2+} or Mn^{2+} cations, the latter being more efficient. However, because of physiological abundance, Mg^{2+} is assumed to be the cofactor in vivo. Whereas the isolated CAT domain is capable of catalyzing the processing reaction, the full-length enzyme is necessary for joining to proceed.

The nature and number of divalent metal cations required for catalysis are still under debate. The general composition of the IN active site (a constellation of acid groups) and the similarity of the catalyzed reactions to those carried out by other nucleotidyl transferases would strongly indicate the two-metal-cation mechanism elaborated by Steitz and Steitz.[56] However, despite numerous attempts, it has never been possible to obtain an IN−Mg^{2+} or IN−Mn^{2+} complex with two metal cations in the active site. On the other hand, it was possible to introduce two cations into the active site by using physiologically irrelevant but stronger binding metals, such as Zn^{2+} with ASV IN,[4] Cd^{2+} with ASV IN[4] and with HIV-1 IN,[12] or even Ca^{2+} with ASV IN.[4] The case of zinc coordination is of special note because, first, Zn^{2+} accepts only four tetrahedrally arranged ligands which are a subset of the octahedral sphere of the other cations; second, although it is not a cofactor of IN catalysis in vivo, it can support endonucleolytic activity in in vitro assays; third, it severely impairs polynucleotidyl transferase activities of IN in vitro; and, fourth, its potential interaction with the CAT domain is complicated by the fact that it is the major physiological cofactor of the N-terminal domain.

The most instructive case is Cd^{2+} coordination by IN. One has to clearly distinguish the HIV-1 IN and ASV IN cases because the C65 residue of HIV-1 IN mentioned above actually functions as a bridge coordinating both metal centers (sites I and II) simultaneously, replacing in this role the catalytic D64, which is forced away from its active conformation.[12] In this light, the structure of the ASV IN−Cd^{2+} complex[4] is more instructive for providing insights about the possible two-metal-cation functional state of the enzyme.

There is striking similarity between the Cd^{2+}-complexed active site of ASV IN and those of other nucleotidyl transferases, most notably of RNase H, which has been described in a ternary complex with Mg^{2+} and an RNA/DNA hybrid[50] (Fig. 4.6). First, the metal−metal distance is nearly identical in both structures and compatible with what Yang et al.[61] predict to be required for effective nucleotide bond hydrolysis (4.0 Å). Additionally, in both cases the two metal centers are connected by two bridging ligands, one of them being a conserved aspartate from the catalytic apparatus (D64 in HIV-1 IN). The other bridge is provided by a water molecule in the ASV IN−Cd^{2+} (and also IN−Zn^{2+}) complex but in the RNaseH complex structure this water is displaced and its role is assumed by an O atom of the scissile phosphate group of the RNA substrate. This phosphate group is even more essential for the integrity of the functional active site of RNaseH because it also fills (albeit with less ideal stereochemistry) an additional site in the coordination sphere of site B of RNaseH. (Moreover, the next phosphate of the RNA substrate participates in the activation of the water nucleophile presented in the coordination sphere of site A.) Overall, site B has much less regular stereochemistry, in contrast to the nearly perfect geometry of site A. While the simplicity and approximate mirror symmetry of the two-metal-cation active sites would allow two alternative mappings of the metal centers between

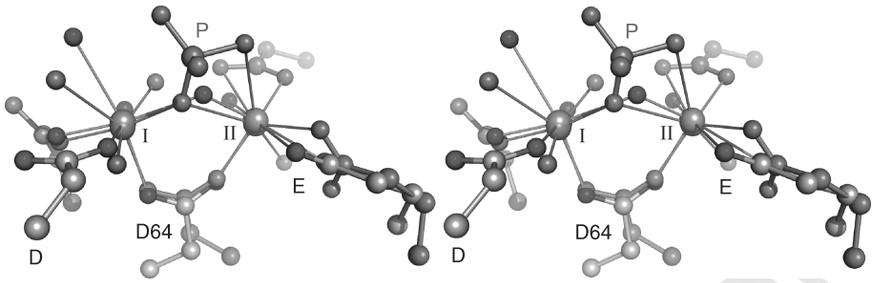

Figure 4.6 Divalent metal cation binding sites in integrases. Comparison of the two sites found in Cd^{2+} complex of ASV IN (PDB code 1VSJ, color code according to element type)[4] and in Mg^{2+}/RNA/DNA complex of RNaseH (1ZBL, molecule A, orange).[50] The active sites were superposed by a simple least-squares fit of the metal sites and the carboxylate O atoms of the bridging aspartate residue, which in ASV IN is the first element (D64) of the D,D(35)E active site. The metal−metal distance in both structures is nearly identical, 4.10 Å in the RNaseH complex and 4.05 Å in the IN complex. Site I, denoted A in Refs. 50 and 51, which in retroviral integrase structures was also seen with the catalytic Mg^{2+} or Mn^{2+} cations, has a more regular octahedral coordination. The coordination spheres I in the two structures are similar, except that two ligands in the equatorial plane, an aspartate (D121 in ASV IN) OD atom and a water molecule, have swapped places (in 1ZBL, the aspartate in question has been mutated to asparagine, D192N). Another important difference is that the bridging water molecule in the ASV IN complex is replaced in the RNaseH complex by an O atom from the scissile phosphate group (P) of the RNA substrate. The phosphate group has a particularly important role in the formation of site II, as it provides two of the ligands. Site II (denoted B in Refs. 50 and 51) has a far less regular geometry; the coordination sphere is incomplete in 1VSJ or highly distorted in 1ZBL. Site II has never been seen occupied by Mg^{2+} or Mn^{2+} in metal complex structures of retroviral INs, and the glutamic acid which takes part in its formation (E157 in ASV IN) is the most mobile element of the IN active site. From this comparison, it is very likely that a proper site II of retroviral integrase, occupied by a catalytically competent metal cation (Mg^{2+} or Mn^{2+}), could only be formed with the participation of a DNA substrate.

RNaseH and retroviral IN, there is little doubt that the correct mapping is A−I and B−II. This is because, as in the RNaseH structure, site I of the IN CAT domain has a nearly perfect coordination sphere, while site II is far less regular, with a missing ligand, large scatter of the Cd^{2+}−O distances, and large angular distortions. If this analogy between the active sites of IN and RNaseH is correct, then the catalytic metal cation at site I of IN would participate in activating a nucleophilic group (e.g., a water molecule) for attack on a substrate DNA phosphate group. Metal II, on the other hand, would play a role in destabilizing the enzyme−substrate complex, that is, in driving the reaction forward. At the completion of a reaction cycle one or both metal cations would probably dissociate as their effective binding (especially at site II) critically depends on the presence of substrate DNA. The parallel between RNaseH and retroviral IN has also a chemical aspect because coordination of two Mg^{2+} cations by RNaseH was easy and occurred at low metal ion concentration only in the presence of the RNA/DNA substrate. With the enzyme alone, the effective Mg^{2+} concentration had to be much higher at nonphysiological levels.[26] With ASV IN, it has not been possible to introduce a catalytic metal cation at site II, despite a thorough experimental survey, in which elevated metal concentrations were used.[4] This difficulty is correlated with the flexibility of the glutamate element of the active site which participates in the formation of site II. It may be necessary for the enzyme to use external means, such as substrate assistance, to sequester an Mg^{2+} cation in site II, with subsequent or simultaneous stabilization of the glutamate side chain.

4.12 CONCLUDING REMARKS

An initial spur of activity in the years 1994−2001 that resulted in a wealth of crystal and NMR structures of retroviral INs was followed by five quiet years, but the pace of structural studies of this enzyme has now accelerated. Approval of strand transfer inhibitors of integrase such as raltegravir as anti-HIV drugs provided clear validation of the efforts that have gone into studies of this enzyme. However, many questions, particularly those regarding the structure of the full-length HIV-1 IN and the multiprotein−DNA preintegration complexes, still remain to be answered. Integrase continues to be an important target for designing new anti-HIV drugs, and it is clear that further studies of its structure and function are warranted.

ACKNOWLEDGMENT

This project was supported in part by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Center for Cancer Research. The research of MJ was supported by a Faculty Scholar Fellowship from the National Cancer Institute and by a subsidy from the Foundation for Polish Science.

REFERENCES

1. Bao, K. K.; Wang, H.; Miller, J. K.; Erie, D. A.; Skalka, A. M.; Wong, I. Functional oligomeric state of avian sarcoma virus integrase. *J. Biol. Chem.* **2003**, 278, 1323−1327.
2. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res*. **2000**, 28, 235−242.
3. Brese, N. E.; O'Keeffe, M. Bond-valence parameters for solids. *Acta Crystallogr*. **1991**, B47, 192−197.
4. Bujacz, G.; Alexandratos, J.; Wlodawer, A.; Merkel, G.; Andrake, M.; Katz, R. A.; Skalka, A. M. Binding of different divalent cations to the active site of avian sarcoma virus integrase and their effects on enzymatic activity. *J. Biol. Chem.* **1997**, 272, 18161−18168.
5. Bujacz, G.; Alexandratos, J.; Zhou-Liu, Q.; Clement-Mella, C.; Wlodawer, A. The catalytic domain of human inmmunodeficiency virus integrase: Ordered active site in the F185H mutant. *FEBS Lett*. **1996**, 398, 175−178.
6. Bujacz, G.; Jaskólski, M.; Alexandratos, J.; Wlodawer, A.; Merkel, G.; Katz, R. A.; Skalka, A. M. High resolution structure of the catalytic domain of the avian sarcoma virus integrase. *J. Mol. Biol.* **1995**, 253, 333−346.
7. Bujacz, G.; Jaskólski, M.; Alexandratos, J.; Wlodawer, A.; Merkel, G.; Katz, R. A.; Skalka, A. M. The catalytic domain of avian sarcoma virus integrase: Conformation of the active-site residues in the presence of divalent cations. *Structure* **1996**, 4, 89−96.
8. Bushman, F. D.; Wang, B. Rous sarcoma virus integrase protein: Mapping functions for catalysis and substrate binding. *J. Virol.* **1994**, 68, 2215−2223.
9. Cai, M.; Huang, Y.; Caffrey, M.; Zheng, R.; Craigie, R.; Clore, G. M.; Gronenborn, A. M. Solution structure of the His12 --> Cys mutant of the N-terminal zinc binding domain of HIV-1 integrase complexed to cadmium. *Protein Sci.* **1998**, 7, 2669−2674.
10. Cai, M.; Zheng, R.; Caffrey, M.; Craigie, R.; Clore, G. M.; Gronenborn, A. M. Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nature Struct. Biol.* **1997**, 4, 567−577.
11. Camarasa, M. J.; Velazquez, S.; San Felix, A.; Perez-Perez, M. J.; Gago, F. Dimerization inhibitors of HIV-1 reverse transcriptase, protease and integrase: A single mode of inhibition for the three HIV enzymes? *Antiviral Res.* **2006**, 71, 260−267.
12. Chen, J. C.; Krucinski, J.; Miercke, L. J.; Finer-Moore, J. S.; Tang, A. H.; Leavitt, A. D.; Stroud, R. M. Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: A model for viral DNA binding. *Proc. Natl. Acad. Sci. USA* **2000**, 97, 8233−8238.
13. 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. 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. *J. Mol. Biol.* **2000**, 296, 521−533.
14. Cherepanov, P.; Ambrosio, A. L.; Rahman, S.; Ellenberger, T.; Engelman, A. Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. *Proc. Natl. Acad. Sci. USA* **2005**, 102, 17308−17313.
15. Christ, F.; Voet, A.; Marchand, A.; Nicolet, S.; Desimmie, B. A.; Marchand, D.; Bardiot, D.; Van, d., V; Van, R. B.; Strelkov, S. V.; De, M. M.; Chaltin, P.; Debyser, Z. Rational design of small-molecule inhibitors of the LEDGF/p75-integrase interaction and HIV replication. *Nat. Chem. Biol.* **2010**, 6, 442−448.
16. Coffin, J. M.; Hughes, S. H.; Varmus, H. E. *Retroviruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, **1997**.
17. Davies, D. R.; Goryshin, I. Y.; Reznikoff, W. S.; Rayment, I. Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science* **2000**, 289, 77−85.
18. DeLano, W. L. *The PyMOL Molecular Graphics System*. DeLano Scientific, San Carlos, CA, **2002**.
19. Deprez, E.; Tauc, P.; Leh, H.; Mouscadet, J. F.; Auclair, C.; Brochon, J. C. Oligomeric states of the HIV-1 integrase as measured by time-resolved fluorescence anisotropy. *Biochemistry* **2000**, 39, 9275−9284.
20. Dyda, F.; Hickman, A. B.; Jenkins, T. M.; Engelman, A.; Craigie, R.; Davies, D. R. Crystal structure of the catalytic domain of HIV-1 integrase: Similarity to other polynucleotidyl transferases. *Science* **1994**, 266, 1981−1986.
21. Eijkelenboom, A. P.; Lutzke, R. A.; Boelens, R.; Plasterk, R. H.; Kaptein, R.; Hård, K. The DNA-binding domain of HIV-1 integrase has an SH3-like fold. *Nature Struct. Biol.* **1995**, 2, 807−810.
22. Eijkelenboom, A. P.; Sprangers, R.; Hard, K.; Puras Lutzke, R. A.; Plasterk, R. H.; Boelens, R.; Kaptein, R. Refined solution structure of the C-terminal DNA-binding domain of human immunovirus-1 integrase. *Proteins* **1999**, 36, 556−564.
23. Eijkelenboom, A. P.; van den Ent, F. M.; Vos, A.; Doreleijers, J. F.; Hård, K.; Tullius, T. D.; Plasterk, R. H.; Kaptein, R.; Boelens, R. The solution structure of the amino-terminal HHCC domain of HIV-2 integrase: A three-helix bundle stabilized by zinc. *Curr. Biol.* **1997**, 7, 739−746.
24. Eijkelenboom, A. P.; van den Ent, F. M.; Wechselberger, R.; Plasterk, R. H.; Kaptein, R.; Boelens, R. Refined solution structure of the dimeric N-terminal HHCC domain of HIV-2 integrase. *J. Biomol. NMR* **2000**, 18, 119−128.
25. Gao, K.; Butler, S. L.; Bushman, F. Human immunodeficiency virus type 1 integrase: Arrangement of protein domains in active cDNA complexes. *EMBO J*. **2001**, 20, 3565−3576.

26. Goedken, E. R.; Marqusee, S. Native-state energetics of a thermostabilized variant of ribonuclease HI. *J. Mol. Biol.* **2001**, 314, 863−871.
27. Goldgur, Y.; Craigie, R.; Cohen, G. H.; Fujiwara, T.; Yoshinaga, T.; Fujishita, T.; Sugimoto, H.; Endo, T.; Murai, H.; Davies, D. R. Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: A platform for antiviral drug design. *Proc. Natl. Acad. Sci. USA* **1999**, 96, 13040−13043.
28. Goldgur, Y.; Dyda, F.; Hickman, A. B.; Jenkins, T. M.; Craigie, R.; Davies, D. R. Three new structures of the core domain of HIV-1 integrase: An active site that binds magnesium. *Proc. Natl. Acad. Sci. USA* **1998**, 95, 9150−9154.
29. Grandgenett, D. P.; Vora, A. C.; Schiff, R. D. A 32, 000-dalton nucleic acid-binding protein from avian retrovirus cores possesses DNA endonuclease activity. *Virology* **1978**, 89, 119−132.
30. Greenwald, J.; Le, V.; Butler, S. L.; Bushman, F. D.; Choe, S. The mobility of an HIV-1 integrase active site loop is correlated with catalytic activity. *Biochemistry* **1999**, 38, 8892−8898.
31. Hare, S.; Gupta, S. S.; Valkov, E.; Engelman, A.; Cherepanov, P. Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* **2010**, 464, 232−236.
32. Hare, S.; Shun, M.-C.; Gupta, S. S.; Valkov, E.; Engelman, A.; Cherepanov, P. A novel co-crystal structure affords the design of gain-of-function lentiviral integrase mutants in the presence of modified PSIP1/LEDGF/p75. *PLOS Pathogens* **2009**, 5, e1000259.
33. Hayouka, Z.; Rosenbluh, J.; Levin, A.; Loya, S.; Lebendiker, M.; Veprintsev, D.; Kotler, M.; Hizi, A.; Loyter, A.; Friedler, A. Inhibiting HIV-1 integrase by shifting its oligomerization equilibrium. *Proc. Natl. Acad. Sci. USA* **2007**, 104, 8316−8321.
34. Heuer, T. S.; Brown, P. O. Mapping features of HIV-1 integrase near selected sites on viral and target DNA molecules in an active enzyme-DNA complex by photo-cross-linking. *Biochemistry* **1997**, 36, 10655−10665.
35. Heuer, T. S.; Brown, P. O. Photo-cross-linking studies suggest a model for the architecture of an active human immunodeficiency virus type 1 integrase-DNA complex. *Biochemistry* **1998**, 37, 6667−6678.
36. Jayaram, M. The cis-trans paradox of integrase. *Science* **1997**, 276, 49−51.
37. Katz, R. A.; Skalka, A. M. The retroviral enzymes. *Annu. Rev. Biochem.* **1994**, 63, 133−173.
38. Leh, H.; Brodin, P.; Bischerour, J.; Deprez, E.; Tauc, P.; Brochon, J. C.; LeCam, E.; Coulaud, D.; Auclair, C.; Mouscadet, J. F. Determinants of Mg^{2+}-dependent activities of recombinant human immunodeficiency virus type 1 integrase. *Biochemistry* **2000**, 39, 9285−9294.
39. Lodi, P. J.; Ernst, J.; Kuszewski, J.; Hickman, A. B.; Engelman, A.; Craigie, R.; Clore, G. M.; Gronenborn, A. M. Solution structure of the DNA binding domain of HIV-1 integrase. *Biochemistry* **1995**, 34, 9826−9833.
40. Lubkowski, J.; Dauter, Z.; Yang, F.; Alexandratos, J.; Merkel, G.; Skalka, A. M.; Wlodawer, A. Atomic resolution structures of the core domain of avian sarcoma virus integrase and its D64N mutant. *Biochemistry* **1999**, 38, 13512−13522.
41. Lubkowski, J.; Yang, F.; Alexandratos, J.; Merkel, G.; Katz, R. A.; Gravuer, K.; Skalka, A. M.; Wlodawer, A. Structural basis for inactivating mutations and pH-dependent activity of avian sarcoma virus integrase. *J. Biol. Chem.* **1998**, 273, 32685−32689.
42. Lubkowski, J.; Yang, F.; Alexandratos, J.; Wlodawer, A.; Zhao, H.; Burke, T. R., Jr.; Neamati, N.; Pommier, Y.; Merkel, G.; Skalka, A. M. Structure of the catalytic domain of avian sarcoma virus integrase with a bound HIV-1 integrase-targeted inhibitor. *Proc. Natl. Acad. Sci. USA* **1998**, 95, 4831−4836.
43. Lutzke, R. A.; Eppens, N. A.; Weber, P. A.; Houghten, R. A.; Plasterk, R. H. Identification of a hexapeptide inhibitor of the human immunodeficiency virus integrase protein by using a combinatorial chemical library. *Proc. Natl. Acad. Sci. USA.* **1995**, 92, 11456−11460.
44. Lutzke, R. A.; Plasterk, R. H. HIV integrase: A target for drug discovery. *Genes Funct*. **1997**, 1, 289−307.
45. Lutzke, R. A.; Plasterk, R. H. 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.* **1998**, 72, 4841−4848.
46. Maignan, S.; Guilloteau, J. P.; Zhou-Liu, Q.; Clement-Mella, C.; Mikol, V. 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.* **1998**, 282, 359−368.
47. Makhija, M. T. Designing HIV integrase inhibitors—Shooting the last arrow. *Curr. Med. Chem.* **2006**, 13, 2429−2441.
48. Maroun, R. G.; Gayet, S.; Benleulmi, M. S.; Porumb, H.; Zargarian, L.; Merad, H.; Leh, H.; Mouscadet, J. F.; Troalen, F.; Fermandjian, S. Peptide inhibitors of HIV-1 integrase dissociate the enzyme oligomers. *Biochemistry* **2001**, 40, 13840−13848.
49. Molteni, V.; Greenwald, J.; Rhodes, D.; Hwang, Y.; Kwiatkowski, W.; Bushman, F. D.; Siegel, J. S.; Choe, S. Identification of a small-molecule binding site at the dimer interface of the HIV integrase catalytic domain. *Acta Crystallogr*. **2001**, D57, 536−544.
50. Nowotny, M.; Gaidamakov, S. A.; Crouch, R. J.; Yang, W. Crystal structures of RNase H bound to an RNA/DNA hybrid: Substrate specificity and metal-dependent catalysis. *Cell* **2005**, 121, 1005−1016.
51. Nowotny, M.; Yang, W. Stepwise analyses of metal ions in RNase H catalysis from substrate destabilization to product release. *EMBO J*. **2006**, 25, 1924−1933.
52. Ren, J.; Stammers, D. K. HIV reverse transcriptase structures: Designing new inhibitors and understanding mechanisms of drug resistance. *Trends Pharmacol. Sci.* **2005**, 26, 4−7.
53. Saenz, D. T.; Poeschla, E. M. FIV: From lentivirus to lentivector. *J. Gene Med.* **2004**, 6 (Suppl. 1), S95−104.

54. Sarafianos, S. G.; Das, K.; Hughes, S. H.; Arnold, E. Taking aim at a moving target: Designing drugs to inhibit drug-resistant HIV-1 reverse transcriptases. *Curr. Opin. Struct. Biol.* **2004**, 14, 716−730.

55. Sourgen, F.; Maroun, R. G.; Frere, V.; Bouziane, M.; Auclair, C.; Troalen, F.; Fermandjian, S. A synthetic peptide from the human immunodeficiency virus type-1 integrase exhibits coiled-coil properties and interferes with the in vitro integration activity of the enzyme. Correlated biochemical and spectroscopic results. *Eur. J. Biochem.* **1996**, 240, 765−773.

56. Steitz, T. A.; Steitz, J. A. A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. USA* **1993**, 90, 6498−6502.

57. Vondrasek, J.; van Buskirk, C. P.; Wlodawer, A. Database of three-dimensional structures of HIV proteinases. *Nature Struct. Biol.* **1997**, 4, 8.

58. Wang, J. Y.; Ling, H.; Yang, W.; Craigie, R. Structure of a two-domain fragment of HIV-1 integrase: Implications for domain organization in the intact protein. *EMBO J.* **2001**, 20, 7333− 7343.

59. Wielens, J.; Headey, S. J.; Jeevarajah, D.; Rhodes, D. I.; Deadman, J.; Chalmers, D. K.; Scanlon, M. J.; Parker, M. W. Crystal structure of the HIV-1 integrase core domain in complex with sucrose reveals details of an allosteric inhibitory binding site. *FEBS Lett.* **2010**, 584, 1455−1462.

60. Wlodawer, A.; Vondrasek, J. Inhibitors of HIV-1 protease: A major success of structure-assisted drug design. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, 27, 249−284.

61. Yang, W.; Lee, J. Y.; Nowotny, M. Making and breaking nucleic acids: Two-Mg^{2+}-ion catalysis and substrate specificity. *Mol. Cell* **2006**, 22, 5−13.

62. Yang, Z. N.; Mueser, T. C.; Bushman, F. D.; Hyde, C. C. Crystal structure of an active two-domain derivative of Rous sarcoma virus integrase. *J. Mol. Biol.* **2000**, 296, 535−548.

63. Zhao, L.; O'Reilly, M. K.; Shultz, M. D.; Chmielewski, J. Interfacial peptide inhibitors of HIV-1 integrase activity and dimerization. *Bioorg. Med. Chem. Lett.* **2003**, 13, 1175−1177.