Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry 12 (2004) 6371–6381
Bioorganic & Medicinal Chemistry

Active site binding modes of the β-diketoacids: a multi-active site approach in HIV-1 integrase inhibitor design

Raveendra Dayam and Nouri Neamati*

Department of Pharmaceutical Sciences, University of Southern California, School of Pharmacy, 1985 Zonal Avenue, PSC 304, Los Angeles, CA 90089, USA

Received 13 August 2004; revised 20 September 2004; accepted 21 September 2004

Abstract—Predicting a bioactive conformation of a ligand is of paramount importance in rational drug design. The task becomes very difficult when the receptor site possesses a region with unusual conformational flexibility. Significant conformational differences are present in the active site regions in the available crystal structures of the core domains of HIV-1 integrase (IN). Among all reported IN inhibitors, the β-diketoacid class of compounds has proved to be of most promise and indeed S-1360 was the first IN inhibitor to enter clinical studies. With an aim to predict the bioactive (active site bound) conformation of S-1360, we performed extensive docking studies using three different reported crystal structures where the active site or partial active site region was resolved. For comparison we extended our studies to include 5CITEP (the first compound cocrystallized with IN core domain) and a bis-diketoacid (BDKA). We found that the conformation of S-1360 when bound in one of the active sites matches that of the experimentally observed results of IN escape mutants resistant to S-1360. Therefore, we propose that this active site conformation is the biologically relevant conformation and can be used for the future structure-based drug design studies selectively targeting IN.

1. Introduction

The *pol* gene of HIV-1 encodes three essential enzymes: reverse transcriptase (RT), protease (PR), and integrase (IN). Currently 17 FDA approved drugs targeting RT and PR are available and are administered in various combinations. The combination of highly active antiretroviral therapy (HAART) in compliant patients come close to stopping virus evolution, however, eradication of the infection has not been achieved because of the persistence of latent HIV-1 in resting memory CD4+ T cells. Moreover, several factors including the emergence of multi-drug-resistant HIV strains, drug toxicity, the patient's ability to adhere to the prescribed therapy, and expensive medication have necessitated a reason to develop novel drugs, which target other viral replication processes.[1] IN is one such target, as it is a vital enzyme in the viral replication process and has no cellular homologue.

IN catalyses two crucial steps required for the integration of viral DNA into the host chromosome. In the first step, while in the cytoplasm of an infected cell, IN selectively cleaves two nucleotides (GT) from the 3′ terminal of the viral cDNA in a reaction known as 3′-processing.[2,3] Immediately after translocation to the nucleus as a component of the pre-integration complex, IN randomly inserts the viral cDNA into the host genome, and this reaction is referred to as strand transfer or integration.[4] Hence, the interactions of IN with viral cDNA are sequence specific, the interaction of IN with the host DNA is sequence independent. Consequently, their respective binding sites are expected to be geometrically and spatially different within the catalytic site of IN. DNA mapping and docking studies support that the viral cDNA and host DNA binding sites are indeed located on different parts of IN surface.[5,6] During the catalytic process, IN interacts with both the viral cDNA and the host DNA. IN has to undertake extensive conformational changes to ensure the favorable interactions with the substrates, the viral cDNA, and the target host DNA. These interactions are vital for the smooth and effective yield of the catalytic process. Divalent metals such as Mg^{2+} or Mn^{2+} are required for both 3′-processing and strand transfer activities of IN. The highly conserved

Keywords: HIV-1 integrase; Inhibitor design.
* Corresponding author. Tel.: +1 323 442 2341; fax: +1 323 442 1390; e-mail: neamati@usc.edu

0968-0896/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2004.09.035

DDE motif (a triad of acidic residues D64, D116, and E152 in HIV-1 IN) involves in the binding of metal cofactors. The coordination of metal cofactors with the DDE motif potentially induces certain conformational changes in the catalytically active region of IN, which in turn makes IN ready for its catalytic process.[7,8] Even though the metal cofactor has such a determinant role in the IN catalytic processes, the exact number of metal atoms involved is unknown.[9] Only one metal atom is present in all currently available X-ray structures of IN.

Many of the known PR and RT inhibitors were designed based on information from cocrystal structures. To date no X-ray structure of the entire IN with or without substrate DNA is available. Fifteen X-ray structures of the core domain, with and without intact C or N terminal domains have been solved (PDB accession codes: 1B92, 1B9F, 1B9F, 1BHL, 1BI4, 1BL3, 1BIS, 1BIU, 1BIZ, 1EX4, 1EXQ, 1ITG, 1K6Y, 1QS4, 2ITG).[10–17] An X-ray structure of the core domain of IN in complex with an inhibitor (5CITEP) was reported.[15] Of the several core domain structures available only two have completely resolved active sites (chain B in PDB 1BIS and chain C in PDB 1BL3).[14,17] Remarkable differences are observed in the conformation of the active site in reported X-ray structures of IN. A flexible loop consisting of amino acid residues 138–150 is one of the major structural components of the active site of IN. E152, one of the three components of the DDE motif resides on a helix near to the flexible loop. In many X-ray structures, including the first complex structure of IN with 5CITEP, a part of the flexible loop was not resolved. Furthermore, the position of the ligand in the IN–5CITEP complex is ambiguous. It is believed that crystal packing effects influenced the orientation and position of the ligand.[18–20] The conformation of the flexible loop in two X-ray structures where it is completely resolved, is quite different and which resulted in large variations in the conformation of the entire active site of these two substrate free structures of IN. The conformation of the flexible loop as well as the entire active site in the IN–5CITEP complex also largely differs from the above two structures. The active site regions from these three structures are considered for further analysis and are reported here as active site A (from chain A in PDB 1QS4), active site B (from chain B in PDB 1BIS), and active site C (from chain C in PDB 1BL3). The active sites from all these three crystal structures are complexed with an Mg^{2+} ion. The carbonyl oxygen atoms of residues D64 and D116 occupied two of the six coordination positions of the Mg^{2+} ion.

Previously, several attempts have been made to exploit the available information on the structure of IN in order to design inhibitors and most were performed using one of the above-mentioned active sites (for recent reviews see Refs. 21 and 22). In some studies, molecules were docked into the active site A where missing residues were added to complete the active site model, and the active site C was used in others studies.[20,23–26] We also used the crystal structure of a naphthalene disulfonate, Y3,[27] which we originally identified as HIV-1 IN inhibitor in complex with ASV integrase[28] for high throughput docking studies.[29,30] That study showed ASV integrase could serve as a surrogate for HIV-1 IN structure-based drug design. Attempts were also made to incorporate the dynamic nature of the active site into the rational design process.[31–33] The outcome of these studies helped to understand the complexity of the problem. However, in most studies the critical role of the metal ion in terms of enzyme inhibitor interactions were not addressed constructively. A recent molecular dynamics study on the 5CITEP–IN complex and a double mutated IN complexed with 5CITEP revealed that a region of the IN active site has high conformational flexibility and the region's conformational flexibility is very important for efficient biological activity of the IN.[34]

Apart from incomplete structural information of IN, the lack of a *bona fide* lead molecule also greatly hampered the drug discovery efforts against IN (for recent reviews see Refs. 22,35–37). Recent discovery of the β-diketoacids as IN inhibitors provided a valid lead molecule.[38] The β-diketoacid inhibitors emerged as a potent and promising class of compounds and two IN inhibitors entered clinical trials against HIV-1 infection. So far

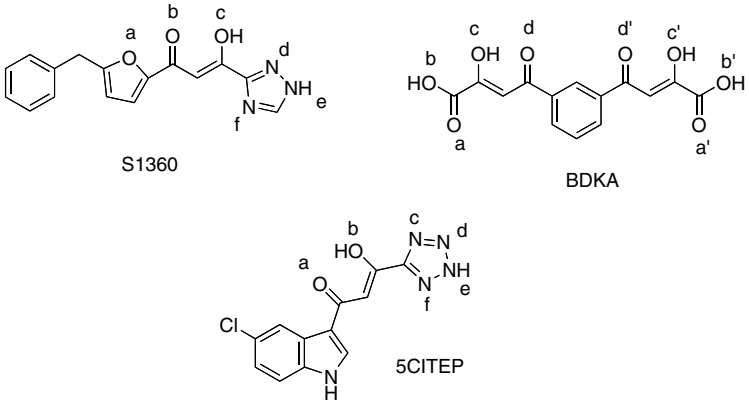

Figure 1. Structures of the β-diketoacids. Letters indicate the position of the H-bond donors or acceptors sites (see also Tables 2 and 3).

no high throughput docking studies using the β-diketo-acid inhibitors have been reported.

With an aim to predict the biologically relevant conformation of three reported IN inhibitors (see Fig. 1) when bound to the IN active site, we used GOLD software package[39] to dock these molecules into the IN active site. S-1360, an orally active IN inhibitor, is a bioisostere of β-diketoacid.[40] 5CITEP is a β-diketoacid bioisostere possessing a 5-chloro-indole moiety, and was cocrystallized with HIV-1 IN (PDB 1QS4).[14] Both 5CITEP and S-1360 are selective inhibitors of strand transfer reaction. BDKA with two β-diketoacid functional groups on the 1,3-positions of the central phenyl ring, inhibits both the 3′-processing and stand transfer reactions of IN with a comparable IC_{50} values.[26,41] BDKA adopted fairly similar orientation and occupied same location in the active sites from three crystal structures of IN while S-1360 adopted somewhat similar bound conformation in the active site A and C but obtained very different bound conformation in the active site B. Apart from the difference in bound orientation, S-1360 occupied quite different location in the active site B.

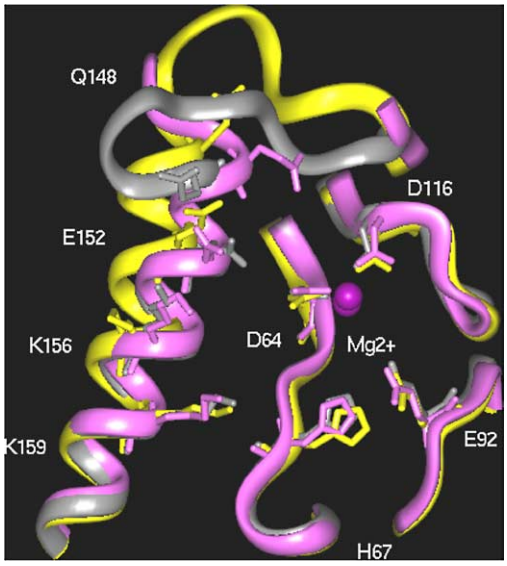

Figure 2. The superimposed backbone ribbon models of the active site regions from three different crystal structures. The pink (PDB 1QS4), yellow (PDB 1BIS), and gray (PDB 1BL3) ribbons represent the active site A, B, and C, respectively. The Mg^{2+} ions are shown in magenta.

2. Results and discussion

In our continuous efforts in developing second and third generation inhibitors based on S-1360 structure we wanted to understand three questions. (1) Why are some diketoacids selective for the strand transfer reaction? (2)[42] Why do very few diketoacids show antiviral activity, even though they have very similar physicochemical properties? Finally, (3) is there a unique IN–drug binding site for those diketoacid analogues that show antiviral activity? Although 15 crystal structures of IN currently exist, it is important to know which one to select for detailed structure-based drug design studies. Our primary objective was to identify the structure that most closely validates the biological observations from mutational studies. We will address these issues using three different IN crystal structures and three known inhibitors.

2.1. The dynamic nature of the active site that influences the DNA–IN and the drug–IN interactions

Of several available X-ray structures of the core domain of IN, two structures have a completely resolved core domain (B chain in PDB 1BIS and C chain in PDB 1BL3). Significant conformational differences were noticed in the flexible loop regions of the active sites of these two structures as well as the 5CITEP–IN crystal structure (PDB 1QS4).

Superimposed backbone ribbon models of the active site regions for the three different crystal structures are shown in Figure 2. The pink, yellow, and gray ribbons represent active sites A, B, and C, respectively. Some of the important active site residues are rendered as stick models and their color indicates the active site in which they reside. Active site B (yellow ribbon) has more room when compared to the other two active sites because of the expanded conformation of the flexible loop. In the other two active sites (A and C) the orientation of the flexible loop is very different from active site B.

We used inter-residue distances to explain certain structural differences that exist in the active site from three IN crystal structures. In Figure 3a, we show distances between backbone carbonyl oxygen of various active site amino acid residues and the Mg^{2+} ion. Large differences are observed in conformations of backbone of the flexible loop region of the active sites A, B, and C. The position of backbone carbonyl oxygen atoms in the active site B are represented by red full circles while the backbone carbonyl oxygen atoms in the active site A and C are represented by blue open circles and the olive diamonds, respectively. Especially, after K159, large deviations are observed in the position of the red full circles, olive diamonds, and the blue open circles, which indicate the flexible nature of this part of the active site (Fig. 3a).

The variation in the relative position and conformation of side chains of amino acid residues throughout IN active site in three crystal structures is depicted in Figure 3b. The distance between one of the terminal heavy atoms in side chain of various amino acid residues and the Mg^{2+} ion is used to distinguish the existing conformational and positional differences of amino acid side chains of IN active site in the three crystal structures. Remarkable differences in the position of the terminal heavy atoms of the amino acid side chains are found. Especially, for active site B, which differs a lot from the others, in particular the differences in the flexible loop region are quite distinct (compare the position of red full circles with the blue open circles and olive diamonds in Fig. 3b).

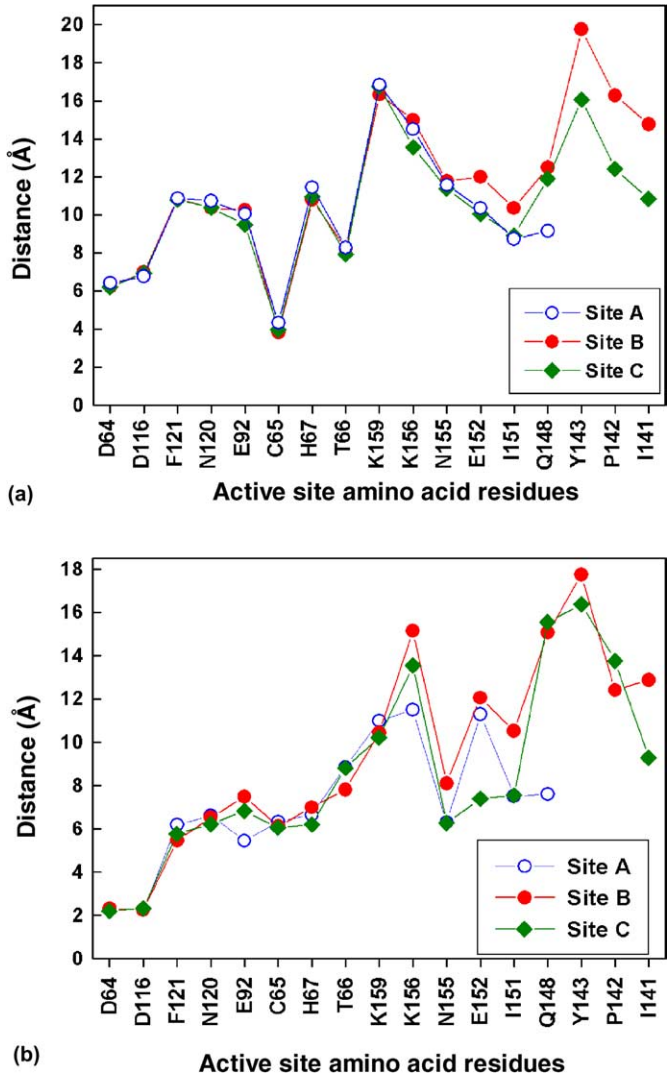

Figure 3. (a) The differences observed in conformation of the backbone of the active site regions in three different crystal structures of IN. The distance between backbone carbonyl oxygen atom of the active site amino acid residues and the Mg^{2+} ion (y-axis) is plotted against the individual active site amino acid residues. (b) The differences observed in the side chain conformations throughout the active site region in the three different crystal structures of IN. The distance between terminal heavy atom of the side chain of the active site amino acid residues and the Mg^{2+} ion (y-axis) is plotted against the individual active site amino acid residues.

The active site is located in the core domain of IN and the major part of the active site is spread on the surface of the core domain. The active site is open and has two cavities that extend perpendicularly to each other and provide a solvent accessible depth to the active site. A schematic representation of the IN active site is shown in Figure 4. The thick line shows the surroundings of the entire active site. One of the two cavities of active site that is covered by a part of the flexible loop (F139–I151) and spreads into an area close to α4 helix (E152 to K159) is designated as cavity ST of the active site. The thin line in Figure 4 covers cavity ST of the active site. The second cavity in the IN active site starts near the amino acid residue K159 and extends into an area surrounded by the amino acid residues D64, D116, N120, E92, and H67. The amino acid residues T66 and C65 provide bottom part to this cavity. This cavity is located almost perpendicular to cavity ST and it is referred as cavity 3P. We named these cavities based on a hypothesis that the 3′-processing reaction should be carried out in a region of the active site where a divalent metal ion is located. As the presence of a divalent metal ion is essential for the 3′-processing activity of the IN, it is believed that the 3′-processing reaction occurs in the proximity to the location of the divalent metal ion. Hence, we refer to this cavity as cavity 3P where metal ion is found in some of the reported crystal structures of IN. The dashed line in Figure 4 denotes cavity 3P of the active site. In its strand transfer reaction IN integrates the viral cDNA into the human DNA. Hence, IN inevitably interacts with both the human DNA and the viral cDNA at a time, which requires a lot of conformational flexibility in the active site of IN. We believe that the cavity which has a flexible loop as one of its edges should harbor the strand transfer reaction of IN so that the unusual flexible nature of the loop helps IN to hold two macromolecules. Furthermore, the region between F139 and K159 was identified as one of the IN peptide

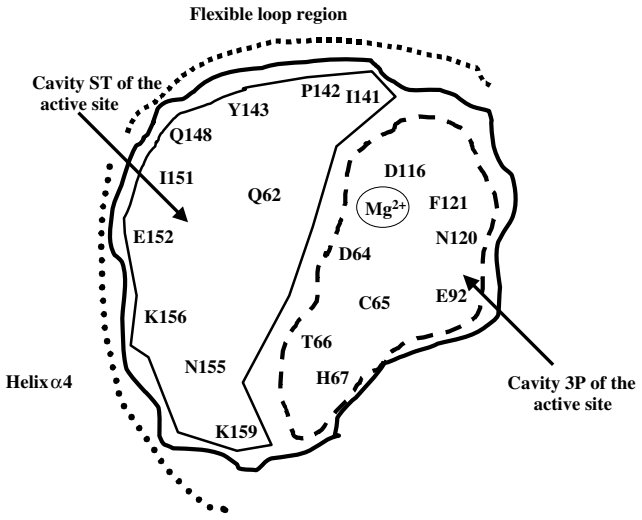

Figure 4. The schematic representation of the active site of HIV-1 IN. The amino acid residues involved in 3′-processing and strand transfer reactions are depicted in two closely juxtaposed cavities 3P and ST, respectively.

regions that interact with DNA.[43] Therefore we refer to this cavity as cavity ST (Fig. 4).

3. Docking studies

3.1. The bound conformation of S-1360 and its interactions with IN active site in three different crystal structures

In order to predict the binding mode of S-1360, 104 unique conformations within a 20 kcal/mol energy range were generated using catConf module of the Catalyst (Accelrys, Inc).[44] All the 104 conformations of S-1360 were docked into the active sites A, B, and C and 20 bound conformations were generated for each conformer in each active site. Based on the GOLD fitness score 10 conformers with high scores were selected for further analysis. S-1360 adopted quite different binding orientations inside the active sites A, B, and C.

In the active site A, the best bound conformation (the bound conformation with highest GOLD fitness score) is found 102 times out of 200 bound conformations (Table 1). In the bound conformation, S-1360 occupies a space near to residues D64, D116, N120, and Mg^{2+} ion in cavity 3P. The 4-flouro-phenyl ring of S-1360 is positioned in cavity ST of the active site A, in an area surrounded by residues K156, K159, and N155 (Tables 2 and 3). The bound orientations of S-1360 in the active

Table 1. Number of best bound conformations (total 200) of three diketoacid analogs

Compound	Active site A		Active site B		Active site C	
	No. Conf.	Score	No. Conf.	Score	No. Conf.	Score
S-1360	102	62	62	56	172	58
BDKA	22 (39)[a]	63 (60)[a]	60 (28)[a]	54 (50)[a]	70	55
5CITEP	99	52	45 (30)[a]	46 (44)[a]	55	47

[a] The second best bound conformation (the bound conformation with second highest GOLD score) and associated GOLD score.

Table 2. H-bonding interactions observed between β-diketoacid analogs and the active sites A, B, and C

Compound	Active site amino acid residues that contribute H-bonding interactions		
	Active site A	Active site B	Active site C
S1360	$O^{a}\cdots$HN C65 (2.83) $OH^{c}\cdots$O=C N120 (1.54) $N^{d}\cdots$HN N120 (1.99) F···HN K159 (2.45)	$OH^{c}\cdots$O=C D64 (3.09) $N^{d}\cdots$HN N155 (3.26) $NH^{e}\cdots$O=C V75 (1.69)	$C{=}O^{b}\cdots$HN H67 (3.51) $NH^{e}\cdots$O=C N120 (1.60)
BDKA	$O^{a}\cdots$HN H67 (2.30) $O^{b}\cdots$HN C65 (2.70) $OH^{c}\cdots$O=C D64 (2.10) $C{=}O^{d}\cdots$HN N120 (3.35) $C{=}O^{d'}\cdots$HN N117 (2.85) $OH^{c}\cdots$O=C E152 (3.38)[aa] $O^{b}\cdots$HN Q148 (2.14)[aa] $O^{b}\cdots$HN Q148 (2.15)[aa] $C{=}O^{d'}\cdots$HN H67 (2.07)[aa] $OH^{c'}\cdots$O=C E92 (1.53)[aa]	$OH^{c}\cdots$O=C D64 (2.17) $HO^{c}\cdots$HN N155 (2.46) $O^{b}\cdots$HN N155 (2.10) $O^{a}\cdots$HN K159 (1.78) $OH^{c'}\cdots$O=C D116 (2.16) $O^{a'}\cdots$HN N120 (3.14) $O^{a}\cdots$HN H67 (2.76)[aa]	$O^{a}\cdots$HN K159 (1.64) $HO^{c}\cdots$HN N155 (2.33) $OH^{c}\cdots$O=C E152 (3.04) $C{=}O^{d}\cdots$HN N155 (3.69) $O^{a'}\cdots$HN N120 (1.95) $O^{a'}\cdots$HN S119 (2.44) $O^{b'}\cdots$HO S119 (2.80)
5CITEP	$OH^{b}\cdots$O=C D116 (2.47) $N^{d}\cdots$HN S119 (2.68) $N^{d}\cdots$HN N120 (2.50) Cl···HN N155 (2.38)	$N^{e}\cdots$HN K159 (2.42) $NH^{g}\cdots$O=C E92 (1.98) $NH^{g}\cdots$OH T66 (2.29)[bb] $OH^{b}\cdots$O=C Q68 (1.53)[aa] $N^{e}\cdots$HN Q68 (1.55)[aa]	$OH^{b}\cdots$O=C D64 (1.95) $NH^{g}\cdots$O=C E92 (1.46) $NH^{d}\cdots$HN K159 (1.48) $NH^{e}\cdots$HN K159 (2.29)

The superscripted alphabet indicates the position of the H-bond donor or acceptor in a molecule (see Fig. 1) and the value in parenthesis is the H-bonding distance in Å. BDKA was docked as dianion and 5CITEP was docked as an anion.

[aa] The H-bonding interaction is observed between the second best bound conformation of the BDKA and the active site A and B.

[bb] The H-bonding interaction is observed between the second best bound conformation of the 5CITEP and the active site B.

Table 3. Ligand interactions with Mg^{2+}

Compound	Active site A	Active site B	Active site C
S1360	$O^{a}\cdots Mg^{2+}$ (1.75) $C{=}O^{b}\cdots Mg^{2+}$ (1.98)	—	$O^{a}\cdots Mg^{2+}$ (1.88) $HO^{c}\cdots Mg^{2+}$ (2.22)
BDKA	$O^{b}\cdots Mg^{2+}$ (2.08) $HO^{c}\cdots Mg^{2+}$ (2.18) $O^{b'}\cdots Mg^{2+}$ (2.38) $HO^{c'}\cdots Mg^{2+}$ (2.26) $O^{b'}\cdots Mg^{2+}$ (1.62)[aa]	$O^{b'}\cdots Mg^{2+}$ (1.90) $HO^{c'}\cdots Mg^{2+}$ (1.76) $O^{b}\cdots Mg^{2+}$ (1.89)[aa] $HO^{c}\cdots Mg^{2+}$ (1.46)[aa]	$C{=}O^{d'}\cdots Mg^{2+}$ (3.21) $C{=}O^{d'}\cdots Mg^{2+}$ (2.06)
5CITEP	$C{=}O^{a}\cdots Mg^{2+}$ (1.91) $HO^{b}\cdots Mg^{2+}$ (1.82)	$C{=}O^{a}\cdots Mg^{2+}$ (2.44)	$C{=}O^{a}\cdots Mg^{2+}$ (2.45)

The superscripted letters indicate the position of the H-bond donor or acceptor in a molecule (see Fig. 1). The value in parenthesis is the distance between donor atom and Mg^{2+} expressed in Å.

[aa] The metal–ligand interaction is observed with the second best bound conformation of BDKA in the active site A and B.

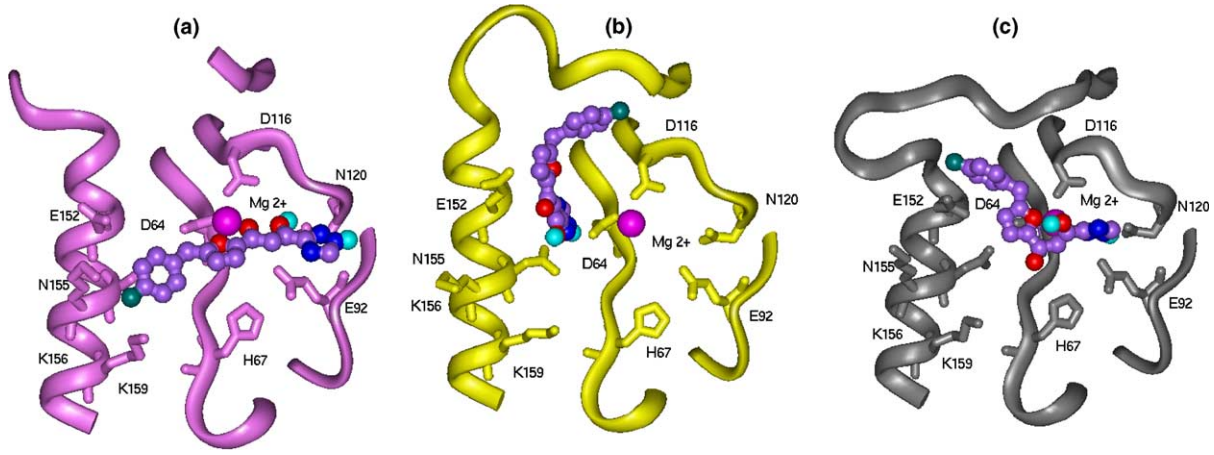

Figure 5. The predicted bound conformation of the S-1360 inside the (a) active site A, (b) active site B, and (c) active site C. The ribbon model shows the backbone of the active site region. All the prominent amino acid residues are rendered as stick models while the S-1360 is shown as a violet ball and stick model. The Mg^{2+} ions are shown in magenta.

sites A, B, and C are shown in Figure 5a–c. In the active site B the highly favorable bound conformation of S-1360 is found 62 times. S-1360 occupies a wide area in cavity ST and establishes favorable van der Waals interactions with various amino acid residues. The triazole and the diketoacid moiety of S-1360 occupy a deep cavity surrounded by residues I151, N155, V75, and the Q62. The triazole and the diketoacid and the furan groups involve favorable van der Waals as well as favorable electrostatic interactions with residues D64, I151, E152, and N155. The H-bonding interactions are observed between the enol hydroxyl, triazole of S-1360, and amino acid residues D64, V75, and N155 (Table 2). The furan ring and the 4-flouro-phenyl ring of S-1360 occupy a space near to residues I151, Q148, P142, I141, F139, D116, T115, and H114. The highly favorable binding conformation of S-1360 inside the active site C is found 172 times. The large part of the molecule occupies an area in cavity 3P while the 4-flouro-phenyl ring is placed in a space near to E152 in cavity ST. Two H-bonding interactions are observed between the carbonyl oxygen atom of the diketoacid moiety, triazole NH of S-1360, and the NH of residue H67 (C═O···H–N H67) and the side chain carbonyl oxygen of N120, respectively (Table 2).

We also compared the best binding orientations achieved by S-1360 in the active sites from the three different crystal structures of IN. In the active site A, S-1360 achieves a planar conformation and explores favorable interactions with various amino acid residues throughout the active site. In this bound orientation, S-1360 forms H-bonding interactions with residues K159 and N120 and two oxygen atoms (oxygen atom in the furan ring and keto group) form coordinate bonds with Mg^{2+} ion. However, it does not interact with amino acid residue E152 and energetically it appears to be a stable binding conformation (found 102 times) in this active site. The best binding site conformation of S-1360 inside the active site C is also very stable and it is found 172 times. Even though the predicted bound conformation of S-1360 in the active site A and C appears to be energetically stable, this bound conformation is not in line with experimentally observed results. S-1360 selectively inhibits strand transfer reaction of the IN but in its best bound conformation inside the active site A and C, S-1360 did not form noticeable interactions in cavity ST; on the other hand it formed very strong interactions with various amino acid residues and Mg^{2+} ion in cavity 3P where 3′-processing reaction of the IN is believed to be carried out.

A unique binding conformation for S-1360 in the active site B is predicted by GOLD. The entire S-1360 molecule fits in the cavity ST of the active site B. A part of the molecule, the triazole ring and the diketoacid moiety occupies a space deep inside of cavity ST, near to residues Q64, V75, N155, I151, and E152. The remaining part of the molecule, the furan ring and the 4-flourophenyl ring occupy the wide area near to residues Q148, P142, P139, I141, T115, and H114. Furthermore, in this bound orientation, S-1360 does not interact with the catalytic triad and the Mg^{2+} ion. As S-1360 was under clinical studies, it is worth to compare the docking results with the observed experimental results. S-1360 resistant mutants are isolated in vitro and the amino acid substitutions responsible for the drug resistance are identified. The three major resistant mutants are Q148K, I151L, and N155S.[40] All the three amino acid residues Q148, I151, and N155 are right in the active site, which is considered for the docking studies. It is interesting to note that in the highly favorable binding orientation in active site B, S-1360 occupies cavity ST of the active site and forms favorable van der Waals interactions with Q148, I151 amino acid residues, and an H-bonding interaction with N155, which are substituted by K, L, and S, respectively in S-1360 resistant mutants. The highly favorable active site binding orientation of the S-1360 inside the active site B has therefore biological relevance. Hence, it appears that the conformation of the active site B (PDB 1BIS) is the biologically active conformation of the active site of IN, wherein the flexible loop adopts an extended conformation.

3.2. The bound conformation of BDKA and its interactions with IN active site in three different crystal structures

Catalyst produced 227 unique conformations and all these conformers were docked into the above-mentioned active sites. The highly favorable bound conformation of BDKA is found 22 times in active site A and 60 times in active site B and 70 times in active site C out of the 200 bound conformations that are considered for analysis inside each active site (Table 1). The best bound orientations of BDKA inside the active sites A, B, and C are shown in Figure 6a–c. Of the 200 GOLD generated bound orientations, the best bound conformation of BDKA inside the active site A is found only 22 times, hence all the remaining bound conformations are analyzed to identify the other plausible binding orientations. The second best bound conformation of BDKA in active site A (the bound conformation which has second best fitness score) is shown as a green stick model in Figure 6a, found 39 times. In the first best bound conformation, BDKA occupies an area in 3P cavity of active site A and the two β-diketoacid wings of BDKA (denoted by letters a–d and a′–d′ in Fig. 1) interact with Mg^{2+} ion. The BDKA adopts a high-energy bent conformation inside active site A due to strong metal ligand interactions and which makes this bound conformation less abundant. In the second best bound orientation, one of the two β-diketoacid wings occupies ST cavity and the other wing occupies an area in 3P cavity near the Mg^{2+} ion. In active site B and C, BDKA adopts similar bound conformations wherein one of the two β-diketoacid wings occupies ST cavity while the other β-diketoacid wing occupies a place close to Mg^{2+} in 3P cavity. In ST cavity of active site B and C, one of the two β-diketoacid wings occupies an area surrounded by amino acid residues H67, K159, K156, N155, E152, and forms strong H-bonding interactions with E152, N155, and K159 (Table 2). While, the second diketoacid wing adopts slightly different conformation within the cavity 3P of active site B and C, and forms interactions with Mg^{2+} ion (Table 3). The oxygen atoms of the β-diketoacid wing form H-bonding interactions with residue N120 in active site B, and residues N120, S119 in active site C.

Even though certain differences are noticed in the conformations of one of the β-diketoacid wings of bound BDKA in cavity 3P, the overall binding conformation and binding location of BDKA inside active sites B and C are very similar, while BDKA adopted entirely different bound conformation in active site A. The

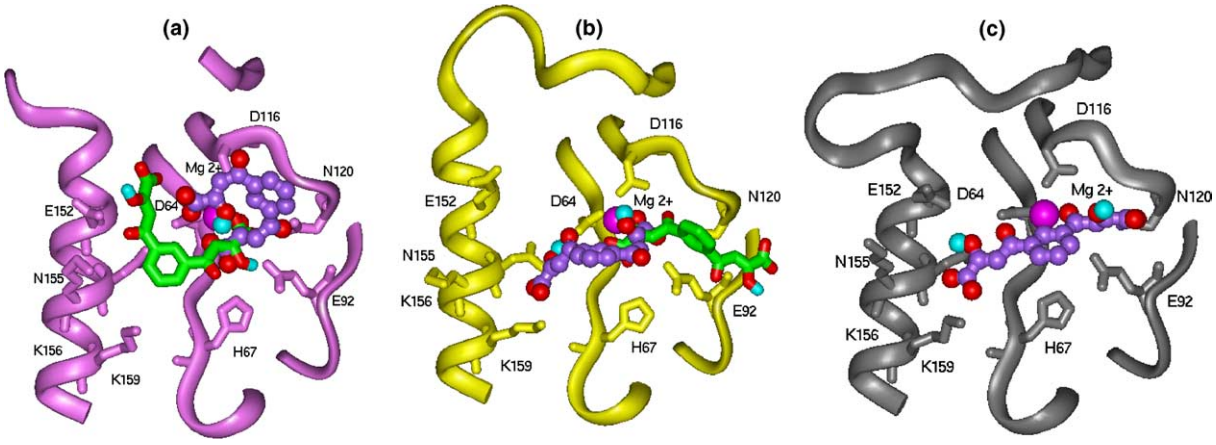

Figure 6. The predicted bound conformation of the BDKA inside the (a) active site A, (b) active site B, and (c) active site C. The ribbon model shows the backbone of the active site region. All the prominent amino acid residues are rendered as stick models while the BDKA is shown as a violet ball and stick model. The Mg^{2+} ions are shown in magenta. The second best bound conformation of the BDKA inside the active site A (a) and the active site B (b) is shown as a green stick model.

second best bound conformation of BDKA inside active site A is to a large extent similar to the best binding conformations in active site B and C. In the best bound conformation, BDKA could possess strong interactions with all the catalytically important amino acid residues and with Mg^{2+}. The predicted bound conformation of BDKA demonstrates its interactions with both cavity ST and cavity 3P and supports the experimentally observed inhibitory activities. BDKA inhibits the 3′-processing and strand transfer reactions of IN with a similar range of IC_{50} values.[26,41] This could be possible because of its ability to establish strong interactions with all the prominent amino acid residues on the IN active site such as D64, D116, E152, K156, K159, and Mg^{2+} which are believed to be participating in both 3′-processing and strand transfer activities of IN. Thus, docking results strongly support that the BDKA inhibits both 3′-processing and strand transfer reactions of IN by blocking catalytically vital regions of IN active site. Hence, it is feasible that an inhibitor which could interact or establish strong interactions with both cavities 3P and ST would inhibit both catalytic reactions of IN.

3.3. The bound conformation of 5CITEP and its interactions with IN active site in three different crystal structures

Catalyst generated 21 unique conformations for 5CITEP. The predicted bound conformations of 5CITEP in the active sites A and C are different from the observed bound conformation of 5CITEP in the IN–5CITEP complex crystal structure (PDB 1QS4). However, the second best bound conformation of 5CITEP in the active site B to some extent shares the binding location with bound location of 5CITEP in the IN–5CITEP complex crystal structure.

The highly favorable bound conformation of 5CITEP in the active site A is found 99 times out of the 200 predicted bound conformations (Table 1). The active site bound conformation of 5CITEP is shown in Figure 7a along with the orientation of 5CITEP in IN–5CITEP complex crystal structure. The orientation of 5CITEP in the crystal structure is represented by the green stick model while the violet ball and stick model represents the predicted bound orientation of 5CITEP in the active site A. In the active site A, the GOLD placed 5CITEP in both parts of IN active site so that the 5-chloro-indole moiety is placed in cavity ST and the β-diketoacid group and the tetrazole ring are positioned in cavity 3P. This bound orientation is almost perpendicular to the orientation of 5CITEP in the IN–5CITEP complex structure.

The highly favorable bound conformation of 5CITEP in the active site B is found 45 times and the second best bound conformation is observed 31 times and both these bound conformations are shown in Figure 7b. It is interesting to note that of the 600 bound orientations of 5CITEP that are generated by GOLD inside the active site of three different IN structures (200 conformations in each active site), none of them exactly resemble the orientation of 5CITEP in crystal structure. However, in the active site B, the second best bound location of 5CITEP is close to that of the 5CITEP in crystal structure but the relative orientation of various groups is different. The 5-chloro-indole moiety is placed near to residues K156, N155, E152, and I151 while the tetrazole and diketoacid groups are positioned in the proximity to residues I141, D116, T115, and H114 (green stick model in Fig. 7b). In this bound conformation no interactions are observed between the catalytic triad of IN and 5CITEP. Additionally, this binding location of 5CITEP in the active site B is close to that of S-1360 in the active site B. This similarity in the binding mode of S-1360 and 5CITEP inside active site B indicates that these molecules should proceed with similar mechanism of action. The highly favorable binding conformation of 5CITEP in the active site C is found 55 times and is to certain extent similar to the highly favorable bound orientation of 5CITEP in the active site B (Fig. 7c).

3.4. The differences in the predicted and X-ray structure conformation of 5CITEP

Of the 600 predicted bound conformations of 5CITEP none exactly resembles the bound conformation in X-ray structure of IN–5CITEP complex. The predicted

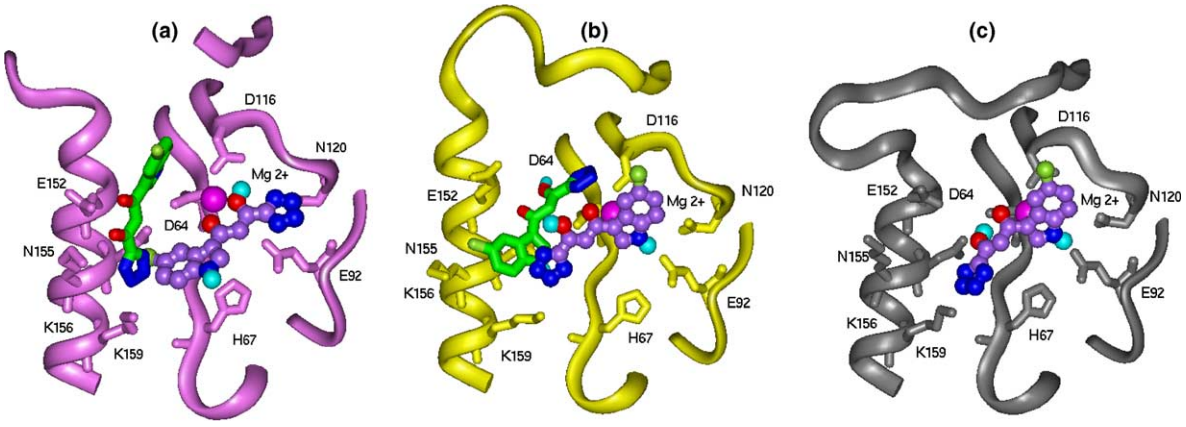

Figure 7. The predicted bound conformation of the 5CITEP inside the (a) active site A, (b) active site B, and (c) active site C. The ribbon model shows the backbone of the active site region. All the prominent amino acid residues are rendered as stick models while the 5CITEP is shown as a violet ball and stick model. The Mg^{2+} ions are shown in magenta. The green stick model in (a) shows the crystallographically determined bound orientation of 5CITEP (PDB 1QS4). The second best bound conformation of the 5CITEP inside the active site B (b) is shown as a green stick model.

bound conformations of 5CITEP in the active site A and C are very different from crystal structure conformation while the second best bound location of 5CITEP in active site B is close to that of the crystal structure. The orientation of 5CITEP in the IN–5CITEP complex crystal structure is ambiguous.[18–20] It is widely believed that the crystal packing effects influenced the position and orientation of 5CITEP in X-ray structure.[18–20,34,45,46] The molecular dynamics studies on the 5CITEP–IN complex and docking studies on IN core domain found different bound orientations for the 5CITEP inside the IN active site.[20,23,31,34] Similar bound conformation to the second best bound conformation of 5CITEP in active site B is found in a recent independent docking study wherein 5CITEP was docked onto snapshots of IN active site region that were collected from a molecular dynamics simulation.[45] This indicates that the availability of multiple conformations of the active site region allows 5CITEP to adopt a bound conformation free from crystal packing effects. The flexible loop in the active site A (a part of the loop is missing) and C adopted a closed conformation while in the active site B it adopted an extended (open) conformation. The cavity ST is more spacious in the active site B due to the open conformation of the flexible loop. The second best bound conformation of 5CITEP in the active site B is achieved due to the availability of an additional area in cavity ST on the other hand such bound conformations are not observed in active site A and C wherein the flexible loop adopted a closed conformation and which reduced the size of cavity ST. Therefore, the geometric and spatial constraints that exist in the respective active site regions in the three different crystal structures of IN forced 5CITEP to adopt energetically favorable bound conformations.

4. Conclusions

The multi-active site approach allows an understanding of how certain existing differences in enzyme active site influence the predicted binding site orientations of a ligand. Due to the fact that binding location of BDKA is not strongly influenced by the conformational flexibility of the flexible loop; it adopted fairly similar bound conformation in the active sites A, B, and C and established strong interactions with catalytically important residues in cavities ST and 3P. The predicted bound conformation strongly supports its in vitro activities in a sense that the BDKA inhibited both 3′-processing and strand transfer activities of IN within a similar range of potency. Therefore, the inhibitor that could form strong interactions with amino acid residues in both cavities ST and 3P of the IN active site would inhibit both 3′-processing and strand transfer reactions of IN. Interestingly, none of the GOLD predicted bound conformations of 5CITEP precisely resemble that of bound conformation in the reported crystal structure of the IN–5CITEP complex. The bound location of the 5CITEP in active site B (second best bound conformation) is close to the bound location of 5CITEP in the crystal structure. S-1360 adopted similar bound conformations in the active site A and C but it adopted quite different bound orientation in active site B. The bound orientation of the S-1360 in the active site B and the experimental results on drug resistant mutation studies indicate that the conformation of the active site B is biologically relevant active site conformation of IN in which the flexible loop adopted an extended conformation. The predicted bound orientations of BDKA, S-1360, and 5CITEP inside the active site B clearly distinguish their inhibitory potency and selectivity toward the IN catalytic reactions. Studies are underway to use active site B for high throughput docking using the multi-conformer databases to identify potent IN inhibitors.

5. Modeling procedures

5.1. Molecular structure and conformation of molecules

The structures of the two molecules (Fig. 1) were built and thoroughly minimized using the Build and Discover modules of the Insight II (Accelrys, Inc.)[47] running on a 24-processors Silicon Graphics Onyx instrument. For the 5CITEP, coordinates were directly taken from the X-ray structure of the IN–5CITEP complex (PDB 1QS4).[15] The well-minimized molecular structures were exported to Catalyst (Accelrys, Inc.) to generate a set of unique conformations that can explore the accessible conformational flexibility of a molecule. The poling algorithm implemented within Catalyst was used to generate conformations.[48] The poling algorithm promotes high conformational variation and assures broad coverage of low energy conformational space.[48,49] For each molecule all feasible unique conformations were generated over a 20 kcal/mol range of energies using the best flexible conformation generation method in Catalyst. This multi-conformer approach is adapted to provide a variety of starting molecular conformations for docking program.

5.2. Protein structures

Active sites from three different core domain X-ray structures of the IN are considered for docking purpose. The active sites of subunit A of the IN–5CITEP cocrystallized complex (PDB 1QS4), subunit B of the core domain X-ray structure in which all the active site amino acid residues are resolved (PDB 1BIS) and subunit C of the core domain in PDB 1BL3 were chosen for docking studies. The active site of the subunit A in PDB 1QS4 is complexed with a ligand while the active sites in subunit B of PDB 1BIS and subunit C of PDB 1BL3 are substrate free. A Mg^{2+} ion is placed in the active site B, between carbonyl oxygen atoms of amino acid residues D64 and D116 considering the geometry of the Mg^{2+} ion that is present in the active site A and C. All the water molecules present in proteins were removed and hydrogen atoms were added to the three proteins considering appropriate ionization states for both the acidic and basic amino acid residues at pH 7.0.

5.3. Docking

Docking was performed using version 1.2 of the GOLD (Genetic Optimization for Ligand Docking) software package running on an SGI Onyx instrument. GOLD is

an automated docking program that uses genetic algorithm to explore the ligand conformational flexibility with partial flexibility of the active site.[50] The algorithm was tested on a dataset of over 300 complexes extracted from the Brookhaven Protein Data Bank.[51] GOLD succeeded in more than 70% cases in reproducing the experimental bound conformation of the ligand. GOLD requires a user defined binding site. A 20 Å radius active site was defined considering the carboxyl carbonyl oxygen atom of residue D64 as the center of the active site. It searches for a cavity within the defined area and considers all the solvent accessible atoms in the defined area as active site atoms. All conformers of the three molecules were docked into the active sites A, B, and C. For each conformer 20 bound conformations were generated in each active site. Based on the GOLD fitness score 10 conformers with high fitness score were selected for analysis. All the bound conformations (200) from the 10 conformers were analyzed. The bound conformation with highest GOLD fitness score was identified and the number of times it occurred was reported for the three molecules in the active sites A, B, and C. All docking runs were carried out using standard default settings with a population size of 100, a maximum number of 100,000 operations, a mutation, and crossover rate of 95. At the end of each run, GOLD reports all the predicted bound conformations in an order of their rank based on the fitness score. The fitness function that is implemented in GOLD consists basically of H-bonding, complex energy, and ligand internal energy terms.

Acknowledgements

We thank Dr. Eric J. Chambers for helpful advice. This work was supported by funds from the Gustavus & Louise Pfeiffer Foundation and the GlaxoSmithKline Drug Discovery Award to N.N.

References and notes

1. Richman, D. D. HIV chemotherapy. *Nature* **2001**, *410*, 995–1001.
2. Brown, P. O. Integration. In *Retroviruses*; Coffin, J. C., Hughes, S. H., Varmus, H. E., Eds.; Cold Spring Harbor Press: Cold Spring Harbor, 1999.
3. Asante-Appiah, E.; Skalka, A. M. HIV-1 integrase: structural organization, conformational changes, and catalysis. *Adv. Virus Res.* **1999**, *52*, 351–369.
4. Engelman, A.; Mizuuchi, K.; Craigie, R. HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell* **1991**, *67*, 1211–1221.
5. Dirac, A. M.; Kjems, J. Mapping DNA-binding sites of HIV-1 integrase by protein footprinting. *Eur. J. Biochem.* **2001**, *268*, 743–751.
6. Perryman, A. L.; McCammon, J. A. AutoDocking dinucleotides to the HIV-1 integrase core domain: exploring possible binding sites for viral and genomic DNA. *J. Med. Chem.* **2002**, *45*, 5624–5627.
7. Asante-Appiah, E.; Skalka, A. M. A metal-induced conformational change and activation of HIV-1 integrase. *J. Biol. Chem.* **1997**, *272*, 16196–16205.
8. Bujacz, G.; Jaskolski, 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.
9. Grobler, J. A.; Stillmock, K.; Hu, B.; Witmer, M.; Felock, P.; Espeseth, A. S.; Wolfe, A.; Egbertson, M.; Bourgeois, M.; Melamed, J.; Wai, J. S.; Young, S.; Vacca, J.; Hazuda, D. J. Diketo acid inhibitor mechanism and HIV-1 integrase: implications for metal binding in the active site of phosphotransferase enzymes. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6661–6666.
10. 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.
11. Bujacz, G.; Alexandratos, J.; ZhouLiu, Q.; ClementMella, C.; Wlodawer, A. The catalytic domain of human immunodeficiency virus integrase: ordered active site in the F185H mutant. *FEBS Lett.* **1996**, *398*, 175–178.
12. 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.
13. 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. U.S.A.* **2000**, *97*, 8233–8238.
14. 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. U.S.A.* **1998**, *95*, 9150–9154.
15. 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. U.S.A.* **1999**, *96*, 13040–13043.
16. 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.
17. 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.
18. Keseru, G. M.; Kolossvary, I. Fully flexible low-mode docking: application to induced fit in HIV integrase. *J. Am. Chem. Soc.* **2001**, *123*, 12708–12709.
19. Sotriffer, C. A.; Ni, H.; McCammon, A. J. HIV-1 integrase inhibitor interactions at the active site: prediction of binding modes unaffected by crystal packing. *J. Am. Chem. Soc.* **2000**, *122*, 6136–6137.
20. Sotriffer, C. A.; Ni, H.; McCammon, J. A. Active site binding modes of HIV-1 integrase inhibitors. *J. Med. Chem.* **2000**, *43*, 4109–4117.
21. Neamati, N. Structure-based HIV-1 integrase inhibitor design: a future perspective. *Expert Opin. Inv. Drugs* **2001**, *10*, 281–296.
22. Dayam, R.; Neamati, N. Small-molecule HIV-1 integrase inhibitors: the 2001–2002 update. *Curr. Pharm. Des.* **2003**, *9*, 1789–1802.
23. Keseru, G. M.; Kolossvary, I. Fully flexible low-mode docking: application to induced fit in HIV integrase. *J. Am. Chem. Soc.* **2001**, *123*, 12708–12709.
24. Sotriffer, C. A.; Ni, H. H.; McCammon, J. A. HIV-1 integrase inhibitor interactions at the active site:

prediction of binding modes unaffected by crystal packing. *J. Am. Chem. Soc.* **2000**, *122*, 6136–6137.
25. Buolamwini, J. K.; Assefa, H. CoMFA and CoMSIA 3D QSAR and docking studies on conformationally-restrained cinnamoyl HIV-1 integrase inhibitors: exploration of a binding mode at the active site. *J. Med. Chem.* **2002**, *45*, 841–852.
26. Pais, G. C. G.; Zhang, X. C.; Marchand, C.; Neamati, N.; Cowansage, K.; Svarovskaia, E. S.; Pathak, V. K.; Tang, Y.; Nicklaus, M.; Pommier, Y.; Burke, T. R. Structure activity of 3-aryl-1,3-diketo-containing compounds as HIV-1 integrase inhibitors. *J. Med. Chem.* **2002**, *45*, 3184–3194.
27. Nicklaus, M. C.; Neamati, N.; Hong, H.; Mazumder, A.; Sunder, S.; Chen, J.; Milne, G. W.; Pommier, Y. HIV-1 integrase pharmacophore: discovery of inhibitors through three-dimensional database searching. *J. Med. Chem.* **1997**, *40*, 920–929.
28. 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. U.S.A.* **1998**, *95*, 4831–4836.
29. Chen, I. J.; Neamati, N.; Nicklaus, M. C.; Orr, A.; Anderson, L.; Barchi, J. J., Jr.; Kelley, J. A.; Pommier, Y.; MacKerell, A. D., Jr. Identification of HIV-1 integrase inhibitors via three-dimensional database searching using ASV and HIV-1 integrases as targets. *Bioorg. Med. Chem.* **2000**, *8*, 2385–2398.
30. Chen, I. J.; Neamati, N.; Mac Kerell, A. D., Jr. Structure-based inhibitor design targeting HIV-1 integrase. *Curr. Drug Targets Infect. Disord.* **2002**, *2*, 217–234.
31. Ni, H. H.; Sotriffer, C. A.; McCammon, J. A. Ordered water and ligand mobility in the HIV-1 integrase–5CITEP complex: a molecular dynamics study. *J. Med. Chem.* **2001**, *44*, 3043–3047.
32. Carlson, H. A.; Masukawa, K. M.; Rubins, K.; Bushman, F. D.; Jorgensen, W. L.; Lins, R. D.; Briggs, J. M.; McCammon, J. A. Developing a dynamic pharmacophore model for HIV-1 integrase. *J. Med. Chem.* **2000**, *43*, 2100–2114.
33. Lins, R. D.; Briggs, J. M.; Straatsma, T. P.; Carlson, H. A.; Greenwald, J.; Choe, S.; McCammon, J. A. Molecular dynamics studies on the HIV-1 integrase catalytic domain. *Biophys. J.* **1999**, *76*, 2999–3011.
34. Barreca, M. L.; Lee, K. W.; Chimirri, A.; Briggs, J. M. Molecular dynamics studies of the wild-type and double mutant HIV-1 integrase complexed with the 5CITEP inhibitor: mechanism for inhibition and drug resistance. *Biophys. J.* **2003**, *84*, 1450–1463.
35. Neamati, N. Patented small molecule inhibitors of HIV-1 integrase: a 10-year saga. *Expert Opin. Ther. Pat.* **2002**, *12*, 709–724.
36. Johnson, A. A.; Marchand, C.; Pommier, Y. HIV-1 integrase inhibitors: a decade of research and two drugs in clinical trial. *Curr. Top. Med. Chem.* **2004**, *4*, 1059–1077.
37. Billich, A. S-1360 Shionogi-GlaxoSmithKline. *Curr. Opin. Investig. Drugs* **2003**, *4*, 206–209.
38. Hazuda, D. J.; Felock, P.; Witmer, M.; Wolfe, A.; Stillmock, K.; Grobler, J. A.; Espeseth, A.; Gabryelski, L.; Schleif, W.; Blau, C.; Miller, M. D. Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. *Science* **2000**, *287*, 646–650.
39. GOLD Cambridge Crystallographic Data Centre (CCDC), V1.2, UK, 2001.
40. Yoshinaga, T. S. A.; Fujishita, T.; Fujiwara, T. In vitro activity of a new HIV-1 integrase inhibitor in clinical development. In 9th Conference on Retroviruses and Opportunistic Infections, Seattle, USA, 2002.
41. Marchand, C.; Zhang, X. C.; Pais, G. C. G.; Cowansage, K.; Neamati, N.; Burke, T. R.; Pommier, Y. Structural determinants for HIV-1 integrase inhibition by beta-diketo acids. *J. Biol. Chem.* **2002**, *277*, 12596–12603.
42. Long, Y. Q.; Jiang, X.-H.; Dayam, R.; Sanchez, T.; Shoemaker, R.; Sei, S.; Neamati, N. Rational design and synthesis of novel dimeric diketoacid containing inhibitors of HIV-1 integrase: implication for binding to two metal ions on the active site of integrase. *J. Med. Chem.* **2004**, *47*, 2561–2573.
43. 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.
44. Accelrys, Inc. Catalyst; V.4.7, San Diego.
45. Schames, J. R.; Henchman, R. H.; Siegel, J. S.; Sotriffer, C. A.; Ni, H.; McCammon, J. A. Discovery of a novel binding trench in HIV integrase. *J. Med. Chem.* **2004**, *47*, 1879–1881.
46. Ni, H.; Sotriffer, C. A.; McCammon, J. A. Ordered water and ligand mobility in the HIV-1 integrase–5CITEP complex: a molecular dynamics study. *J. Med. Chem.* **2001**, *44*, 3043–3047.
47. Accelrys, I. InsightII, San Diego.
48. Smellie, A.; Teig, S. L.; Towbin, P. Poling-promoting conformational variation. *J. Comput. Chem.* **1995**, *16*, 171–187.
49. Smellie, A.; Kahn, S. D.; Teig, S. L. Analysis of conformational coverage. 1. Validation and estimation of coverage. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 285–294.
50. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748.
51. Nissink, J. W.; Murray, C.; Hartshorn, M.; Verdonk, M. L.; Cole, J. C.; Taylor, R. A new test set for validating predictions of protein–ligand interaction. *Proteins* **2002**, *49*, 457–471.