Active Site Binding Modes of HIV-1 Integrase Inhibitors

Christoph A. Sotriffer,* Haihong Ni, and J. Andrew McCammon

Departments of Chemistry and Biochemistry and of Pharmacology, Howard Hughes Medical Institute, University of California, San Diego, La Jolla, California 92093-0365

Received May 4, 2000

Using the crystal structure of the first complex of the HIV-1 integrase catalytic core domain with an inhibitor bound to the active site, structural models for the interaction of various inhibitors with integrase were generated by computational docking. For the compound of the crystallographic study, binding modes unaffected by crystal packing have recently been proposed. Although a large search region was used for the docking simulations, the ligands investigated here are found to bind preferably in similar ways close to the active site. The binding site is formed by residues 64−67, 116, 148, 151−152, 155−156, and 159, as well as by residue 92 in case of the largest ligand of the series. The coherent picture of possible interactions of small-molecule inhibitors at the active site provides an improved basis for structure-based ligand design. The recurring motif of tight interaction with the two lysine residues 156 and 159 is suggested to be of prime importance.

Introduction

HIV-1 integrase (IN) is responsible for the integration of reversely transcribed viral DNA into host cell DNA. It catalyzes two reactions referred to as 3′-processing and strand transfer.[1−3] In 3′-processing two nucleotides are cleaved from the 3′-end of the viral DNA. The processed ends subsequently undergo strand transfer, a transesterification reaction in which the 3′-viral ends are covalently joined to 5′-phosphates in host DNA at the site of integration. As these reactions are essential for the life cycle of the virus, IN represents an attractive target for the treatment of HIV infections. However, no clinically useful inhibitor seems to be available yet, even though a large number of compounds that inhibit IN have already been identified.[4] Accordingly, current AIDS drugs target only two HIV enzymes, reverse transcriptase and protease, while IN remains still unexploited.

A problem for the design of strong and selective IN inhibitors is the lack of detailed structural information about the interaction between IN and its substrates or inhibitors. So far, drug design efforts have mostly relied on pharmacophore hypotheses derived from the structures of known inhibitors.[5−8] Recently the first receptor-based pharmacophore model has been presented, developed according to a new "dynamic pharmacophore technique" to incorporate effects of conformational flexibility at the active site.[9−11] Although these models were successful in discovering some new inhibitors by database searches, uncertainties remain about the details of the protein−ligand interaction and the effects of the molecular framework not covered by the pharmacophore model.

The situation has somewhat improved with the appearance of the first complex structure of the IN catalytic core domain with an inhibitor bound to the active site,[12] allowing first insights about the actual interaction sites and representing a starting point for structure-based design efforts. However, the information provided by this structure is not unequivocal since the position of the ligand appears to be significantly influenced by a crystal packing effect and could be determined for only one of the three monomers in the asymmetric unit. In a recent study, we could provide support for this assumption, showing that only in the presence of the symmetry-related neighbor the ligand position observed in the crystal is energetically preferred, whereas different orientations are favored if the crystal environment is not taken into account.[13]

We now expand our initial study on ligand binding by IN to address the potential binding modes of various other IN inhibitors for which no experimental structures are available, in an attempt to derive guidelines for the design of new inhibitory compounds. For this purpose, comparative docking studies on the IN catalytic core domain have been performed using the latest version of the AutoDock program with an empirical free energy function and a Lamarckian genetic search algorithm.[14] With recent improvements in search algorithms and energy functions, computational docking methods have become a valuable tool to probe the interaction between an enzyme and its inhibitors in the absence of detailed experimental data and can contribute significantly to the understanding of its structural and energetic basis.[14−19]

The structures of the investigated ligands are shown in Figure 1. 1-(5-Chloroindol-3-yl)-3-(tetrazolyl)-1,3-propanedione enol (5CITEP) is the inhibitor for which the crystal structure of the IN complex has been obtained.[12] Due to the available experimental information and our previous findings for this compound, it will serve as a reference for the analysis of the other inhibitors: 1,2,5,8-tetrahydroxyanthraquinone (quinalizarin, QLZ), 3,3,3′,3′-tetramethyl-1,1′-spirobis(indan)-5,5′,6,6′-tetrol (TMS), L-dicaffeoyltartaric acid (L-chicoric acid, LCH), and 4-acetylamino-5-hydroxynaphthalene-2,7-disulfonic acid (abbreviated as Y-3 by Lubkowski et al.[20]). QLZ is

* Corresponding author. Phone: 858-822-0255. Fax: 858-534-7042. E-mail: csotriff@mccammon.ucsd.edu.

10.1021/jm000194t CCC: $19.00
Published on Web 10/12/2000

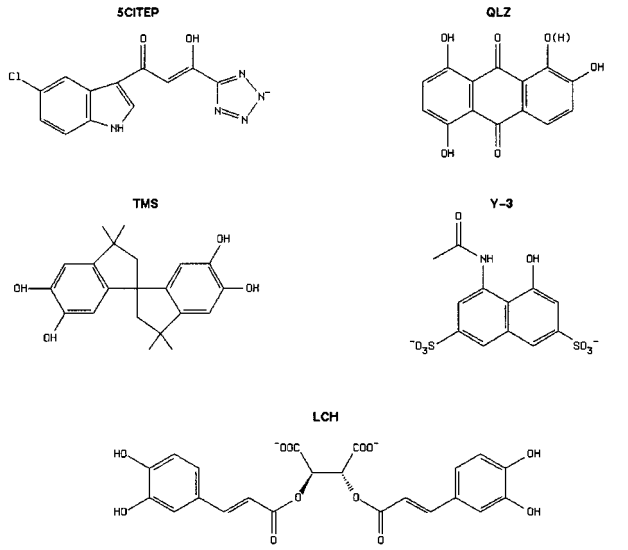

Figure 1. Structures of the investigated ligands.

among the most active inhibitors in the class of hydroxylated aromatics and anthraquinones.[21,22] TMS is the lead compound of a recently presented new class of IN inhibitors;[23] its rigid and bulky nature makes it an interesting target for docking studies. LCH shows submicromolar IC_{50} values for inhibition of 3′-processing and strand transfer; it is one of the most potent IN inhibitors discovered so far and one of the few which are also active against virus in vivo. Y-3, finally, is an inhibitor both of IN and of avian sarcoma virus (ASV) integrase, and a crystal structure is available for the complex with ASV integrase.[20]

Methods

Docking. Docking was performed with version 3.0 of the program AutoDock.[14] It combines a rapid energy evaluation through precalculated grids of affinity potentials with a variety of search algorithms to find suitable binding positions for a ligand on a given protein. While the protein is required to be rigid, the program allows torsional flexibility in the ligand.

Docking to IN was carried out using the new empirical free energy function and the Lamarckian genetic algorithm, applying a standard protocol, with an initial population of 50 randomly placed individuals, a maximum number of 1.5×10^6 energy evaluations, a mutation rate of 0.02, a crossover rate of 0.80, and an elitism value of 1. Proportional selection was used, where the average of the worst energy was calculated over a window of the previous 10 generations. For the local search, the so-called pseudo-Solis and Wets algorithm was applied using a maximum of 300 iterations per local search. The probability of performing local search on an individual in the population was 0.06, and the maximum number of consecutive successes or failures before doubling or halving the local search step size was 4.

50 independent docking runs were carried out for each ligand. Results differing by less than 1.5 Å in positional root-mean-square deviation (rmsd) were clustered together and represented by the result with the most favorable free energy of binding.

Ligand Setup. The structures of the ligands were generated with QUANTA and optimized with the implemented CHARMm force field (programs distributed by Molecular Simulations Inc.). Coordinates for Y-3 were taken directly from the X-ray structure of the ASV complex (PDB 1A5V).[20] For the chiral spiro compound TMS, both enantiomers were generated and analyzed by docking. All ligands were modeled in the protonation state shown in Figure 1, with QLZ handled in both the neutral and deprotonated forms at the potentially acidic position 1 (the pK_a for the hydroxy group in position 1 has been calculated to be between 5 and 7; P. I. W. de Bakker and W. Weber, unpublished results). Atomic charges were assigned using the Gasteiger−Marsili formalism,[24] which is the type of atomic charges used in calibrating the AutoDock empirical free energy function.[14] Finally, the compounds were setup for docking with the help of AutoTors, the main purpose of which is to define the torsional degrees of freedom to be considered during the docking process. Besides the hydroxyl rotors, the following number of flexible torsions were defined for each ligand: 1 in 5CITEP, 0 in QLZ and TMS, 5 in LCH, and 2 in Y-3.

Table 1. Results of 50 Independent Docking Runs for Each Ligand[a]

ligand	N_{tot}	f_{occ}	ΔG_{bind}	contacting residues
QLZ(0)	6	20	−6.6	D64, C65, **T66**, H67, **D116**, E152, N155, **K159**
QLZ(−1)	5	25	−7.1	D64, **C65**, T66, **D116**, E152, N155, K159
(*S*)-TMS	4	38	−7.2	**T66**, **Q148**, **E152**, N155, K156
(*R*)-TMS	5	14	−7.3	**D64**, **C65**, **T66**, H67, **Q148**, I151, E152, N155, K156, K159
LCH	13	18	−7.5	C65, T66, **H67**, Q92, **Q148**, E152, **K156**, K159
Y-3	5	10	−6.9	D64, C65, T66, H67, D116, N155, **K156**, **K159**
5CITEP	5	36	−6.9	D64, T66, H67, **D116**, Q148, E152, K159

[a] N_{tot} is the total number of clusters; the number of results in the top cluster is given by the frequency of occurrence, f_{occ}; ΔG_{bind} is the estimated free energy of binding for the top cluster results and is given in kcal/mol. The last column shows the contacting residues for the binding mode of the top cluster. Only residues with at least 5 van der Waals contacts to the ligand are shown. Residues that form hydrogen bonds with the ligand are highlighted in bold.

Protein Setup. For the IN catalytic core domain, the structure obtained by X-ray analysis of the complex with 5CITEP was used (PDB 1QS4).[12] For the purpose of docking, subunit A was selected, which is the only monomer in the asymmetric unit where the position of the ligand could be determined. The missing residues at positions 141−144 in this subunit were incorporated from monomer B of the IN structure PDB 1BIS[25] after superposition of the backbones of residues 135−140 and 145−150.

The structure was setup for docking as follows: polar hydrogens were added using the PROTONATE utility (written by D. A. Case, K. Cross, and G. P. Gippert and distributed with AutoDock[14] and AMBER[26]). Histidine protonation states were modeled as in related MD simulation studies.[27] To optimize the inserted loop residues and the hydrogen positions, the structure was subjected to a short energy minimization using the SANDER program of AMBER5.0[26] and the AMBER united atom force field,[28] in accordance with the type of force field and protein charges of the AutoDock empirical free energy function. Solvation parameters were added to the final protein file using the ADDSOL utility of AutoDock3.0.

The grid maps representing the protein in the actual docking process were calculated with AutoGrid. The grids (one for each atom type in the ligand, plus one for electrostatic interactions) were chosen to be sufficiently large to include not only the active site but also significant portions of the surrounding surface. The dimensions of the grids were thus 30 Å × 30 Å × 30 Å, with a spacing of 0.375 Å between the grid points and the center close to C^β of the catalytic residue Asp 64.

Results

Docking of the compounds shown in Figure 1 revealed a consistent set of recurring binding modes. For all investigated ligands well-clustered docking results could be obtained. As shown in Table 1, the 50 independent

docking runs carried out for each ligand generally converged to a small number of different positions ("clusters" of results differing by less than 1.5 Å rmsd). A higher-than-average number of clusters is reported only for the larger and more flexible ligand LCH, but even here most of the results are found in the top two clusters (18 in the top cluster and 17 in the cluster ranked second). Generally, the top clusters (i.e. those with the most favorable ΔG_{bind}) are also associated with the highest frequency of occurrence which suggests a good convergence behavior of the search algorithm. As far as the estimated free energies of binding are concerned, they compare well with experimental inhibition constants, correctly suggesting affinities in the low-micromolar range. The best results in terms of free energy of binding are all located in a similar position at the active site. The most important interactions are summarized in Table 1, and a graphical representation of the binding modes is given in Figures 2 and 3 (coordinates in PDB format are available from the authors. For each inhibitor, the results are briefly described in the following section.

QLZ. The most favorable result for QLZ is placed "horizontally" within the active site (i.e. perpendicular to the direction of the Glu 152 side chain) and interacts with all three catalytic residues. The neutral form reaches an estimated ΔG of −6.6 kcal/mol. Three of the four hydroxy groups are involved in hydrogen bonds with the ammonium of Lys 159, the side chain oxygen of Thr 66, and the carboxylate of Asp 116. The oxygen of the hydroxyl interacting with Asp 116 and the adjacent quinone carbonyl oxygen coordinate with the Mg^{2+} ion at a distance of 2.5 and 3.1 Å, respectively. The second carbonyl of the ligand is close to the Asn 155 side chain amide.

Although an identical binding position is also observed for singly deprotonated QLZ (−7.0 kcal/mol), an alternative but largely overlapping position is preferred (−7.1 kcal/mol) and found most frequently (in 25 runs out of 50). The aromatic system occupies the same plane, but the functional groups are oriented differently. Together with one of the carbonyl oxygens, the deprotonated oxygen is coordinated with the Mg^{2+} ion. The hydroxyl ortho to the deprotonated site makes a hydrogen bond with Asp 116. Regarding the remaining two hydroxy groups, a weak hydrogen bond is observed between the backbone carbonyl of Cys 65 and the OH in position 8.

Neither of the two orientations found for QLZ fully explores all potential interaction sites. While in one case major contributions arise from interactions with Lys 159, Thr 66, and His 67, in the other significant energy is gained by placing the area of negative charge density closer to the Mg^{2+} ion. With the probable predominance of the deprotonated form, however, the second variant might be the more relevant binding mode of QLZ.

TMS. TMS is considerably bulkier than the planar system of QLZ (and 5CITEP), which could allow this compound to exploit additional binding regions. Being a chiral spiro compound, both enantiomers were considered for docking (cf. Figures 2 and 3). For the (*S*)-enantiomer, very clear preference for a single position in the active site could be obtained: the result with top binding energy (−7.2 kcal/mol) was found 38 times out of 50. The structure of the ligand fits in a nearly ideal way around the protruding side chain of Glu 152. Most of the van der Waals (vdW) contacts are formed with this residue and Asn 155. No significant interactions are observed with the other two catalytic residues, Asp 64 and Asp 116. Instead, the four hydroxy groups are involved in hydrogen bonds with Thr 66 (with the side chain oxygen), Gln 148 (with the backbone carbonyl and the side chain NH_2), and Glu 152 (with the backbone carbonyl). The side chains of Lys 156 and Lys 159 are somewhat too far away for significant hydrogen-bonding contributions but could, in principle, adopt a slightly different conformation to increase the strength of interaction.

The preference for a single binding position is less clear in case of the (*R*)-enantiomer. Besides the top

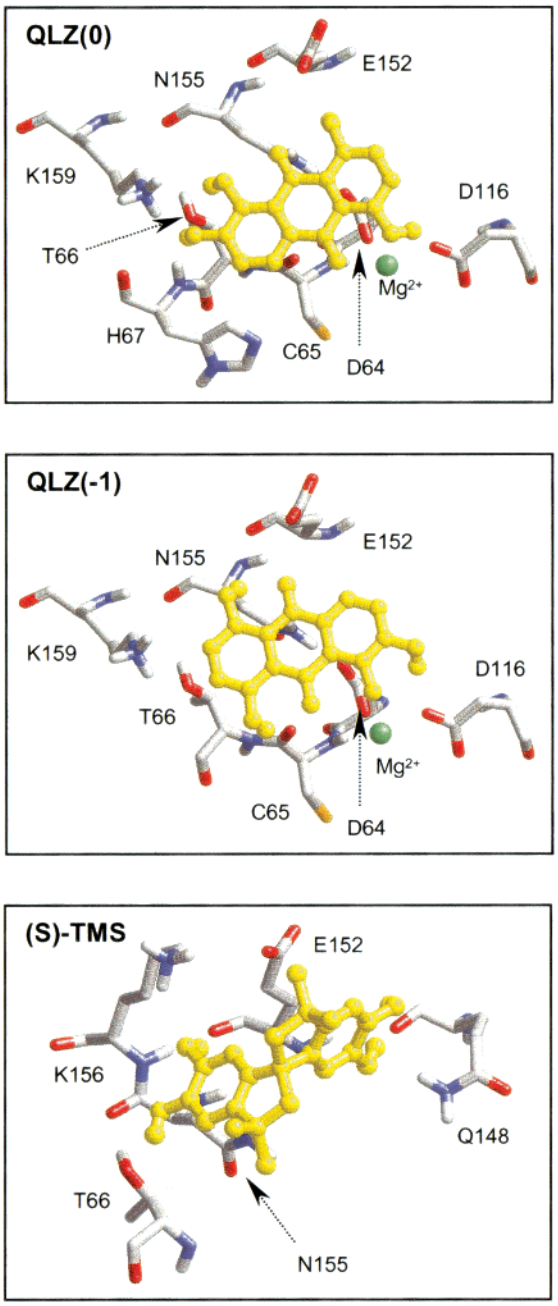

Figure 2. Graphical representation of the top-ranked binding modes obtained for neutral QLZ, deprotonated QLZ, and the (*S*)-enantiomer of TMS, showing the contacting residues given in Table 1.

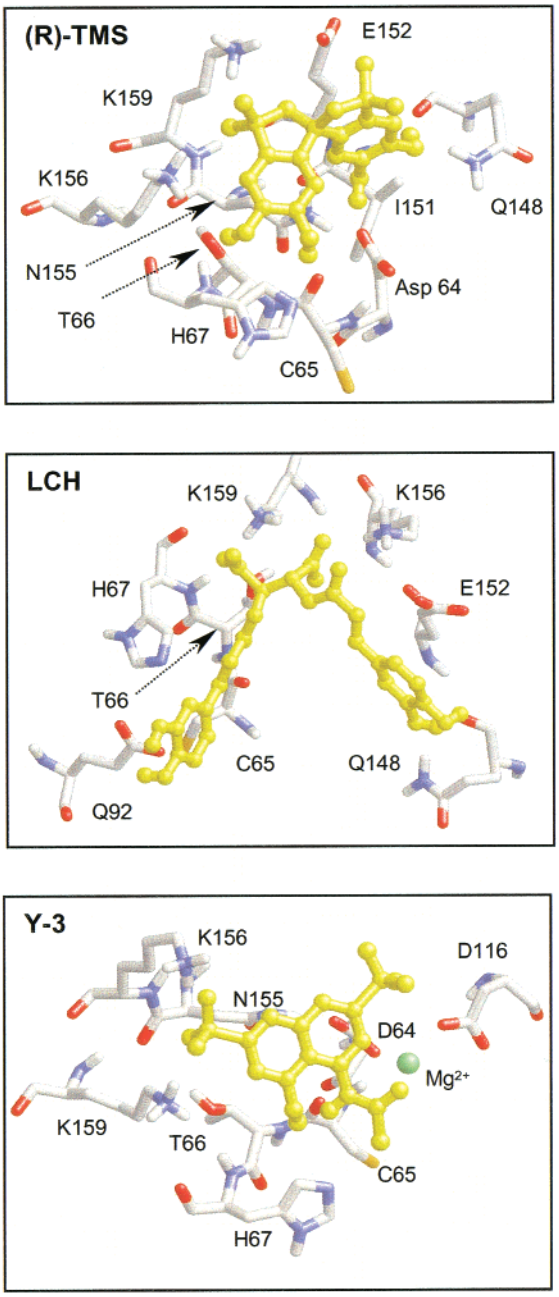

Figure 3. Graphical representation of the top-ranked binding modes obtained for the (*R*)-enantiomer of TMS, LCH, and Y-3, showing the contacting residues given in Table 1.

result with a ΔG of −7.3 kcal/mol (found 14 times out of 50), a second result is obtained with a ΔG of −7.1 kcal/mol (found 20 times). Interestingly, both results are located in the active site and share the binding mode with one or the other half of the (*S*)-spiro compound, respectively. The top result is similarly placed as the ring system of (*S*)-TMS that stretches along the Glu 152 side chain but shows additional contacts with Asp 64, Cys 65, Thr 66, and His 67. All four hydroxyls are involved in hydrogen bonds of varying quality with Asp 64, Cys 65, Thr 66, and Gln 148.

LCH. LCH is significantly larger and more flexible than the molecules discussed so far and thus a more demanding case for docking. Still, 35 of the 50 runs converge to the two top-ranked clusters which share a very similar binding mode (1.6 Å rmsd, 0.1 kcal/mol difference in free energy).

Half of the molecule occupies the same space as TMS, while the other half makes use of the area in front of the two catalytic aspartates but without contacting them. The two carboxylates interact with the lysine residues: one is placed between Lys 159 and Lys 156, the other between Lys 159 and His 67. One of the ester carbonyls is interacting with Lys 156 as well, while the other is buried (somewhat unfavorably) along the backbone of Cys 65 and Thr 66. One catechol ring shows extensive contact with Glu 152 and forms hydrogen bonds with Gln 148. The other catechol ring remains rather exposed and interacts weakly with Glu 92 (in the second ranked result, the interaction with Glu 92 is improved but the rest of the molecule is placed slightly less favorable). In terms of vdW interactions, most contacts are made with Glu 152 and Gln 148, followed by Lys 159 and His 67, but essentially the entire shallow groove surrounding the catalytic center is occupied by the ligand. This presents a plausible mode of interaction and could eventually explain the high inhibitory potency of LCH observed experimentally.

Y-3. A suitable binding position in the central active site is also found for compound Y-3 which is an inhibitor of both ASV and HIV IN. Although in the crystal structure solved for the complex of Y-3 with ASV IN the ligand was found to bind not directly to the active site but to a location close to the flexible loop,[20] docking to HIV IN suggests that inhibition may occur through a binding mode very similar to that observed for the other compounds described so far. The most favorable result (−6.9 kcal/mol, found 19 times out of 50) has its aromatic ring system placed "horizontally" in the active site, above Asp 64. One of the two sulfonates is very favorably placed between Lys 156 and Lys 159, while the other is positioned between the Mg^{2+} ion and the amide of Gln 148 (which is somewhat less favorable due to the vicinity of Asp 64 and Asp 116). The carbonyl oxygen of the ligand is coordinated with the Mg^{2+} ion. A potential hydrogen bond partner for both the amide and the hydroxy group is the backbone carbonyl of Cys 65. Further important interaction partners are Thr 66 and Asn 155.

It is worth noting that the 50 runs produced only one result outside the active site, corresponding to a position behind the flexible loop where one of the two sulfonates is interacting with Tyr 143. This result, however, is ranked last and has a free energy of binding that is more than 2 kcal/mol less favorable.

5CITEP. For comparative purpose, the results obtained for 5CITEP are briefly summarized here (for details and figures see the work mentioned previously[13]). In the most frequently occurring and most favorable result (reported in Table 1), the ligand is found in the center of the active site, the same location as in the crystal structure, but in a different orientation. The main molecular plane is aligned "horizontally" within the active site, providing a large surface area for vdW interactions with buried residues (Thr 66, Asn 155, and Gln 148). The chloroindole ring interacts with Gln 148 (the indole NH with Asp 116), the keto−enol oxygens point toward Glu 152, and the tetrazole ring is located in proximity to Lys 159.

A significant alternative to this position is given by the second cluster, which is found less frequently but

shows the same free energy. Although largely overlapping with the first result, the molecular plane is now oriented in such a way that the chlorine becomes buried in the cleft between Gln 148 and Asp 116, while the keto−enol oxygens are oriented toward the Asn 155 amide. No hydrogen bond is formed between Asp 116 and the indole NH, which remains solvent-exposed. The tetrazole ring is again interacting with Lys 159 but now found between the two lysine residues 156 and 159.

Discussion

The list of known IN inhibitors contains a considerable variety of molecular structures. The IN protein thus seems to be able to accommodate very diverse ligands.[4] One reason for this could simply be given by the fact that the active site is not a deep cleft but a comparatively large and shallow pocket. It remains, however, to be understood how ligands actually bind within this pocket and whether general principles of binding emerge that could be explored for the design of tighter binding compounds. The docking studies presented here should help to shed light on these issues, providing a structural model for IN−inhibitor interactions in the catalytic core domain. To do so, the quality of the computational method and the underlying structural data need first to be critically assessed. This equally applies to the results which are to be compared with available experimental findings.

Docking Method. The docking program AutoDock has repeatedly been shown to successfully reproduce experimental binding positions of protein ligands.[14,29−33] The problem of scoring and ranking, common to all docking programs, has been considerably alleviated in version 3.0 by the incorporation of a new empirical free energy function, which also makes comparisons among the results of different ligands more meaningful. As far as the search methods are concerned, search efficiency in version 3.0 could be significantly improved by the development of a so-called Lamarckian genetic algorithm which is the combination of a traditional genetic algorithm with a local search method.

In tests carried out by ourselves using experimentally known antibody−ligand complexes which had already served as test cases for earlier versions of AutoDock,[33] the Lamarckian genetic algorithm showed indeed the best performance and the new free energy function allowed a correct ranking and reasonably accurate guesses for the actual free energy of binding (results not shown).

Molecular Structures. Regarding the quality of the molecular structures employed for docking, most of the ligands can be modeled in a straightforward way by standard molecular mechanics procedures, with the exception of 5CITEP, for which mesomeric and tautomeric effects had to be considered (as described in detail in our previous study[13]).

Regarding the protein, the structure used in this work (subunit A of 1QS4) has the advantage of representing active site conformations of a complexed state. Therefore the approximation of a rigid protein structure becomes less severe. The normally rather flexible side chains of Glu 152 and Lys 156 show considerably reduced temperature factors in 1QS4. The disordered residues 141−144 could be incorporated from the completely solved structure of subunit B in 1BIS.[25] Due to their distance from the active site center they should not significantly affect the binding mode of ligands at the active site. Structure 1QS4 therefore appears to be a useful template for docking studies and indeed "a platform for antiviral drug design", as stated by the authors of the crystallographic structure analysis.[12]

Comparison with Experimental Data: Energies. The free energies of the docked positions estimated with the AutoDock empirical free energy function compare reasonably well with experimental inhibition data. The IC_{50} values for 5CITEP, QLZ, TMS, and Y-3 are all in the low-micromolar range, between 2 and 17 μM.[12,20−23] With the AutoDock free energy function, the calculated K_d values (at 298 K) for these compounds are in the range of 4−15 μM. The IC_{50} for LCH is the lowest in the series and has been determined to 0.1−0.2 μM.[22,34,35] With −7.5 kcal/mol and a corresponding K_d of 3 μM, the binding strength of LCH may be somewhat underestimated by AutoDock, but nevertheless it is found to be the tightest binding inhibitor in the series.

A comparison beyond these general considerations is not feasible, as IC_{50} and K_d values are not directly comparable and the experimental data were determined by different methods. Furthermore, the energy function is reported to have a residual standard error of about 2 kcal/mol, corresponding to a factor of 30 in relative affinity (K_{d1}/K_{d2}), which of course precludes a more detailed comparison. Nevertheless, this is an excellent performance for an empirical free energy function and superior over earlier scoring schemes, as it allows reasonably accurate estimates of the actual affinity range.

Comparison with Experimental Data: Binding Modes. Structural information about inhibitor binding to IN is still very limited. The only crystal structure of an inhibitor bound to the active site has been obtained for 5CITEP.[12] Accordingly, a comparison of the previously obtained 5CITEP docking results[13] with this structure is of prime interest and briefly summarized here.

The two positions found by docking show an rmsd of 2.8 and 3.2 Å from the X-ray structure. While the overall location at the active site and the relative positioning of the three functional units (tetrazole, keto−enol, chloroindole) are very similar, the orientation within the active site is different. In both docking results, the tetrazole ring is placed closer to the lysine residues 156 and 159, the keto−enol oxygens are moved farther away from the Glu 152 carboxylate, and the chloroindole ring interacts more tightly with the protein. Both docked positions show a significantly more favorable binding energy compared to the experimental position.

As argued in detail in our earlier study,[13] the reason for this discrepancy may ultimately be related to a crystallographic packing effect. In the crystal, the active site of one of the three IN molecules (A) in the asymmetric unit is close to the active site of another IN molecule (A′) in a neighboring cell, with A and A′ related by a crystallographic 2-fold axis. Only for A (and A′, respectively) the position of the ligand could be unambiguously determined, but not for the other two IN molecules in the asymmetric unit. This arrangement of the A and A′ active sites forms a larger binding cavity

which allows the ligand to interact not only with its primary "receptor" but also with the symmetry related protein and the ligand bound to it. These additional contacts seem to contribute significantly to the actual orientation of the ligand. In the absence of the crystallographic neighbor, however, these contacts are lost and the docked positions become energetically preferred. MD simulations are currently underway to further clarify the interaction between IN and 5CITEP.

Regarding the other inhibitors, less detailed information is available about their interaction with IN. For QLZ, binding to the catalytic domain has been detected by competition with nucleotide binding, and it seems likely that it occurs at the active site.[22] Since QLZ can potentially chelate metal ions and was found to be cross-reactive with other metal-containing enzymes, it has been suggested that it may form a ternary complex with the metal and the enzyme.[22] This would be in perfect agreement with the docking result.

Nothing is known yet about the binding site of TMS and the activity of the single enantiomers (reported activity data were obtained for the racemic mixture[23]). Using IN structure 1BL3[36] as template for docking prior to the availability of 1QS4, a slight preference for the (*S*)-enantiomer could be detected (results not shown). This is, however, not confirmed by docking to 1QS4, as for both enantiomers results with almost the same energy (−7.2 and −7.3 kcal/mol) are found. The reason is most probably given by the significantly different orientation of the Glu 152 and Gln 148 side chains in the two protein structures. Experimental binding data for the enantiomerically pure compounds could therefore also help to reveal the active binding conformation of these residues.

Biochemical studies have recently made plausible that LCH acts on IN, not on DNA, and binds to the catalytic core domain.[37] In the same study, it has also been shown that the divalent metal ion is not significantly involved in the inhibition mechanism, which would be in agreement with the binding mode proposed by docking. Furthermore, Cys 65 has been attributed a role in binding LCH: mutation of this residue increased the IC_{50} and a mechanism has been proposed, which would explain the observed irreversible inhibition of LCH by a reaction of the oxidized catechol moiety with the sulfhydryl group.[37] Interestingly, the catechol ring in the docking result is found to be positioned just above the side chain of Cys 65. Another study has also suggested that binding of LCH may occur at the core domain, most likely at residues near the catalytic triad, as a glycine-to-serine mutant at residue 140 displayed reduced sensitivity to LCH.[38] Although in the docked position LCH does not directly interact with Gly 140, this does not necessarily disprove the docking result. Gly 140 is actually critical for loop flexibility, and it has been shown that mutations of this residue lead to impaired catalytic activity while DNA binding affinity is minimally affected.[39]

The binding mode of Y-3 to HIV IN is equally unknown, but since the compound also inhibits ASV IN (with IC_{50} values of 40−350 μM), an X-ray structure of the complex with the ASV IN catalytic domain could be obtained.[20] In the ASV IN complex, Y-3 is not found directly at the active site but in a nearby position separated by the flexible loop. In contrast, docking suggests that binding to HIV IN occurs at the active site, even though the search grid included also the flexible loop region (only a poorly occupied and much less favorable result could be found in that area). A structural analysis points again to crystal packing effects being involved in the interaction between ASV IN and Y-3: in the crystal, the Y-3 molecule stacks against its symmetry-related mate and makes significant contacts with the symmetry-related second IN monomer; visualization of the monomeric complex shows only a comparatively loose fit between the ligand and the protein. Regarding binding of Y-3 to HIV IN, a similar binding mode as in ASV IN appears questionable based on comparisons between the ASV IN complex and the available HIV IN structures. It can, however, not be completely ruled out because the loop region essential for this putative Y-3 binding site is known to be highly flexible. On the other hand, binding to the active site of the 1QS4 HIV IN structure seems readily possible and would also allow a straightforward explanation of inhibition, which is not the case for binding as observed in ASV IN.

Self-Consistency of Docking Results. The binding modes of the ligands show common characteristics, as revealed by superposition of the docking results (cf. Figure 4). The results are "self-consistent" in the sense that functional groups of similar chemical character are placed in similar ways and show comparable interactions with the protein.

QLZ (neutral form) and the tetrazole and keto−enol units of 5CITEP occupy exactly the same plane in the binding site. The amine of the chloroindole and one hydroxyl of QLZ point to the same hydrogen bond acceptor, Asp 116; the negatively charged tetrazole ring overlaps with the 1,2-dihydroxy group, both interacting with Lys 159; the β-hydroxy ketone units of both inhibitors are involved in polar interactions at similar locations. The aromatic system of Y-3 occupies the same plane as 5CITEP and QLZ, and the orientation of the Y-3 functional groups is comparable to the orientation observed for the charged form of QLZ.

TMS enters the active site as deeply as the planar aromatic systems, but obviously in a different way by virtue of its bulky nature. Nevertheless both TMS enantiomers are found to place one of their hydroxy groups (the OH closest to Lys 159) similar to the corresponding hydroxyls of QLZ. Interesting homologies are observed for the TMS (*S*)-enantiomer and LCH: the catechol moiety interacting with Gln 148 is identically placed in both cases, and one carboxylate of LCH is situated in the same area as the second pair of hydroxy groups of TMS. If compared to Y-3, this carboxylate of LCH is exactly at the same position as the corresponding Y-3 sulfonate placed between Lys 156 and Lys 159.

The interaction with the two lysine residues appears to be one of the most essential recurring motifs. Sulfonate and carboxylate groups are exceptionally well-placed here, and hydroxyls as well as the tetrazolium are very favorable interaction partners, too. Based on these findings, it may obviously be suspected that this region of the active site could also serve to keep part of the phosphate backbone of the DNA substrate in place. When dAMP was docked to IN according to the same

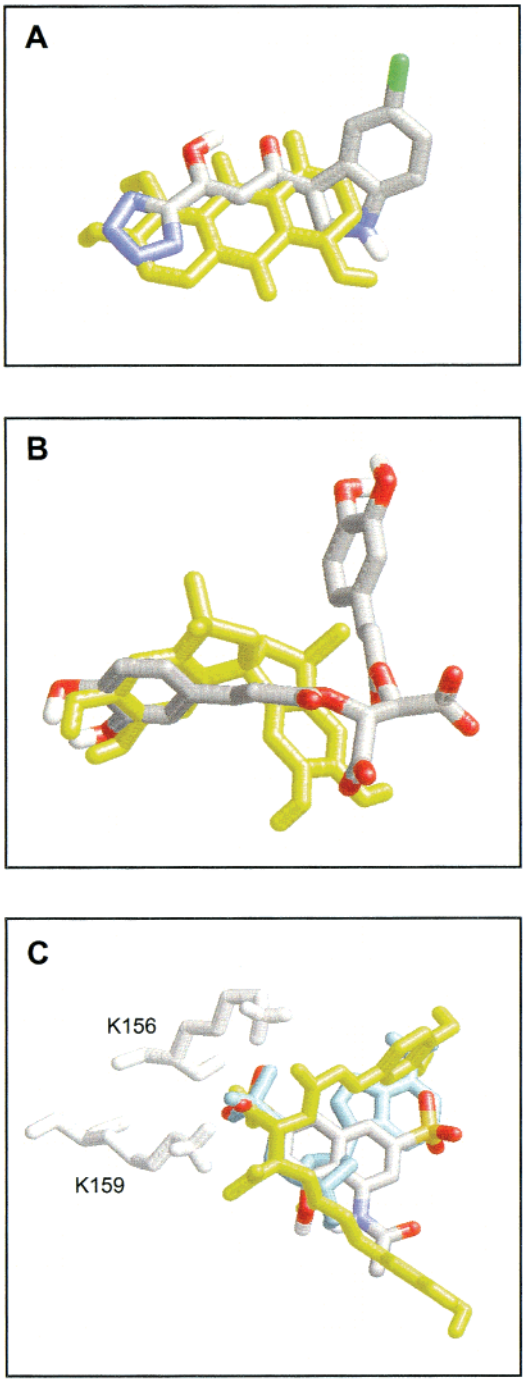

Figure 4. Comparison of ligand-binding modes revealed by docking: (A) superposition of 5CITEP (CPK colors) and QLZ-(0) (yellow); (B) superposition of LCH (CPK colors) and (*S*)-TMS (yellow); (C) superposition of Y-3 (CPK colors), LCH (yellow), and dAMP (light blue); also the two essential lysine residues are shown.

protocol as for the inhibitors, the phosphate was indeed with high preference found at this position (cf. Figure 4). The corresponding dAMP result would also make sense with respect to the fact that some of the inhibitors, most notably 5CITEP, can easily be recognized as nucleotide mimics: the tetrazolium replaces the phosphate, the central polar sugar unit is mimicked by the keto−enol group, and the chloroindole is a surrogate for a purine base. As a matter of fact there is also considerable overlap between the adenine and the chloroindole in the docking result. Support for these preliminary findings for dAMP could also be seen in photo-cross-linking studies, which have attributed a prominent role in substrate/nucleotide binding to Lys 156, Lys 159, and Gln 148,[40−42] although different binding modes were ultimately suggested. We are currently investigating the possibilities for nucleotide binding in a more systematic and comprehensive way.

Guidelines for Structure-Based Ligand Design. Although the current knowledge about structure and function of the IN enzyme in vivo is still rather limited (neither an intact three-domain protein structure nor an IN−DNA complex structure is available at the moment), the experimental results mentioned above indicate that targeting the active site in the catalytic core domain should be a useful strategy for inhibitor design. Based on the first crystal structure of an inhibitor bound to the active site, the docking results presented here support this hypothesis and may indicate directions for further structure based design efforts.

According to the docking results, the inhibitor binding site is formed by Asp 64, Cys 65, Thr 66, His 67, Glu 92, Asp 116, Gln 148, Ile 151, Glu 152, Asn 155, Lys 156, and Lys 159. The site is clearly dominated by polar residues, and the challenge for the design of tighter binding ligands is to make maximum use of the available interaction partners and minimize conflicts, avoiding especially the frustration of potential hydrogen-bonding sites of the ligand.

An interaction site of primary importance, which may serve as anchor for tight binding active site ligands, is given by the two lysines and the neighboring residues. Carboxylates, sulfonates, and phosphates are exceptionally well-placed here, but also a tetrazolium or catechol group can be favorably accommodated. In most of the cases analyzed by docking, about 30% of the total intermolecular energy arises from interactions of the functional groups placed at this position. Ideally, the interaction at this site is complemented by favorable interactions with Thr 66 and His 67 (cf. the binding mode of LCH). Glu 92 bears potential for further exploitation by larger ligands. Cys 65 is occasionally involved in favorable interactions; it needs to be considered when exploring the region around Thr 66, His 67, and Glu 92 to avoid repulsive contacts with the backbone carbonyl. With the Mg^{2+} ion bound between them, the two catalytic residues Asp 64 and Asp 116 are not easily exploited in a straightforward way. Only the carboxylate of Asp 116 appears to be predisposed as single hydrogen bond acceptor for ligand interactions (cf. QLZ). Targeting the metal ion may not be a useful strategy because the corresponding compounds can cross-react with other metal-containing enzymes (as seen, for example, in the work of Farnet et al.[22]). The carboxylate of the third catalytic residue, Glu 152, is, in principle, accessible for ligand binding but too exposed over the active site. None of the investigated ligands shows specific interactions with this carboxylate in the docked binding modes. Special attention is of course required regarding the role of the Glu 152 side chain, since it is known to be rather flexible (MD simulations of IN−inhibitor complexes may help to clarify its role in ligand binding; such simulations are currently underway for the complex with 5CITEP). Finally, Gln 148 could have the potential to be more

specifically explored as second anchor for ligands at the active site. In the conformation observed in the IN structure used for docking, both the backbone carbonyl and the side chain amide (NH_2) point toward the active site and are directly available for polar interactions and hydrogen bonds (cf. (*S*)-TMS and LCH).

Conclusion

Using the first IN crystal structure of an active site inhibitor complex as template, docking with AutoDock provided structural models for inhibitor binding that are largely consistent with available experimental data. For 5CITEP, optimized binding modes were proposed which may be preferred in the absence of the crystallographic environment and thus a better starting point for structure-based design. For the other inhibitors, the docking results represent the first detailed molecular binding model. Altogether a coherent picture of possible interactions at the active site was obtained that should be helpful in future efforts of inhibitor design. An ideal ligand may require a very-well-placed set of polar groups. Tight interactions should most readily be accessible through binding to Lys 156 and Lys 159. Starting from this anchor site, ligands may be constructed that make optimized use of the other available interaction sites. LCH appears to be the most useful lead structure in the investigated set of compounds as it occupies most of the available interaction space and has a well-distributed (though not yet ideal) set of polar groups to interact with the active site residues.

Acknowledgment. C.A.S. is grateful to the Austrian Science Fund for a postdoctoral fellowship (J1758-GEN). H.H.N. is an NIH postdoctoral fellow (F32-GM63094). We thank MSI for the donation of the QUANTA and CHARMm software and Dr. Arthur Olson for the AutoDock program. We also thank Dr. Heather Carlson for helpful comments. This work is supported by NIH Grant GM56553 and through grants of supercomputer time from the NSF Supercomputer Centers.

References

(1) Katz, R. A.; Skalka, A. M. The retroviral enzymes. *Annu. Rev. Biochem*. **1994**, *63*, 133−173.
(2) Farnet, C. M.; Bushman, F. D. HIV cDNA integration: molecular biology and inhibitor development. *AIDS* **1996**, *10* (Suppl. A), S3−S11.
(3) Asante-Appiah, E.; Skalka, A. M. Molecular mechanisms in retrovirus DNA integration. *Antiviral Res.* **1997**, *36*, 139−156.
(4) Pommier, Y.; Neamati, N. Inhibitors of human immunodeficiency virus integrase. *Adv. Virus Res.* **1999**, *52*, 427−458.
(5) Nicklaus, M. C.; Neamati, N.; Hong, H.; Mazumder, A.; Sunder, S.; Chen, J.; Milne, G. W. A.; Pommier, Y. HIV-1 integrase pharmacophore: discovery of inhibitors through three-dimensional database searching. *J. Med. Chem*. **1997**, *40*, 920−929.
(6) Hong, H.; Neamati, N.; Wang, S.; Nicklaus, M. C.; Mazumder, A.; Zhao, H.; Burke, T. R.; Pommier, Y.; Milne, G. W. A. Discovery of HIV-1 integrase inhibitors by pharmacophore searching. *J. Med. Chem*. **1997**, *40*, 930−936.
(7) Neamati, N.; Hong, H.; Mazumder, A.; Wang, S.; Sunder, S.; Nicklaus, M. C.; Milne, G. W. A.; Proksa, B.; Pommier, Y. Depsides and depsidones as inhibitors of HIV-1 integrase: discovery of novel inhibitors through 3D database searching. *J. Med. Chem*. **1997**, *40*, 942−951.
(8) Hong, H.; Neamati, N.; Winslow, H. E.; Christensen, J. L.; Orr, A.; Pommier, Y.; Milne, G. W. A. Identification of HIV-1 integrase inhibitors based on a four-point pharmacophore. *Antiviral Res.* **1998**, *9*, 461−472.
(9) Carlson, H. A.; Masukawa, K. M.; McCammon, J. A. Methods for including the dynamic fluctuations of a protein in computer-aided drug design. *J. Phys. Chem. A* **1999**, *103*, 10213−10219.
(10) 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.
(11) Masukawa, K. M.; Carlson, H. A.; McCammon, J. A. Technique for developing a pharmacophore model that accommodates inherent protein flexibility: an application to HIV-1 integrase. In *Pharmacophore perception, development, and use in drug design*; Güner, O. F., Ed.; International University Line: La Jolla, CA, 2000; pp 409−427.
(12) 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.
(13) 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.
(14) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem*. **1998**, *19*, 1639−1662.
(15) Lybrand, T. P. Ligand−protein docking and rational drug design. *Curr. Opin. Struct. Biol.* **1995**, *5*, 224−228.
(16) Bamborough, P.; Cohen, F. E. Modeling protein−ligand complexes. *Curr. Opin. Struct. Biol.* **1996**, *6*, 236−241.
(17) Lengauer, T.; Rarey, M. Computational methods for biomolecular docking. *Curr. Opin. Struct. Biol.* **1996**, *6*, 402−406.
(18) Clark, D. E.; Murray, C. W.; Lin, J. Current issues in de novo molecular design. In *Reviews in Computational Chemistry;* Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley−VCH: New York, 1997; Vol. 11, pp 67−125.
(19) Murcko, M. A. Recent advances in ligand design methods. In *Reviews in Computational Chemistry;* Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley−VCH: New York, 1997; Vol. 11, pp 1−66.
(20) Lubkowski, J.; Yang, F.; Alexandratos, J.; Wlodawer, A.; Zhao, H.; Burke, Jr., T. R.; 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.
(21) Farnet, C. M.; Wang, B.; Lipford, J. R.; Bushman, F. Differential inhibition of HIV-1 preintegration complexes and purified integrase protein by small molecules. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 9742−9747.
(22) Farnet, C. M.; Wang, B.; Hansen, M.; Lipford, J. R.; Zalkow, L.; Robinson, Jr., W. E.; Siegel, J.; Bushman, F. Human immunodeficiency virus type 1 cDNA integration: new aromatic hydroxylated inhibitors and studies of the inhibition mechanism. *Antimicrob. Agents Chemother*. **1998**, *42*, 2245−2253.
(23) Molteni, V.; Rhodes, D.; Rubins, K.; Hansen, M.; Bushman, F.; Siegel, J. S. A new class of HIV-1 integrase inhibitors. The 3,3,3′,3′-tetramethyl-1,1′-spirobis(indan)-5,5′,6,6′-tetrol family. *J. Med. Chem*. **2000**, *43*, 2031−2039.
(24) Gasteiger, J.; Marsili, M. Iterative partial equilization of orbital electronegativity − a rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219−3228.
(25) 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.
(26) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham III, T. E.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Ferguson, D. M.; Radmer, R. J.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER 5*; University of California: San Francisco, 1997.
(27) 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.
(28) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A new force field for molecular mechanical simulations of nucleic acids and proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765−784.
(29) Goodsell, D. S.; Olson, A. J. Automated docking of substrates to proteins by simulated annealing. *Proteins* **1990**, *8*, 195−202.
(30) Morris, G. M.; Goodsell, D. S.; Huey, R.; Olson, A. J. Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 293−304.
(31) Goodsell, D. S.; Morris, G. M.; Olson, A. J. Automated docking of flexible ligands: applications of AutoDock. *J. Mol. Recogn.* **1996**, *9*, 1−5.
(32) Sotriffer, C. A.; Winger, R. H.; Liedl, K. R.; Rode, B. M.; Varga, J. M. Comparative docking studies on ligand binding to the multispecific antibodies IgE−La2 and IgE−Lb4. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 305−320.

(33) Sotriffer, C. A.; Flader, W.; Winger, R. H.; Rode, B. M.; Liedl, K. R.; Varga, J. M. Automated docking of ligands to antibodies: methods and applications. *Methods: Companion Methods Enzymol.* **2000**, *20*, 280−291.
(34) McDougall, B.; King, P. J.; Wu, B. W.; Hostomsky, Z.; Reinecke, M. G.; Robinson, Jr., W. E. Dicaffeoylquinic and dicaffeoyltartaric acid are selective inhibitors of human immunodeficiency virus type 1 integrase. *Antimicrob. Agents Chemother.* **1998**, *42*, 140−146.
(35) Lin, Z.; Neamati, N.; Zhao, H.; Kiryu, Y.; Turpin, J. A.; Aberham, C.; Strebel, K.; Kohn, K.; Witvrouw, M.; Pannecouque, C.; Debyser, Z.; De Clerq, E.; Rice, W. G.; Pommier, Y.; Burke, Jr., T. R. Chicoric acid analogues as HIV-1 integrase inhibitors. *J. Med. Chem.* **1999**, *42*, 1401−1414.
(36) Maignan, S.; Guilloteau, J.-P.; Zhou−Liu, Q.; Clément−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.
(37) Zhu, K.; Cordeiro, M. L.; Atienza, J.; Robinson, Jr., W. E.; Chow, S. A. Irreversible inhibition of human immunodeficiency virus type 1 integrase by dicaffeoylquinic acids. *J. Virol.* **1999**, *73*, 3309−3316.
(38) King, P. J.; Robinson, Jr., W. E. Resistance to the anti−human immunodeficiency virus type 1 compound L-chicoric acid results from a single mutation at amino acid 140 of integrase. *J. Virol.* **1998**, *72*, 8420−8424.
(39) 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.
(40) Jenkins, T. M.; Esposito, D.; Engelman, A.; Craigie, R. Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photocrosslinking. *EMBO J.* **1997**, *16*, 6849−6859.
(41) Esposito, D.; Craigie, R. Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction. *EMBO J.* **1998**, *17*, 5832−5843.
(42) Drake, R. R.; Neamati, N.; Hong, H.; Pilon, A. A.; Sunthankar, P.; Hume, S. D.; Milne, G. W. A.; Pommier, Y. Identification of a nucleotide binding site in HIV-1 integrase. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 4170−4175.

JM000194T