Letters

Discovery of a Novel Binding Trench in HIV Integrase

Julie R. Schames,* Richard H. Henchman, Jay S. Siegel, Christoph A. Sotriffer, Haihong Ni, and J. Andrew McCammon

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

Received September 11, 2003

Abstract: Docking of the 5CITEP inhibitor to snapshots of a 2 ns HIV-1 integrase MD trajectory indicated a previously uncharacterized trench adjacent to the active site that intermittently opens. Further docking studies of novel ligands with the potential to bind to both regions showed greater selective affinity when able to bind to the trench. Our ranking of ligands is open to experimental testing, and our approach suggests a new target for HIV-1 therapeutics.

Introduction. HIV-1 integrase (IN), the enzyme responsible for the integration of viral DNA into the host genome, represents an attractive yet unexploited target for the treatment of HIV infection. The lack of detailed structural information about IN−ligand interactions has hampered the design of IN inhibitors. There is currently one crystal structure of IN complexed with an inhibitor (5CITEP).[1] However, the information provided by the crystal structure is questionable. There is a disordered flexible loop (Ile141−Asn144) whose structure cannot be precisely determined and is believed to be near the substrate during integration.[2] Integrase is thought to form a large complex with DNA, and this relevant IN-DNA complex structure is unknown.[3] Docking studies have shown that the position of the ligand appears to be influenced by crystal packing effects.[4]

The actual binding site of IN is not understood; only key residues are known. The Relaxed-Complex method was used to elucidate binding modes of the 5CITEP inhibitor to the IN protein, especially in regard to the loop region (residues 141−144). The Relaxed Complex method is a scheme to take protein flexibility into account by docking flexible ligands to protein snapshots taken from a molecular dynamics (MD) simulation.[5]

Docking of 5CITEP. A 2 ns MD simulation using the protein structure of the IN from subunit A of PDB file 1QS4, with all crystallographic waters removed, was run.[6] Covalent bonds to hydrogen atoms were constrained by the SHAKE algorithm, a 2 fs time step was used, a 10 Å cutoff was applied for van der Waals interactions, and the electrostatics were treated by the Particle Mesh Ewald method. The simulation was carried out with the SANDER MD module of AMBER 5.0.[7]

Snapshots from the simulation were taken every 50 ps. The 5CITEP was docked to those MD snapshots using AutoDock 3.0.[8] It allows for torsional flexibility in the ligand while holding the protein structure rigid and also incorporates an empirical free energy function and a Lamarckian genetic search algorithm. Two flexible torsions were specified: one about the tetrazole ring, and one about the indole ring.

Docking 5CITEP to the MD snapshots revealed two predominant docked conformations (Figure 1).

The 5CITEP shown in green is the general orientation found in the crystal structure. This was noted in many cases, with slight variations rotated about a horizontal axis through the ligands. The second conformation was much more intriguing. The 5CITEP shown in yellow we termed the "flipped" conformation. This conformation was noted in a majority of snapshots. The indole ring of 5CITEP is in the same general position, and the whole molecule is flipped so that the keto−enol groups and the tetrazole ring point in the opposite direction.

Evidently, a new space had opened up adjacent to the active site that binds 5CITEP more strongly than the X-ray position. We termed this new space the "trench".

Analysis of the residues (as shown in Figure 1) that make up the trench, His114, Gly140, Phe139, Pro142, and Gln148, revealed that there were two main contributors to the opening and availability of the trench. First, the Gln148 moves back and widens the active site region (when closed, Gln148 and Asp116 appear "pinched", keeping the active site narrow). Second, Pro142 pops out, and allows for significant movement and flexibility of Gly140 and Phe139. The movement of Phe139 reveals an additional opening between it and Asp116, although this gap was not docked into. Whether the trench is open or closed is a subtle function of residue position. Residue RMSDs were found not to be a good indicator of whether the trench is open or closed.

Docking Butterfly Compounds. To take advantage of the two positions into which 5CITEP docks, a series of ligands that we termed "butterfly compounds" (Figure 2) were designed. Having discovered a new potential binding site, we designed ligands to bind to both the crystallographic binding site and the trench. Their structure is based on the 5CITEP. It places two keto−enol and tetrazole groups and in all possible positions around a naphthalene spacer. By placing these two "wings" in all possible ways, we ensured our sample of compounds was unbiased for the trench.

From the 40 MD snapshots 5CITEP was originally docked to, 10 protein conformations were chosen to which we docked our butterfly compounds: 7 MD snapshots in which we had seen flipping and therefore had an open trench, 2 MD snapshots had no flipping and a closed trench, and the original X-ray structure.

We expected that compounds whose wings most resembled both the flipped and crystallographic conformations (Figure 1) would dock to the open snapshots the best, but they would not be favored over other butterfly compounds when docked to closed snapshots.

10.1021/jm0341913 CCC: $27.50
Published on Web 03/09/2004

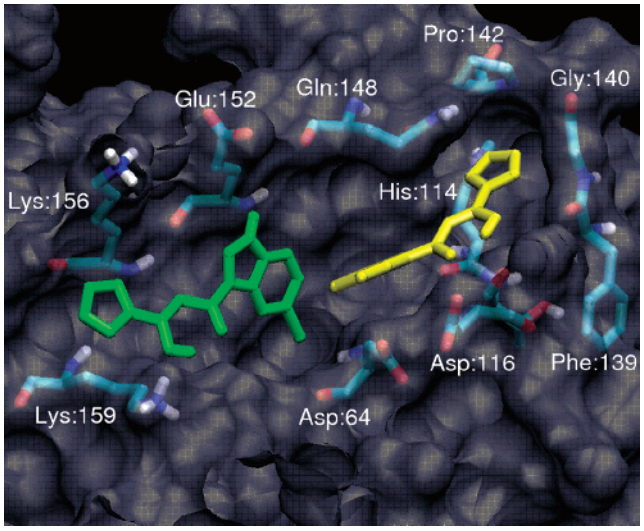

Figure 1. The two predominant docking conformations of 5CITEP to an open MD snapshot of integrase. The ligand in green shows 5CITEP in the orientation similar to the crystal structure of the complex. The ligand in yellow shows 5CITEP in its "flipped" orientation. Residues lining both ligand positions are highlighted.

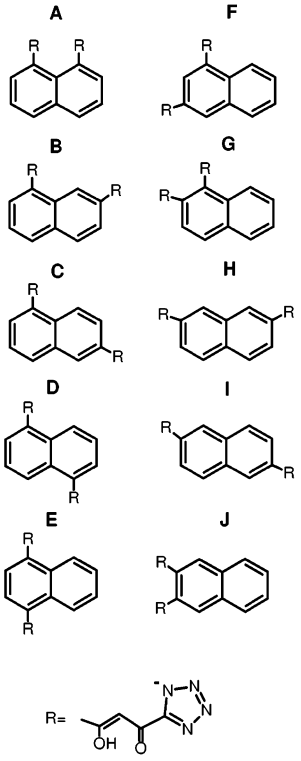

Figure 2. The 10 butterfly compounds. The R group is modeled after the 5CITEP inhibitor. The compounds comprise all possible arrangements of the two R groups.

The butterfly compounds were docked to the protein conformations using AutoDock. The identical docking protocol was used as with the original 5CITEP docking. The same two dihedrals in the tetrazole/keto−enol were allowed to rotate, giving four flexible dihedrals per compound. The results of docking the butterfly compounds to the different protein snapshots are displayed in the histograms in Figure 3. Each histogram is constructed from all the docked energies of a single butterfly compound. The bars in green represent docking to open snapshots, the bars in red represent docking to closed snapshots, and the bars in blue represent docking to the X-ray structure.

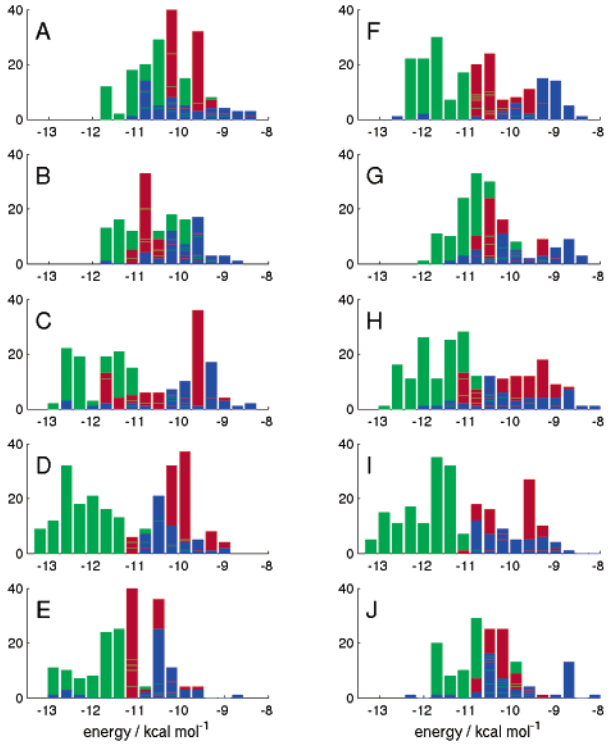

Figure 3. The energy docking histograms for the butterfly compounds. Data from the open snapshots are shown in green, from the closed snapshots in red, and from the X-ray structure in blue. The single horizontal bars indicate overlapping data.

Those butterfly compounds that could take advantage of both the active site and the trench docked to the open MD snapshots at lower energies than those butterfly compounds that could not. The compounds showed no significant energetic difference when docking to MD snapshots of the closed trench, or to the X-ray structure.

All 10 compounds docked with better energies to open snapshots than to closed or X-ray structures. The greatest difference in energies was seen with compounds that could take full advantage of the trench (**D** and **I**, with a ∼2 kcal/mol preference for the open snapshots). Notably, the structures of these ligands are most similar to the two conformations of the 5CITEP that we saw earlier when combined. Figure 4 illustrates a typical docking conformation for these two compounds to an open protein conformation.

The energies for docking to closed snapshots and to the X-ray structure are approximately the same for all 10 compounds. This reinforces the idea that the X-ray structure can be thought of as a closed conformation.

Discussion. The structure of HIV-1 IN in the vicinity of the active site region is not confidently known. By combining MD with flexible-ligand docking, we have shown the existence of a new and possibly important binding region, the trench. This open protein conformation was noted in a majority of the snapshots, suggesting that it is energetically stable. The trench is lined with residues from the loop region that had been built in previously (Ile141−Asn144). This reinforces the usefulness of the approach whereby MD simulations be run on proteins that have ambiguous loops built in and reconstructed.

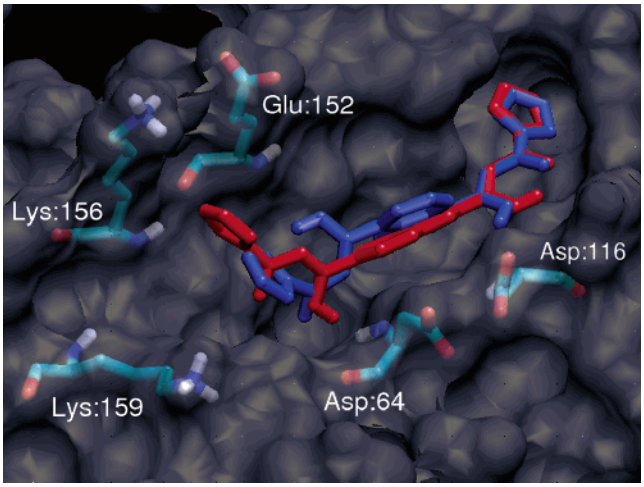

Figure 4. Compounds **D** (blue) and **I** (red) superimposed in the same open MD snapshot. Each ligand samples the active site and the trench for maximal binding energy.

These results bring up some important issues. We have discovered a potentially important part of the IN enzyme which should be considered for drug targeting. Earlier work suggests that residues 141−148 constitute an important region for the enzymatic mechanism, and that its behavior could point to the need for flexibility for efficient catalytic activity.[9] Additionally, the region between residues 139−152 had been identified as the one interacting with DNA.[10]

Some of the butterfly compounds were able to take advantage of the open trench and others were not, providing a testable prediction that we feel is reliable and reproducible, within the limitation of the theory applied. This is especially true because the butterfly compounds showed no significant energetic difference when docking to MD snapshots that were closed.

The work shown here used ligand shape as the optimizing factor. We did not look at variations in functional groups, charge, or spacer length. These are obvious next steps for pharmacophore development of HIV-1 IN. The Relaxed-Complex method has proven an effective tool for the general ranking of compounds within families. Given a new family of inhibitors, we could theoretically rank binding as well.

Acknowledgment. H.H.N. was an NIH postdoctoral fellow (F32-GM63094). C.A.S. is grateful to the Austrian Science Fund for a postdoctoral fellowship (J1758-GEN). We thank Dr. Arthur Olson for the AutoDock program. This work was supported, in part, by grants to JAM from NIH, NSF, and NPACI/SDSC.

References

(1) 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.

(2) Sotriffer, C. A.; Ni, 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.

(3) 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 withy the 5CITEP Inhibitor: Mechanism for Inhibition and Drug Resistance. *Biophys. J.* **2003**, *84*, 1450−1463.

(4) Varmus, H.; Coffin, J. M.; Hughes, S. H. *Retroviruses*; Cold Spring Harbor Laboratory Press: Plainview, NY, 1997.

(5) Lin, J.-H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. Computational Drug Design Accommodating Receptor Flexibility: The Relaxed Complex Scheme. *J. Am. Chem. Soc.* **2002**, *124*, 5632−5633.

(6) 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(19)*, 3043−3047.

(7) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; 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.

(8) 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 free energy function. *J. Comput. Chem.* **1998**, *19*, 1639−1662.

(9) 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.

(10) 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.

JM0341913