Fully Flexible Low-Mode Docking: Application to Induced Fit in HIV Integrase

György M. Keserû*,† and István Kolossváry*,‡

Computer Assisted Drug Discovery
Gedeon Richter Ltd., P.O. Box 27
H-1475 Budapest, Hungary
Novartis Institute for Biomedical Research
556 Morris Avenue, Summit, New Jersey 07901

Received April 12, 2001

HIV integrase is responsible for the integration of viral DNA into host DNA and is an important target in AIDS research.[1,2] A whole arsenal of computer-assisted drug discovery methods has been applied to find potent inhibitors. Recently, structure−activity relationships were established by Zouhiri et al.,[3] a dynamic pharmacophore model was developed by McCammon and coworkers,[4] and a similarity search has been published by Chen et al.[5] In particular, structure-based studies involved MD simulation on the catalytic core domain,[6] and docking calculation using Autodock.[7] Although there are a number of compounds inhibiting HIV integrase in vitro, only 4-aryl-2,4-dioxobutanoic acid derivatives were found to be active in vivo.[8] L-731,988 by Merck is a specific inhibitor of integration and inhibits HIV-1 infectivity in low micromolar range. Considering that the tetrazole moiety is a well-known bioisostere of the carboxylic acid functionality, the structure of the Merck inhibitor closely resembles that of 5-CITEP, an integrase inhibitor of Shionogi.[9] Since the structure of the 5-CITEP/HIV integrase complex has been recently reported (PDB code: 1QS4) we were interested in whether the Merck inhibitor binds in the same mode. X-ray analysis of the complex revealed that a partially resolved flexible loop (Ile141-Gln148) is in contact with the indole ring of 5-CITEP. The conformation of this loop was found to be crucial to the catalytic activity of HIV integrase.[10] The conformation of this loop, however, was only determined in the substrate free form (1BL3 C chain, 1BIS B chain). McCammon inserted the missing Ile141-Asn144 fragment in 1QS4 from the uncomplexed 1BIS structure. Considering the role of this loop we argue that mixing of coordinates from a complexed and a substrate-free structure might be misleading. Moreover, the rigid protein approach applied in Autodock avoids consideration of the conformational mobility of this loop. Finally, docking into a fixed binding site particular to 5-CITEP might lead to artifacts for other ligands interacting with the flexible loop. In contrast to McCammon's approach our docking protocol models induced fit by allowing fully flexible docking of 5-CITEP to the full-length catalytic core domain of HIV integrase (1BL3 C chain).[11] Our own low-mode docking methodology, LMOD,[12] utilizes the concerted atomic motions found in low-frequency vibrational modes. We applied LMOD to the attempted reproduction of the experimental complex (1QS4) and the prediction of the binding mode of the structurally similar inhibitor L-731,988.

We argue in this communication that induced-fit modeling involving significant conformational changes of a protein host, especially when experimental evidence suggests that such conformational changes are correlated with activity, can only be accomplished by fully flexible docking. We utilized LMOD docking combined with recent advances introduced to LMOD in the L-LMOD algorithm.[13] L-LMOD is specifically designed for large-scale calculations involving entire, fully flexible protein molecules. The computations were carried out using a pre-release version of BatchMin 7.5 obtained from Schrödinger, Inc. for beta testing on an Intel-based 450 MHz Pentium II, dual processor Linux platform. The calculations started with manual docking of the inhibitor (5-CITEP and Merck's L-731,988, respectively) into the binding site of the X-ray structure of the *empty* HIV integrase (1BL3). Note that the starting docking mode was intentionally kept dissimilar to that found in the X-ray structure of the cocrystallized 5-CITEP complex (1QS4), both in terms of location and orientation of the inhibitor in the active site. The initial inhibitor−protein complex was minimized with the AMBER* force-field[14,15] with attenuated Coulombic electrostatics ($\epsilon = 4r$) as a simple charge screening function to model solvation effects.[16] In our experience, application of distance-dependent electrostatics with all-atom AMBER* has been proven to be adequate to reproduce experimentally found binding geometries with high accuracy in a variety of protein−ligand complexes.[17] Of course, calculation of binding free energy would require a more sophisticated solvation treatment, but our scope in the present work was aimed at geometry. Both the inhibitor and the protein host were allowed to move freely during minimization except for the Mg ion and its four coordinated water molecules for which the heavy atoms were tethered to preserve the Mg coordination sphere as found in 1BL3. The minimized complex was then subjected to 1000 steps of L-LMOD search to explore low-energy binding modes. Once again, the L-LMOD search was carried out on a fully flexible system, flexible ligand, and flexible protein (except for the tethered catalytic Mg−water coordination sphere). Ligands were used in all-atom representation with 6-31G* ESP charges. L-LMOD docking was combined with explicit translation and rotation of the ligand using BatchMin's MOLS command. Low-energy protein−ligand complexes were saved for visual inspection for our induced-fit study within 50 kJ/mol above the global minimum found by L-LMOD search. Each of the two L-LMOD calculations took approximately 3 CPU days on a single processor of the PC.

The L-LMOD search on the 5-CITEP complex found 54 low-energy binding modes, among which the fifth lowest in energy

* Correspondence can be addressed to either author.
† Gedeon Richter Ltd.
‡ Novartis Institute for Biomedical Research.

(1) Neamati, N.; Turpin, J. A.; Winslow, H. E.; Christensen, J. L.; Williamson, K.; Orr, A.; Rice, W. G.; Pommier, Y.; Garofalo, A.; Brizzi, A.; Campiani, G.; Fiorini, I.; Nacci, V. *J. Med. Chem.* **1999**, *42*, 3334−3341.
(2) Wai, J. S.; Egbertson, M. S.; Payne, L. S.; Fisher, T. E.; Embrey, M. W.; Tran, L. O.; Melamed, J. Y.; Langford, H. M.; Guare, J. P.; Zhuang, L.; Grey, V. E.; Vacca, J. P.; Holloway, M. K.; Naylor-Olsen, A. M.; Hazuda, D. J.; Felock, P. J.; Wolfe, A. L.; Stillmock, K. A.; Schleif, W. A.; Gabryelski, L. J.; Young, S. T. *J. Med. Chem.* **2000**, *43*, 4923−4926.
(3) Zouhiri, F.; Mouscadet, J. F.; Mekouar, K.; Desmaele, D.; Savoure, D.; Leh, H.; Subra, F.; Bret, M. L.; Auclair, C.; d'Angelo, J. *J. Med. Chem.* **2000**, *43*, 1533−1540.
(4) Carlson, H. A.; Masukawa, K. M.; Rubins, K.; Bushman, F. D.; Jorgensen, W. L.; Lins, R. D.; Briggs, J. M.; McCammon, J. A. *J. Med. Chem.* **2000**, *43*, 2100−2114.
(5) Chen, I. J.; Neamati, N.; Nicklaus, M. C.; Orr, A.; Andersom, L.; Barchi, J. J.; Kelley, J. A.; Pommier, Y.; MacKerrel, A. D. *Bioorg. Med. Chem.* **2000**, *8*, 2385−2398.
(6) Lins, R. D.; Briggs, J. M.; Straatsma, T. P.; Carlson, H. A.; Greenwald, J.; Choe, S.; McCammon, J. A. *Biophys. J.* **1999**, *76*, 2999−3011.
(7) Sotriffer, C. A.; Ni, H.; J. A. McCammon *J. Med. Chem.* **2000**, *43*, 4109−4117.
(8) Hazuda, D. J.; Felock, P.; Witmar, M.; Wolfe, A.; Stillmock, K.; Grobler, J. A.; Espeseth, A.; Gabryeslki, L.; Schleif, W.; Blau, C.; Miller, M. D. *Science* **2000**, *287*, 646−650.
(9) Goldur, Y.; Craigie, R.; Cohen, G. H.; Fujiwara, T.; Yoshinaga, T.; Fujishita, T.; Sugimoto, H.; Endo, T.; Murai, H.; Davies, D. R. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 13040−13043.
(10) Greenwald, J.; Le, V.; Butler, S. L.; Bushman, F. D.; Choe, S. *Biochemistry* **1999**, *38*, 8892−8898.
(11) Maignan, S.; Guilloteau, J. P.; Liu, Q. Z.; Mella, C. C.; Mikol, V. *J. Mol. Biol.* **1998**, *282*, 359−368.
(12) Kolossváry, I.; Guida, W. C. *J. Comput. Chem.* **1999**, *20*, 1671−1684.
(13) Kolossváry, I.; Keserû, G. M. *J. Comput. Chem.* **2001**, *22*, 21−30.
(14) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.
(15) McDonald, D. Q.; Still, W. C. *Tetrahedron Lett.* **1992**, *33*, 7747.
(16) Whitlow, M.; Teeter, M. M. *J. Am. Chem. Soc.* **1986**, *108*, 7163.
(17) Keserû, G. M.; Kolossváry, I. *Molecular Mechanics and Conformational Analysis in Drug Design*; Blackwell Science: Oxford, 1999.

10.1021/ja0160086 CCC: $20.00
Published on Web 11/21/2001

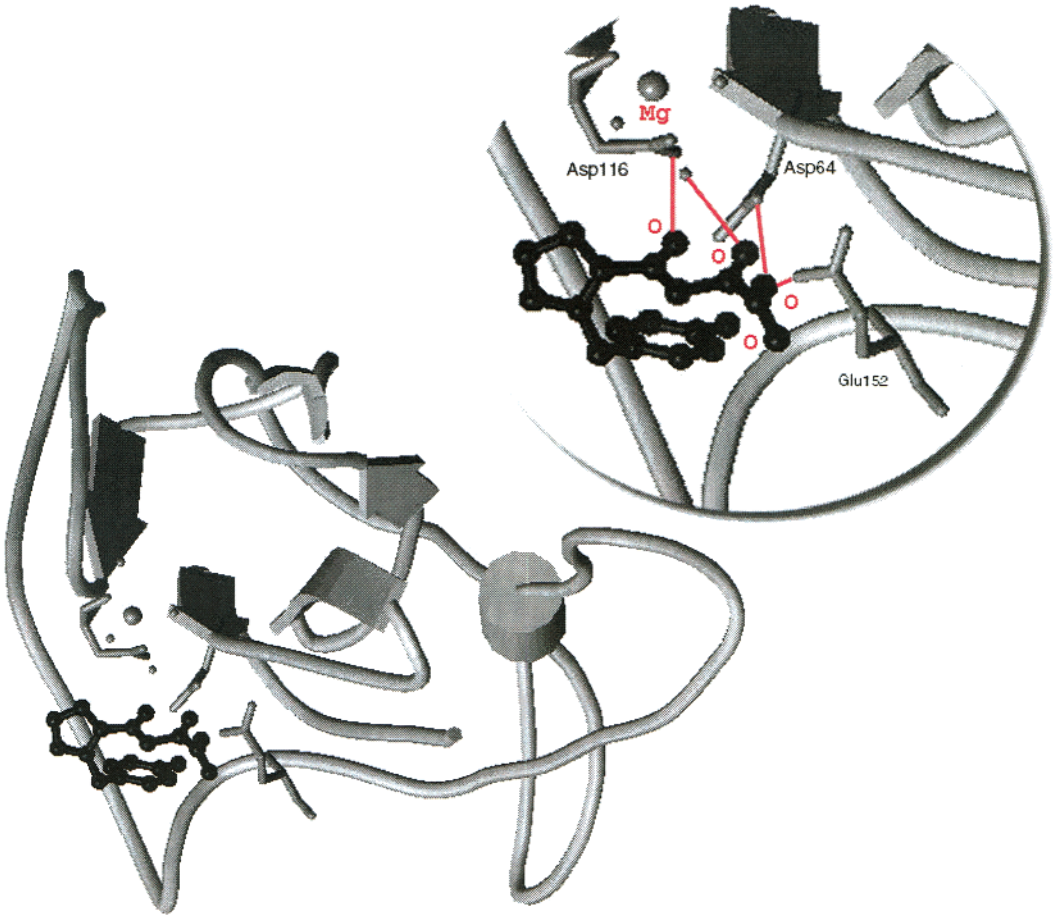

Figure 1. Lowest-energy binding mode of L-731,988 found by L-LMOD search.

Table 1. Calculated and Measured Distances between 5-CITEP and Key Residues in HIV Integrase

residue/atom in 5-CITEP	calcd distance (Å)	measd. distance (Å)
Gln148-O/indole-N	3.4	3.6
Glu152-O/indole-N	3.4	2.9
Thr66-O/tetrazole-N3	3.1	2.7
Asn155-O/tetrazole-N4	4.8	3.5
Lys156-N+/tetrazole-N1	5.6	3.6
Lys159-N+/tetrazole-N2	4.9	2.9

(2 kcal/mol above the global minimum) was the closest to the experimental binding conformation found in 1QS4 (RMSD = 0.392 Å). The effects of induced fit became apparent by including in geometric comparisons not only the ligand but also its contact residues. The average error of experimentally identified key contact distances for this particular binding conformation is 1 Å with respect to 1QS4 (see Table 1). Although the ligand rmsd of the global minimum was even lower (0.272 Å), we found that all of the key contact residues were misoriented.

The L-LMOD search on the Merck inhibitor L-731,988 found 85 low-energy binding conformations among which we found a variety of different binding modes. Our hypothesis of a similar binding mode found for 5-CITEP based on bioisosteric equivalence of the tetrazole moiety in 5-CITEP and the carboxy group of L-731,988 was contradicted by the fact that the lowest-energy such binding mode was only the 20th with a relatively high, 8.7 kcal/mol energy above the global minimum. The global minimum suggests a totally different binding mechanism shown in Figure 1. The carboxy group of L-731,988 forms a strong hydrogen bond with Asp64 and Glu152, and the diketo moiety is tethered by the carboxy side chain of Asp116 and one of the water molecules in the Mg coordination sphere. The global minimum binding mode is in agreement with the findings of Hazuda et al.,[8] suggesting that the Merck inhibitor binds at or near the active site. 5-CITEP is shifted with respect to the suggested binding mode of L-731,-988, away from the catalytic center and binds to key residues listed in Table 1. We suggest that 5-CITEP and the Merck inhibitor bind in different ways, and therefore, in vivo activity of the latter could be attributed to more than just a favorable PK profile. On the computational side, we believe that fully flexible L-LMOD docking follows the trend of successful applications of the low-mode search method in flexible active-site docking,[12] protein loop optimization[13] and will find widespread utility in induced-fit modeling.

Acknowledgment. G.M.K is a Bolyai Reserach Fellow of the Hungarian Academy of Sciences. Financial support of OTKA (Grant. No. F30044 for G.M.K.) is gratefully acknowledged.

JA0160086