Journal of
Medicinal
Chemistry



Volume 44, Number 19 September 13, 2001

# Letters

## Ordered Water and Ligand Mobility in the HIV-1 Integrase-5CITEP Complex: A Molecular Dynamics Study

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

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

*Received May 8, 2001*

**Abstract:** A 2 ns molecular dynamics simulation has been carried out for the HIV-1 integrase-5CITEP complex in order to understand the role of water in defining the ligand's binding mode and to address issues of binding site flexibility and ligand motion. Although the ligand retains considerable mobility within the active site, a structural water molecule bridging 5CITEP with Asp 64 and Asn 155 is identified in the simulation. Consideration of this water molecule could open a route to new HIV-1 integrase inhibitors.

**Introduction.** The structural, dynamical, and functional importance of water molecules for biomolecular structure and recognition is well appreciated. Water is known to contribute significantly to the stability of biomacromolecules and to play a crucial role in molecular association.[1−3] In general, buried water is believed to be involved in local structural stabilization in proteins and DNA[4] and is therefore termed "structural water". Whether particular structural waters are critical for the binding affinity or specificity of protein−ligand complexes is an important issue in the design of new ligands.[5] One prominent example is the observation of a structural water in the crystal structure of nearly every HIV protease−inhibitor complex, which has guided the design of a novel class of tight-binding inhibitors.[6]

X-ray crystallography, NMR spectroscopy, and neutron diffraction are typical experimental methods to analyze water molecules at the atomic level. However, high-resolution structures (better than 1.5 Å) or multiple average-resolution structures are usually preferred for a reliable analysis of structural aspects of ordered water associated with a host protein.[7] In the absence of high-resolution experimental data, theoretical studies, such as molecular dynamics (MD) simulations, represent complementary methods to locate water positions and understand the dynamics and energetics of these water molecules. MD studies of BPTI in solution, for example, were able to reproduce the experimentally determined residence time of water molecules at the protein surface.[8]

HIV-1 integrase (IN), the enzyme responsible for the integration of viral DNA into host cell DNA, represents an attractive, yet unexploited target for the treatment of HIV infections, complementary to reverse transcriptase and protease, the current targets for AIDS drugs. The lack of detailed structural information about IN−ligand interactions has hindered the design of IN inhibitors. The recent appearance of the first and only complex structure of the IN catalytic core domain with an inhibitor (5CITEP) bound to the active site represents, in principle, a starting point for structure-based design efforts.[9] However, the information provided by this structure is not unequivocal. As we have shown in a recent docking study, the position of the ligand appears to be significantly influenced by a crystal packing effect.[10] This study, however, was not suited to obtain information about water interactions in the IN-5CITEP complex since solvation was modeled implicitly, as in most current docking studies. Given the important role that water molecules can play in protein−ligand interactions, it was therefore desirable to apply techniques that go beyond the implicit solvation model.

In the experimentally determined IN complex structure, four water molecules are observed in the vicinity of the ligand. These four water molecules and the side chain oxygen atoms of Asp 64 and Asp 116 are found to be octahedrally coordinated with the active site's $Mg^{2+}$

* Corresponding author: phone, (858)-822-0255; fax, (858)-534-7042; e-mail, hni@mccammon.ucsd.edu.
# Current address: Department of Pharmaceutical Chemistry, Philipps University Marburg, Marbacher Weg 6, D-35032, Marburg, Germany.

10.1021/jm010205y CCC: $20.00 
Published on Web 08/21/2001

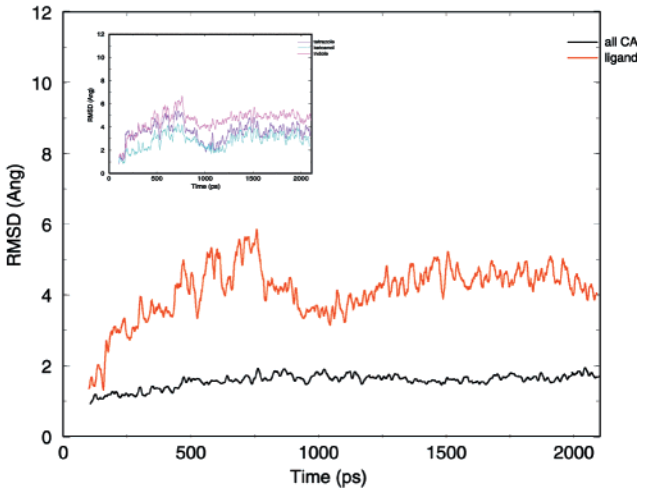


**Figure 1.** Root-mean-square deviation (rmsd) from the starting structure, averaged over 10 ps, measured for all Cα atoms of the protein and all non-H atoms of the ligand. The inset figure shows the rmsd for the three substructures of the ligand: tetrazole ring, keto−enol unit, and chloroindole ring.

ion. One of the four water molecules ("WAT 443") is located between the $Mg^{2+}$ ion and 5CITEP, interacting with the indole ring of the ligand (most probably as a simple van der Waals "spacer" or through a $\pi$-hydrogen bond). This interaction is believed to be a major factor in defining the ligand position (Sugimoto, H., personal communication). However, the experimentally observed position of the ligand in the crystallographic *dimer* could be reproduced by docking studies even without explicit inclusion of this water molecule.[10] In addition, the X-ray structure of the complex was solved at a resolution of 2.1 Å, which may not be sufficient for a reliable, detailed analysis of water molecules.[5] We have therefore carried out a 2 ns MD simulation for the IN-5CITEP complex to better understand the role of water in the binding interface and to analyze protein flexibility and ligand motion.

**Results and Discussion.** As a measure of structural stability, root-mean-square deviations (rmsd) from the starting structure were calculated based on a superposition of all Cα atoms. As shown in Figure 1, the overall Cα-rmsd of IN remains within 1.5−2 Å for the entire simulation, indicating a stable protein structure. On the other hand, the rmsd of the ligand with respect to its starting position occasionally reaches 6 Å, leveling off to 4.5 Å in the second half of the trajectory, suggesting that the ligand displays significant movement within the active site. The rmsd of the tetrazole ring, the keto-enol group, and the chloroindole ring are quite similar (cf. Figure 1), indicating that 5CITEP moves as a whole without major conformational changes during the entire simulation, which is consistent with torsion analyses for 5CITEP. The considerable mobility of the ligand within the active site might explain why the position of the inhibitor could not be unambiguously determined for subunits B and C in the X-ray structure, where in contrast to subunit A there is no symmetry-related neighbor to hold the ligand tightly in place.[9] The movement of the ligand is strongly correlated with Lys 159, Lys 156, and Glu 152, as the distances between the ligand and the three residues remain nearly constant during the simulation. As indicated by the rmsd, the ligand adopts a very different orientation in the MD

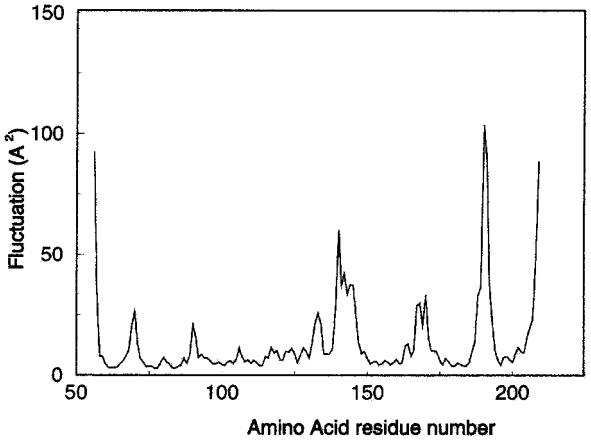


**Figure 2.** Calculated B-factors of the Cα atoms, averaged over subsequent 500 ps time frames.

simulation compared to the X-ray crystal structure. In the MD simulation, the tetrazole ring is placed between Lys 156 and Lys 159, making no interactions with His 67; the keto-enol group is oriented parallel to Glu 152, and the indole ring is located close to Gln 148. The entire ligand forms more interactions with surrounding residues and is more deeply burried in the active site. The observation that in the MD simulation the ligand drifts away from its X-ray starting position, adopting a different orientation within the active site, is consistent with the conclusion of our docking study that the ligand position in the crystal structure is influenced by crystal packing.[10]

To reveal the most flexible regions of the protein structure, B-factors were calculated for the Cα atoms (averaged over three subsequent 500 ps time frames, from 0.5 to 2 ns). As shown in Figure 2, the B-factors display major peaks in the regions of residues 67−71, 138−148, 163−172, and 185−195, with highest B-factors in the latter region. This is consistent with previous observations that the most flexible regions of the enzyme are given by the segments 138−150 and 185−195.[11,12] A comparison of the B-factors calculated here with values obtained from previous MD simulations[13] indicates that the flexible region 138−150 becomes stabilized in the presence of the ligand. The main reason for this stabilization is probably the hydrogen bond network that forms around Gln 148. This residue makes hydrogen bonds with Gln 62, Tyr 143, and Glu 152. The hydrogen bond network seems to be well maintained in the 2 ns simulation, with Gln 148 showing a stable conformation. In addition, the side chain OH of Tyr 143 shows a hydrogen bonding interaction with the side chain OE of Gln 62. The hydrogen bond network in the complex differs significantly from that in the uncomplexed IN form, where a hydrogen bond network among the above-mentioned residues is mediated by water molecules.[14] Since water can exchange among internal hydration sites and possibly exchange with bulk water, the interaction in the uncomplexed form is conceivably less stable than in the complex.

Analyzing the behavior of the active site residues in the MD simulation in comparison to the crystal structure reveals that Asp116 assumes a different conformation, with its $\chi_2$ angle changing from −57° in the crystal to 60° in the first 50 ps of the simulation and the distance between $C_\gamma$ of Asp 64 and Asp 116 decreasing

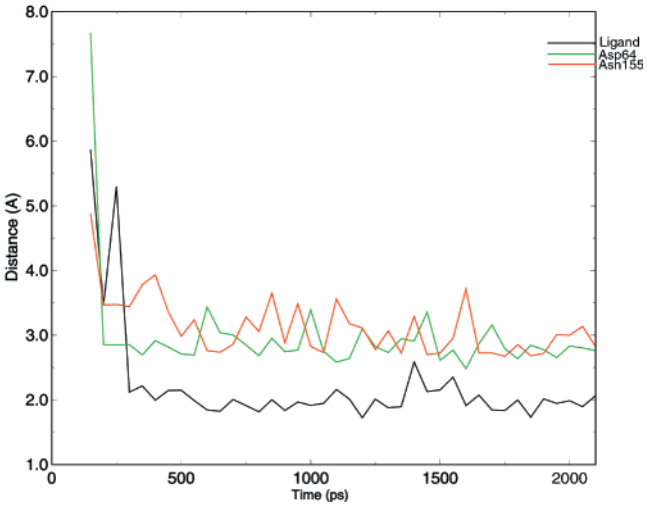


**Figure 3.** Distances between water and HN of the ligand's indole ring (shown in black), between water and O of Asp64 (shown in green), and between water and OD of Asn155 (shown in red), measured at every 50 ps snapshot.

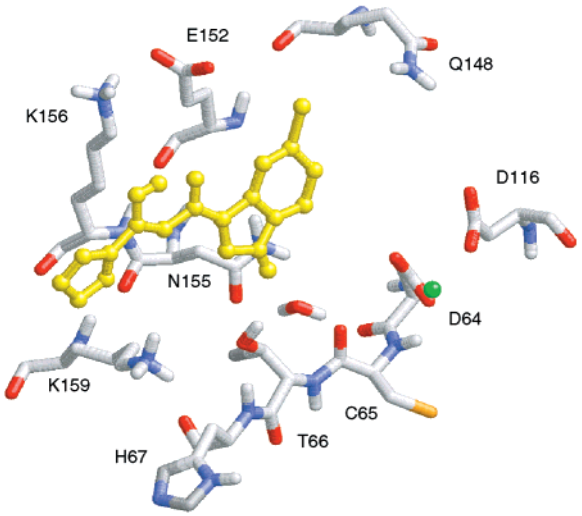


**Figure 4.** Trajectory snapshot of the structural water molecule bound between the ligand (shown in yellow), Asp 64, and Asn 155. Coordinates of this snapshot are available as Supporting Information.

from 5.2 to 3.2 Å. Similar to the X-ray structure, in the MD simulation four water molecules and two oxygens from Asp 64 and Asp 116 are octahedrally coordinated with the $Mg^{2+}$ ion, with an average Mg−O distance of about 2.4 Å. Occasional exchanges of the coordinated water with water molecules from the bulk are observed during the 2 ns of simulation time. It should be noted that none of these four water molecules is in contact with the ligand, a consequence of the different orientation of the ligand and the different conformation of Asp 116 in the simulation.

In the MD simulation, however, another water molecule is found to link the bound inhibitor to IN. This water accepts one hydrogen bond from the NH of the indole ring of 5CITEP and donates two hydrogen bonds to the carbonyl oxygen of Asp 64 and OD of Asn 155. Distance measurements between the oxygen of water and O of Asp 64, OD of Asn 155, and HN of the indole ring are shown in Figure 3, and the snapshot of the structural water molecule bound between the ligand, Asp 64 and Asn 155 is shown in Figure 4. The formation of multiple hydrogen bonds is a known feature of water-mediated protein−ligand complexes.[15] The hydrogen bonds are formed in the first 250 ps and are maintained for the entire course of the simulation. The water molecule shows a remarkably low mobility and does not exchange with bulk water. Apparently, the entropic cost of trapping the water molecule in the binding interface is compensated by the enthalpic gain from making water-mediated hydrogen bonds. The residence time of this water molecule is at least 2 ns, consistent with the NMR observation that the residence time of structural water molecules is greater than 10 ns.[16] However, although this water is trapped in the interface, it is not immobilized. In fact, because of the movement of 5CITEP, this water moves within a maximum range of 3 Å. It is therefore conceivable that similar to the ligand in subunit B and C, also this water molecule could not be traced in the X-ray structure determination. Still, the information about the potential role of a water molecule in mediating IN−ligand interactions should definitely be considered for the design of new inhibitors. Concerning comparisons with water positions in the crystal structure, it should also be emphasized that water molecules observed in the crystal structure did not bias the MD simulation as all crystal waters were deleted during the setup of the simulation. In addition, we have carried out a second MD simulation of equal length where a different ligand position (obtained by docking[10]) was used as starting point: interestingly, the ligand converged to the same binding mode and again a water molecule hydrogen bonding with 5CITEP, Asp 64, and Asn 155 was found, suggesting that the observed structural water position is not merely a consequence of the initial setup. A detailed discussion of this additional simulation will be presented elsewhere.

Although no MD studies have been done to identify structural water in the protein−ligand interface, MD simulations have demonstrated the predictive capability in identifying structural water on protein surfaces. For example, a single conserved water molecule in the family of fatty acid binding proteins has been shown to have a very low mobility in 1 ns MD simulations.[17] Interestingly enough, this water also forms three hydrogen bonds with the protein backbone. In our MD simulation, another water, located outside the active site, was recognized to have a residence time of at least 2 ns. It forms hydrogen bonds with Tyr 83, Ile 84, His 183, and Asp 184. A similar water position has been observed in different crystal structures,[9,11−12] with a B-factor comparable to the average B-factors of the protein atoms. This water may be involved in stabilizing the local structure.

Waters in the binding interface can provide useful information for drug design. Strategies for designing new inhibitors, which take into consideration structural water molecules, include either saturating the water position with appropriate hydrogen bond donors and acceptors to achieve maximum complementarity or replacing it to gain in affinity through entropic effects.[5] As a well-known example, in HIV protease complexes a structural water has been found to play an important role in ligand binding. This water molecule is tetracoordinated with the ligand and the protein by donating two hydrogen bonds to the ligand and accepting two

hydrogen bonds from the protein. The importance of this water molecule has been recognized because of its persistent presence in practically all complex structures of HIV protease and SIV protease. Its incorporation in the design of urea-based ligands resulted in new inhibitors with positive entropic and selectivity benefits.[18,19]

Given the ease by which water can occupy voids and holes in binding sites, it is actually not surprising that water molecules can act as a bridge between ligand and protein and form an integral part of the protein−ligand complex. However, such water molecules need not necessarily be present in the absence of the ligand. In fact, no such water has been observed in the crystal structures of uncomplexed HIV protease, and also for uncomplexed IN this seems to be the case.

**Conclusion.** The presence of water molecules at the binding site has so far mostly been ignored in the design of IN inhibitors because of largely lacking information about water mediation in ligand binding. In this study, we have demonstrated that through an MD simulation of the IN-5CITEP complex a stable water molecule bridging the ligand and IN could be revealed. For the future design of novel specific and tight-binding ligands, it may be important to take this structural water molecule into consideration by either utilizing it as a bridge group or displacing it.

**Methods.** The protein structure used for the MD simulation of the IN-5CITEP complex corresponds to subunit A of PDB file 1QS4,[9] with all crystallographic water molecules removed. It was set up as in the docking study,[10] except that nonpolar hydrogens were added as well to allow the application of the all-atom AMBER95 force field.[20] Parameter assignment for 5CITEP required special attention. Except for the chlorine, all atoms could be assigned standard atom types of the AMBER95 force field. Values for equilibrium bond lengths and angles were taken from the X-ray structure, missing force constants were mainly chosen from a comparison with existing parameters. Atomic charges were calculated by fitting to the electrostatic potential obtained from a HF/6-31G* ab initio calculation with Gaussian98.[21] To check the appropriateness of the ligand parameters, a separate 1 ns MD simulation was run for the ligand in solution. The intramolecular hydrogen bond of the keto-enol group was found to be well-conserved during this simulation. The torsion angles between the keto-enol group and the two aromatic rings were in the expected range according to experimental data and ab initio calculations of torsion potentials carried out previously (Sotriffer and Ni, unpublished results).

Following the initial setup, the complex was energy minimized in vacuo using 50 steps of steepest descent followed by 450 steps conjugate gradient minimization. After solvation in a box of TIP3P water molecules[22] (including the addition of a counterion to ensure neutrality) and minimization of the solvated system (50 steps steepest descent, 250 steps conjugate gradient), the MD simulation was started by heating the solvent to 300 K over a period of 40 ps and cooling to 100 K over a period of 10 ps, keeping the protein complex fixed. After this, the entire system was gradually brought to 300 K over a period of 50 ps. The simulation was then carried on under NPT conditions for 2 ns.

The temperature was kept constant by coupling to a heat bath through the Berendsen algorithm[23] using separate solute and solvent scaling. Pressure was adjusted by isotropic position scaling using a Berendsen-like algorithm. Covalent bonds to hydrogen atoms were constrained by the SHAKE algorithm.[24] A time step of 2 fs was used. A 10 Å cutoff was applied to the van der Waals interactions, while the electrostatics were treated by the Particle Mesh Ewald method.[25] The simulation was carried out with the SANDER MD module of AMBER5.0.[26]

For analysis, energy data were saved every 10 time steps, solute coordinates every 0.1 ps (with snapshots of the solvent every 50 ps). This scheme was changed in the last 100 ps, where coordinates of the entire system (i.e., solute and solvent) where saved every 1 ps. All the results presented in this report refer to the 2.0 ns trajectory which excludes the first 100 ps (temperature adjustment), but includes the last 100 ps with more frequent solvent coordinate saving.

**Acknowledgment.** H.H.N. is an NIH postdoctoral fellow (F32-GM63094). C.A.S. is grateful to the Austrian Science Fund for a postdoctoral fellowship (J1758-GEN). We thank John Mongan and Dr. Richard Henchman for valuable contributions. This work is supported by Grants from NIH, SDSC, and the W.M. Keck Foundation.

**Supporting Information Available:** Coordinates of the snapshot in Figure 4. This material is available free of charge via the Internet at http://pubs.acs.org.

## References

(1) Westhof, E. *Water and Biological Macromolecules*; CRC Press: Boca Raton, FL, 1993.
(2) Karplus, P. A.; Faerman, C. Ordered water in macromolecular structure. *Curr. Opin. Struct. Biol.* **1994**, *4*, 770−776.
(3) Levitt, M.; Park, B. H. Water: now you see it, now you don't. *Structure* **1994**, *1*, 223−226.
(4) Williams, M. A.; Goodfellow, J. M.; Thornton, J. M. Buried waters and internal cavities in monomeric proteins. *Protein Sci.* **1994**, *3*, 1224−1235.
(5) Ladbury, J. E. Just add water! The effect of water on the specificity of protein−ligand binding sites and its potential application to drug design. *Chem. Biol.* **1996**, *3*, 973−980.
(6) Lam, P. Y. S.; Jadhav, P. K.; Eyermann, C. J.; Hodge, C. N.; Ru, Y.; Bacheler, L. T.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Wong, Y. N.; Chang, C. H.; Weber, P. C.; Jackson, D. A.; Sharpe, T. R.and Eickson-Viitanen, S. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* **1994**, *263*, 380−384.
(7) Koellner, G.; Kryger, G.; Millard, C. B.; Silman, I.; Sussman, J. L.; Steiner, T. Active site gorge and buried water molecules in crystal structures of Acetylcholinesterase from Torpedo California. *J. Mol. Biol.* **2000**, *296*, 713−735.
(8) Brunne, R. M.; Liepinsh, E.; Otting, G.; Wuthrich, R.; van Gunsteren, W. F. Hydration of proteins: a comparison of experimental residence times of water molecules solvating the bovine pancreatic trypsin inhibitor with theoretical model calculations. *J. Mol. Biol.* **1993**, *23*, 1040−1048.
(9) Goldgur, Y.; Craigie, R.; Cohen, G. H.; Fujiwara, T.; Yoshinaga, T.; Fujishita, T.; Sugimoto, H.; Endo, T.; Murai, H.; Davis, D. R. Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: A platform for antiviral drug design. *Proc. Natl. Acad. Aci. U.S.A.* **1999**, *96*, 13040−13043.
(10) 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.
(11) Goldgur, Y.; Dyda, F.; Hickman, A. B.; Jenkins, T. M.; Craigie, R.; Davis, 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.
(12) 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. *J. Mol. Biol.* **1998**, *282*, 359−368.

(13) Lins, R. D.; Adesokan, A.; Soares, T. A.; Briggs, J. M. Investigations on human immunodeficiency virus type 1 integrase/DNA binding interactions via molecular dynamics and electrostatics calculations. *Pharmacol. Ther.* **2000**, *85*, 123−131.
(14) 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.
(15) Poornina, C. S.; Dean, P. M. Hydration in drug design. 1. multiple hydrogen-bonding features of water molecules in mediating protein−ligand interactions. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 500−512.
(16) Wiesner, S.; Kurian, E.; Prendergast, F. G.; Halle, B. Water molecules in the binding cavity of intestinal fatty acid binding protein: dynamic characterization by water O-17 and H-2 magnetic relaxation dispersion. *J. Mol. Biol.* **1999**, *286*, 233−246.
(17) Likic, V. A.; Juranic, N.; Macura, S.; Prendergast, F. G. A "structural" water molecule in the family of fatty acid binding proteins. *Protein Sci.* **2000**, *9*, 497−504.
(18) Hodge, C. N.; Straatsma, T. P.; McCammon, J. A.; Wlodawer, A. Rational design of HIV protease inhibitors. *Structural Biology of Viruses*; Chiu, W., Burnett, R. M., Garcia, R., Eds.; Oxford University Press: Oxford, U.K., 1997; pp 451−473.
(19) De Lucca, G. V.; Jadhav, P. K.; Waltermire, R. E.; Aungst, B. J.; Erickson-Viitanen, S.; Lam, P. Y. S. De novo design and discovery of cyclic HIV protease inhibitors capable of displacing the active site structural water molecule. *Pharm. Biotechnol.* **1998**, *11*, 257−284.
(20) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation force-field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(21) Frisch, M. J.; et al. *Gaussian 98*; Gaussian, Inc.: Pittsburgh, PA, 1998.
(22) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. P.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(23) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(24) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(25) Darden, T.; York, D.; Pedersen, L. Partical mesh Ewald-An N.log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(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.

JM010205Y