Binding Free Energy Simulations of the HIV-1 Protease and Hydroxyethylene Isostere Inhibitors

Youngdo Won

Department of Chemistry, Hanyang University, Seoul 133-791, Korea
Received October 6, 2000

The free energy simulation technique is used to evaluate the relative binding affinity of a set of hydroxyethylene isostere inhibitors of the HIV-1 protease. The binding reactions and an alchemical mutation construct the thermodynamic cycle, which reduces the free energy difference of the binding interactions into that of the alchemical processes. In the alchemical process, a methyl group is mutated into a hydrogen atom. Albeit the change is a small perturbation to the inhibitor-protease complex, it results in 25 fold difference in the binding constants. The simulation reproduces the experimentally measured binding affinities within 2% of the free energy difference. The protonation state of the catalytic aspartic acid residues is also investigated through the free energy simulations.

Introduction

The HIV-1 protease catalyzes the cleavage process of the polyprotein precursor and produces proteins needed to mature infectious HIV particles.[1] As it is a unique viral enzyme, inhibition of the enzyme is a promising target for chemotherapy of AIDS. There have been intensive efforts to characterize the enzyme-substrate interactions: several hundred oligopeptide analogs have been synthesized and more than 200 crystal structures of HIV-1 protease complexes have been solved.[2,3] Theoretical investigations have been made to define the quantitative structure-activity relationship and to understand the mode of action at the active site.[4,5] In spite of these efforts, we still do not have a detailed understanding of interactions between the inhibitor and the enzyme.

Molecular dynamics simulations may provide information on the nature of the protease-inhibitor interactions. One calculates the binding free energy difference using the thermodynamic perturbation on dynamics trajectories and directly relates it to the relative binding affinity of inhibitors.[6] The free energy perturbation method has been successfully applied to several ligand-receptor systems,[7,8] including HIV-1 protease complexes.[9-12] The free energy simulation technique was used to predict the relative binding affinity of the *S* and *R* isomers of the JG365 compound[11,12] and the results agreed fairly well with the experimental values reported later.[13] Chen and Tropsha investigated the influence of the active site protonation state while calculating the relative binding affinity of the *S* and *R* isomers of the U85548E inhibitor.[9]

As an aspartic protease, the HIV-1 protease has the highly conserved pair of catalytic triads, Asp25-Thr26-Gly27 and Asp125-Thr126-Gly127. The designation Asp125 is according to Jaskolski *et al.*[14] The protonation state of the catalytic aspartic acids is the key to explain the catalytic mechanism, either general acid-base catalysis or catalysis *via* a covalently bound intermediate. Chatfield and Brooks showed that both types of mechanism were plausible in their molecular dynamics simulation.[15] They suggested that Asp125 was protonated when the reaction began and Asp25 was protonated when the reaction was in progress. Although the neutron scattering experiment may reveal the protonation state, it has not been carried out. In a variety of molecular simulations, both aspartic acid residues were treated as neutral species,[16,17] only one aspartic acid residue was considered neutral,[11,12] or both were treated as negatively charged.[18] Despite progress in molecular simulations and experimental studies, the question of the protonation state remains controversial.

Dryer and coworkers synthesized a series of hydroxyethylene isostere inhibitors and performed structure-activity analysis using enzyme kinetics, X-ray crystallography and infected T-cell assays.[19] The inhibitors Boc-Phe-ψ(CHOH-CH_2)-Ala-Val-NH_2 and Boc-Phe-ψ(CHOH-CH_2)-Gly-Val-NH_2 are different by 25 folds in the binding affinity. The residue Ala is replaced by Gly in the other inhibitor. This is relatively a small change as an "alchemical" perturbation; the methyl group is changed into a hydrogen atom. The small change results in an order of magnitude difference in the binding constant. Furthermore, the crystallographic structure of the HIV-1 protease binding the inhibitor Ala-Ala-Phe-ψ(CHOH-CH_2)-Ala-Ala-Val-OMe has been reported (PDB1AAQ).

The experimental data combined with the crystallographic structure provide us an ideal testing ground to demonstrate the accuracy of the current free energy simulation methods. In this work, the HIV-1 protease complexed with Boc-Phe-ψ(CHOH-CH_2)-Ala-Val-NH_2 is modeled based on the 1AAQ structure and the relative binding free energy is evaluated against that of the Boc-Phe-ψ(CHOH-CH_2)-Gly-Val-NH_2 complex. With minimal use of computational resources, we have been able to reproduce the binding constant ratio within the experimental error limit and to suggest the feasible protonation state of the catalytic aspartic acids through the free energy simulations.

Computational Methods

We obtain thermodynamic properties, including the free

energy, as an ensemble average of the observable in the phase space. The ensemble is generated through molecular dynamics or Monte Carlo simulations. For macromolecular systems such as proteins and nucleic acids in solution, a molecular mechanical model is normally used to represent the hamiltonian of the system. In this work, we used the standard CHARMM version 22 all hydrogen potential function and performed molecular dynamics simulations.[20,21] The dynamics procedure results in the trajectory of atoms, which constructs the configurational space.

The classical configurational partition function of the system is the state sum in the configurational space.[6]

$$Z^C = \sum_k e^{-\beta V_k(\mathbf{r})} = \int d\mathbf{r}^N \exp[-\beta V(\mathbf{r}^N)] \quad (1)$$

$V(r^N)$ is the potential energy of the system and $\beta = 1/k_B T$. The Helmholtz free energy of an NVT ensemble is

$$A = C(N, V, T) - k_B T \ln Z^C \quad (2)$$

where $C(N, V, T)$ is a constant related to the kinetic energy of the system. k_B is the Boltzmann constant and T is the absolute temperature. We consider the Gibbs free energy, G, of the NPT ensemble equivalent to the Helmholtz energy obtained from the NVT ensemble.[22]

The molecular system of interests includes the enzyme (E), the inhibitor (I) and water molecules as the solvent. The binding process is the reaction

$$E(aq) + I(aq) \rightarrow E \cdot I(aq)$$

The binding constant, K_i, is defined as the equilibrium constant of the reverse reaction and is directly related to the Gibbs free energy of the binding process, $\Delta G = RT \ln K_i$. R is the gas constant. While there are some simulation techniques, *e.g.*, utilization of an umbrella potential or a bias potential, it is much difficult to simulate the binding process. No inhibitor diffuses out or into the active site of the enzyme during a finite sampling period.

As we intend to obtain the thermodynamic state function, an elegant way to approach the problem is to devise the thermodynamic cycle of Figure 1. The free energy difference $\Delta\Delta G = \Delta G_1 - \Delta G_2$ is exactly equal to $\Delta G_a - \Delta G_b$ to complete the cycle. Instead of the processes 1 and 2, the processes *a* and *b* are used to yield the difference in the thermodynamic property. The diffusion processes can be replaced by the I_1 to I_2 mutation processes; free in solution (*a*) and bound to the enzyme (*b*). The free energy difference directly yields the ratio of the binding constants.

$$\Delta\Delta G = \Delta G_1 - \Delta G_2 = RT \ln \frac{K_1}{K_2} \quad (3)$$

K_1 and K_2 are binding constants of inhibitors I_1 and I_2, respectively.

The free energy change ΔG is calculated in the thermodynamically perturbed molecular dynamics simulation. Here we compare the binding affinity of Boc-PheψAla-Val-NH_2(I_1) and Boc-PheψGly-Val-NH_2(I_2) and need to construct molecular models for the inhibitor I_1 solution and the I_1 and HIV-1 protease complex. The stochastic boundary molecular dynamics (SBMD) simulation technique is applied to the molecular systems. The inhibitor I_1 is generated and immerged into a water sphere with the diameter of 18.5 Å. The center of the coordinates of the inhibitor is placed at the origin of the sphere. The water sphere was pre-equilibrated at the simulation temperature, 300 K. The overlapping water molecules with the inhibitor are removed and the system is briefly minimized in energy to ensure the optimal position of water molecules surrounding the inhibitor. A water molecule is removed if its oxygen atom is placed away more than 12 Å from the center of the sphere. The stochastic boundary potential is imposed on the 12 Å spherical boundary. The outside of the 12 Å radius sphere is considered to act as a heat bath, the "reservoir" region. The Langevin

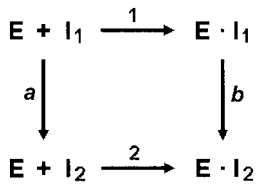

Figure 1. The thermodynamic cycle of inhibitor binding processes. The processes 1 and 2 represent complex formation of the enzyme with the inhibitor I_1 and I_2, respectively. In the processes *a* and *b*, the inhibitor I_1 is mutated to I_2.

Figure 2. The schematic drawing of the HIV-1 protease structure. helices are shown in black, the A chain is in dark grey and the B chain is in light grey. The catalytic aspartates are shown in sticks in the center cavity. The ball and stick model represents the inhibitor Boc-PheψAla-Val-NH_2.. The crystal water molecule is also shown in the active site.

Table 1. Feasible hydrogen bonds of titratable residues[a]

H-Bond Donor	H-Bond Acceptor	D-A distance (Å)	A-H-D angle (deg.)
Asp25 OD1-HD1	1AAQ ψ O	2.901	158.8
Asp25 OD2-HD2	1AAQ ψ O	3.062	149.6
Asp125 OD1-HD1	1AAQ ψ O	2.868	154.2
Asp125 OD2-HD2	1AAQ ψ O	2.760	139.6
His69 ND1-HD1	Cys67 O	2.683	123.1
His169 ND1-HD1	Cys167 O	2.748	126.1

[a]1AAQ ψ O is the hydroxyl oxygen atom of the inhibitor Ala-Ala-Phe-ψ(CHOH-CH_2)-Ala-Val-Val-OMe and the Cys O is the carbonyl oxygen atom of the peptide bond.

equation of motion is applied to the atoms in between 10 Å and 12 Å from the origin, which is the "buffer" region of the spherical stochastic boundary model. For atoms within the 10 Å radius sphere, the "reaction" region, normal Newtonian equation of motion is used to produce the trajectory. The leap-frog integrator is used to integrate the equations of motion.

The molecular model of the I_1 and enzyme complex is based on the 1AAQ coordinates obtained from the protein data bank. The 1AAQ coordinates are the X-ray crystallography data at the 2.5 Å resolution.[19] It contains the HIV-1 protease complexed with the hydroxyethylene isostere inhibitor, Ala-Ala-Phe-ψ(CHOH-CH_2)-Ala-Ala-Val-OMe. It is therm P_{61} crystal with the crystal constants $a = b =$ 63.3 Å, c = 83.6 Å, $\alpha = \beta = 90^o$, and $\gamma = 120^o$. There is one crystal water molecule in the structure and it contains total of 1,560 heavy atom coordinates. The HIV-1 protease is a homodimer of the polypeptides of 99 residues. Each chain contains one α helix of Arg87 through Gln92, 9 β sheets and four turns. There is no disulfide linkage. The 1AAQ inhibitor is modified to Boc-Phe-ψ(CHOH-CH_2)-Ala-Val-NH_2 (I_1). The amino terminal Ala-Ala is removed and capped with the *tert*-butyloxycarbonyl (Boc) group and the carboxyl terminal Ala-Val-OMe is replaced with the valine residue capped with the amino group. The schematic drawing of the protease-inhibitor complex is in Figure 2.

As the X-ray technique determines no hydrogen coordinates, one needs to evaluate the protonation state of titratable residues, His, Glu, Asp, and Lys. pK_a of Lys ε-NH_3^+ in an aqueous solution is 10.5. 12 lysine residues of the protease are with ε-NH_3^+. There are only two histidines, His69 and His169. The imidazole has pK_a = 6.4 and needs close examination. We protonated both imidazole nitrogens and examined hydrogen bonding feasibility. Table 1 lists feasible hydrogen bonds of His69 and His169, which shows that the histidines are protonated at the δ-N atom. The side chain carboxyl group of aspartic and glutamic acids has pK_a = 4 approximately.[23] The residue is in the carboxylate form in an aqueous solution. When these residues are imbedded in the protein matrix and interact with a hydrogen bond acceptor, they may have a proton. We examined hydrogen bonding feasibility of each carboxyl group of 8 aspartic acids and 8 glutamic acids and found only four feasible hydrogen bonds listed in Table 1. Both catalytic aspartic acids interact with the hydroxyl group of the hydroxyethylene isostere inhibitor. One hydrogen on Asp25 or on Asp125 may constructs the hydrogen bonding network in the active site. The protonation state of the catalytic aspartates has been in question.[9,14] One way to address the question is to carry out simulations of the protease complex at each different protonation state. We performed five set of free energy simulations in this work; one with no hydrogen placed on the catalytic aspartic acids and four cases with one carboxylic oxygen protonated, respectively.

With the protonation state of titratable residues is determined, hydrogens are placed into the protease-inhibitor complex structure by using the CHARMM HBUILD function. The entire complex is translated so that the center of coordinates of the inhibitor may be at the coordinate origin. The system is hydrated with the 18.5 Å water sphere. Water molecules overlapping with a heavy atom of the complex are removed. A brief energy minimization (100 steps of the adopted basis Newton-Raphson method) is performed to remove any abnormal strains in the resultant structure. We employ 15 Å radius spherical boundary with the 2 Å crust buffer region. All atoms within 13 Å from the origin (in the reaction region) move under the normal Newtonian mechanics and atoms between 13 Å and 15 Å spheres (in the buffer region) evolve according to the Langevin equations of motion subject to the Langevin random force, the frictional force proportional to the velocity and the stochastic boundary force. Atoms outside of the 15 Å radius sphere (in the reservoir region) are fixed during the dynamics integration. The Leap-Frog integrator is employed in the dynamics simulation.

The free energy changes of the processes *a* and *b* of Figure 1 are calculated in the thermodynamically perturbed molecular dynamics simulations of the inhibitor solution and the solvated protease-inhibitor complex, respectively. The process is essentially the mutation of alanine to glycine; the methyl group is mutated to a hydrogen. The I_1 to I_2 mutation entails change, creation and annihilation of atoms. As atoms are represented by the mass, the charge, connectivity and interaction energy parameters, we may accomplish such mutation in several ways. The "alchemical" mutation scheme depends on the free energy simulation method. There are three free energy codes implemented in CHARMM. If the perturbation (PERT) code is used, one possibility is to annihilate three methyl hydrogens and mutate the methyl carbon into a hydrogen. For the block hamiltonian method (BLOCK) and the thermodynamic simulation method (TSM) codes, the methyl group of alanine is put into a block (H_{Ala}) and the alpha hydrogen of glycine is put into another block (H_{Gly}) of the hamiltonian. The charge on the alpha carbon is adjusted as the mutation progresses.

In this work, the TSM function of CHARMM is used. The alanine to glycine mutation is performed by the PATCH command of CHARMM. The patch modifies the molecular model of Boc-PheψAla-Val-NH_2 by adding a hydrogen atom connected to the alpha carbon of the alanine residue.

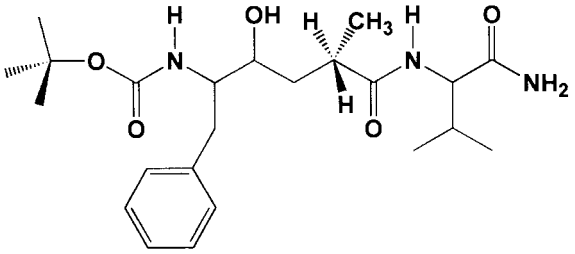

Figure 3. The inhibitor Boc-PheψAla-Val-NH_2 patched for the Ala to Gly mutation. The methyl group attached to the alpha carbon at the P_1' position gradually disappears and the hydrogen atom grows as the perturbation progresses. The valence of the alpha carbon is effectively four as it should be.

The total hamiltonian is

$$H = H_0 + (1-\lambda)\, H_{\mathrm{Ala}} + \lambda H_{\mathrm{Gly}} \tag{4}$$

H_{Ala} represents the alanine residue (CH_3) and H_{Gly} is the mutant glycine residue (H). H_0 is the hamiltonian of the rest of the system. λ is the perturbation parameter increasing from zero to one as the mutation progresses. $\lambda = 0$ is the states of the wild type residue and $\lambda = 1$ is the state of the mutant residue. The I_1 to I_2 mutation corresponds to the $\lambda = 0$ to $\lambda = 1$ perturbation in the molecular dynamics simulation. In the TSM code, the alpha carbon of the mutating residue is the co-located (COLO) atom common to both the wild and mutant residues. Its charge is scaled by the perturbation parameter λ, $(1-\lambda)(-0.09) + \lambda(-0.18)$, and set by the TSM COLO command. Figure 3 depicts the inhibitor Boc-PheψAla-Val-NH_2 with the patch for alanine to glycine mutation.

The potential energy corresponding to the hamiltonian in Eq. (4) can be written as

$$V = V_0 + \lambda V_\lambda \tag{5}$$

where V_0 is the unperturbed potential energy and V_λ is the perturbation, $V_\lambda = V_{\mathrm{Gly}} - V_{\mathrm{Ala}}$. In general, the unperturbed state serves as the reference state. The configurational partition function is

$$Z_\lambda^C = \int d\boldsymbol{r}^N \exp[-\beta\{V_0(\boldsymbol{r}^N) + \lambda V_\lambda(\boldsymbol{r}^N)\}] \tag{6}$$

By division and multiplication with the configurational partition function of the reference state,

$$Z_0^C = \int d\boldsymbol{r}^N \exp[-\beta V_0(\boldsymbol{r}^N)] \tag{7}$$

the configurational partition function in Eq. (6) is reduced to

$$Z_\lambda^C = Z_0^C \langle \exp[-\beta\lambda V_\lambda(\boldsymbol{r}^N)]\rangle_0 \tag{8}$$

$\langle ... \rangle_0$ denotes the canonical average at the reference state. The free energy due to the perturbation λ on the reference state is

$$\Delta A_\lambda = A_\lambda - A_0 = -k_B T \ln \langle \exp[-\beta\lambda V_\lambda(\boldsymbol{r}^N)]\rangle_0 \tag{9}$$

When λ and/or V_λ are small, the free energy is reduced to

$$\Delta A_\lambda = \lambda \langle V_\lambda(\boldsymbol{r}^N)\rangle_0 \tag{10}$$

which is proportional to the canonical average of the perturbation at the reference state.

The thermodynamic perturbation is performed through the perturbation windows at $\lambda = 0.1$, 0.3, 0.5, 0.7 and 0.9. These five states are the reference states and the forward (+0.1) and backward (–0.1) summations yield the free energy ΔA. The initial coordinates of each window are set to the equilibrated coordinates of the previous window except those of $\lambda = 0.1$ which are set to the thermalized coordinates of the unperturbed system ($\lambda = 0$). For each window, the equilibration dynamics is performed at 300K for 5ps with the initial coordinates, then sampling is carried out through the SBMD simulation. In order to calculate the free energy of Eq. (10), the perturbation in potential energy is saved at each dynamics step, *i.e.*, in every 1fs.

Results and Discussion

The unperturbed systems, Boc-PheψAla-Val-NH_2 and its HIV-1 complex solvated in the 13 Å and 15 Å radius water spheres, respectively, are thermally equilibrated at 300 K. The Langevin heat bath is set to 300 K and the equilibration dynamics is performed for 10ps. Figure 4 shows the temperature and the potential energy as the dynamics progresses. The first 10ps is for the thermal equilibration phase and the following 15ps is the dynamics profile of the first perturbation window ($\lambda = 0.1$). Both systems are quickly thermalized in 3ps and equilibrated for the rest 7ps. At the end of the thermalization dynamics of 10ps, we turn on the perturbation by applying the alanine to glycine mutation patch with λ set to 0.1. The perturbed system is equilibrated for 5ps and the ensemble of configuration is generated with the equilibrated system.

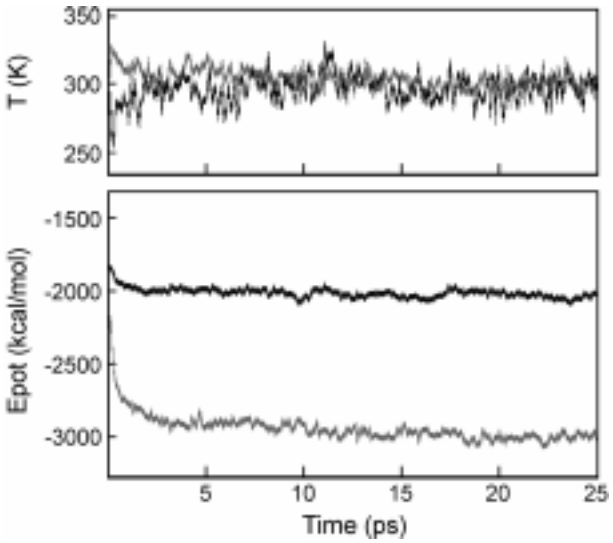

Figure 4. Dynamics profiles of the solution (black solid lines) and the complex (grey lines) systems showing temperature and the potential energy. 0 to 10ps is for the thermal equilibration stage of the unperturbed system and 10 to 15ps is for the equilibration of the perturbed system with $\lambda = 0.1$. Sampling dynamics of the $\lambda = 0.1$ system runs from 15ps.

Table 2. The average and the fluctuation in the temperature and the potential energy of 10ps sampling dynamics[a]

λ	Solution		Complex	
	T(K)	E_{pot} (kcal/mol)	T (K)	E_{pot} (kcal/mol)
0.1	295 ± 10	−2049 ± 27	298 ± 5	−3007 ± 32
0.3	293 ± 9	−2053 ± 19	297 ± 6	−3028 ± 26
0.5	292 ± 10	−2060 ± 18	300 ± 5	−3023 ± 28
0.7	290 ± 13	−2083 ± 29	304 ± 5	−2992 ± 26
0.9	298 ± 11	−2047 ± 17	301 ± 5	−3005 ± 26

[a]the standard deviation is given as the error estimate.

The $\lambda = 0.1$ dynamics profile from 15 to 25ps in Figure 4 shows that sampling of solution configurations is performed at 295 ± 10 K with the potential energy of −2049 ± 27 kcal/mol. The complex system is equilibrated at 298 ± 5 K and with the potential energy of −3007 ± 32 kcal/mol. We have observed similar dynamics behavior for other windows as shown in Table 2. Thermal fluctuation of the solution dynamics is as twice large as that of the complex dynamics. It requires longer sampling run for the solution dynamics. Note that the potential energy is not a linear functional of the perturbation parameter, λ. The energetics must be complemented by the entropy of the system to reveal the thermodynamics of the process.

The convergence of the free energy is checked by extending the dynamics sampling by 10ps. Table 3 summarizes the free energies with the standard deviation obtained from 10ps, 20ps and 30ps dynamics sampling. Based on the data of Table 3, we perform 30ps sampling for the solution free energy simulation and 10ps for the complex free energy simulations. The cumulative free energy changes for the solution and the complex obtained from the 30ps and 10ps trajectories, respectively, are visualized in Figure 5. Each window yields two free energy differences, both forward and backward; $\lambda \pm \Delta\lambda$. Here we used five windows ($\lambda = 0.1$, 0.3, 0.5, 0.7 and 0.9) with $\Delta\lambda = 0.1$.

The HIV-1 protease has two catalytic aspartates, Asp25 and Asp125. The protonation state has not been clearly determined. The above free energy simulation is performed with both aspartates unprotonated. Four different protonation possibilities listed in Table 1 are examined by carrying out free energy simulations. In each simulation, the molecular model of the HIV-1 protease is prepared with one of the carboxylate oxygens protonated. The same protocol as that for "both aspartate" simulation is applied to these simulations. The resultant free energy changes are collected in Table 4. The $\Delta\Delta G$ ($= \Delta G_{solution} - G_{complex}$) values are also shown in the table.

Table 3. The free energy change and the fluctuation[a]

Sampling Time	Solution		Complex	
	ΔG	Std. Dev.	ΔG	Std. Dev.
10ps	−28.59	0.33	−27.44	0.14
20ps	−29.22	0.24	−27.59	0.13
30ps	−29.33	0.23	–	–

[a]all values are given in kcal/mol.

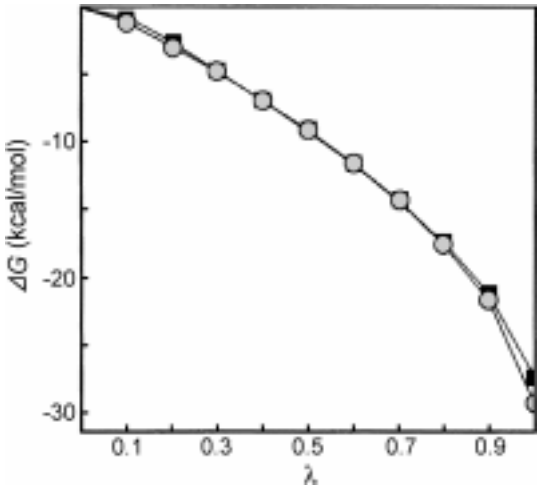

Figure 5. The cumulated free energy of the inhibitor mutation in solution (○) and in the solvated complex (■) as the alanine to glycine mutation progresses. Both catalytic aspartates of the protease are not protonated.

Table 4. The free energy changes of the inhibitor mutation in solution and in the solvated complexes

Simulation	ΔG (kcal/mol)	$\Delta\Delta G$ (kcal/mol)
Solution	−29.33	–
Both aspartate	−27.44	−1.89
Asp25-OD1-H	−32.46	3.13
Asp25-OD2-H	−27.54	−1.79
Asp125-OD1-H	−29.06	−0.28
Asp125-OD2-H	−31.94	2.61

The experimentally determined binding constants are $K_1 = 260 \pm 28$ for Boc-PheψAla-Val-NH_2 and $K_2 = 6500 \pm 700$ for Boc-PheψGly-Val-NH_2. The ratio is $K_1/K_2 = 0.04$ and the experimental free energy difference is $\Delta\Delta G = (1.987 \times 10^{-3})(300) \ln 0.04 = -1.92$ kcal/mol. This value is reproduced by the free energy simulation of the complex with no aspartate protonated, $\Delta\Delta G = -1.89$ kcal/mol within the experimental uncertainly. If one aspartate oxygen is protonated, the hydrogen should be on the Asp25 OD2 atom ($\Delta\Delta G = -1.79$ kcal/mol) when the HIV-1 protease binds the hydroxyethylene inhibitor.

Conclusion

Thermodynamic perturbation drives the "alchemical" alanine to glycine mutation in the molecular dynamics simulation under the stochastic boundary. In SBMD, the system size is effectively reduced to 701 atoms for the solution and 1927 atoms for the complex, respectively. The enzyme with the 15 radius water sphere contains about six thousand atoms. Complete solvation of the enzyme-inhibitor complex in a normal periodic boundary would result in several hundred thousands of atoms in the system. In order to ensure the convergence of the free energy, 30ps sampling dynamics is

needed for each window of the solution simulation. For the complex system, 10ps sampling is enough to produce similar or better statistics. While the system size of the complex is larger than that of the solution sphere, the reservoir region atoms are fixed in the complex system, which results in smaller fluctuation. The total simulation time for a complex system adds up to 75ps (5ps equilibration and 10ps sampling in each window) and costs approximately 15 CPU hours on a Cray Y-MP C90 platform.

While the computational resource is easily affordable, the free energy simulation in this work reproduces successfully the experimental value, −1.92 kcal/mol. The simulation yields −1.89 kcal/mol and the difference is only 0.03 kcal/mol, which is within 2% error. The experimental binding constants are measured with more than 10% error. Although the current free energy simulation methods do not yield the absolute free energy, they are reliable enough to estimate the binding affinity difference.

Four possible protonation states of the catalytic aspartic acid residues of the HIV-1 protease were separately incorporated into the complex system and free energy simulation was performed respectively. Only the case with Asp25 OD2 protonation results in the reasonable binding affinity. As the hydroxyethylene inhibitor is a transition state mimic, the free energy simulation results indicate that Asp25 is likely protonated at the second step of the general acid-general base mechanism with the neutral intermediate described in ref. 15.

Acknowledgment. This work has been supported by the Basic Science Research Institute Program, the Korea Research Foundation, and the Cray University Research and Development Grant. The facility of the KORDIC super computing center was extensively used to run the simulation code.

References

1. Krausslich, H. G.; Wimmer, E. *Ann. Rev. Biochem.* **1988**, *57*, 701.
2. Wlodawer, A.; Erickson, J. W. *Ann. Rev. Biochem.* **1993**, *62*, 543.
3. Kuntz, I. D.; Meng, E. C.; Shoichet, B. K. *Accounts Chem. Res.* **1994**, *27*, 117.
4. Hyland, L. J.; Tomasek Jr., T. A.; Meek, T. D. *Biochemistry* **1991**, *30*, 8454.
5. Goldblum, A.; Rayan, A.; Fliess, A.; Glick, M. *J. Chem. Inf. Computer Sci.* (USA) **1993**, *33*, 270.
6. Brooks III, C. L.; Karplus, M.; Pettitt, B. M. *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics Adv. Chem. Phys. LXXI*; John Wiley & Sons, Inc.: New York, 1988; p 66.
7. McCammon, J. A.; Straatsma, T. P. *Ann. Rev. Phys. Chem.* **1992**, *43*, 407.
8. Kollman, P. A. *Chem. Rev.* **1993**, *93*, 2395.
9. Chen, X.; Tropsha, A. *J. Med. Chem.* **1995**, *38*, 42.
10. Reddy, M. R.; Viswanadhan, V. N.; Weinstein, J. N. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 10287.
11. Tropsha, A.; Hermans, J. *Protein Eng.* **1992**, *51*, 29.
12. Ferguson, D. M.; Radmer, R. J.; Kollman, P. A. *J. Med. Chem.* **1991**, *34*, 2654.
13. Rich, D. H.; Sun, C.; Vara-Prasad, J. V.; Pathiasseril, A.; Toth, M. V.; Marshall, G. R.; Clare, M.; Mueller, R. A.; Houseman, K. *J. Med. Chem.* **1991**, *34*, 1222.
14. Jaskolski, M.; Tomasselli, A. G.; Sawyer, T. K.; Staples, D. G.; Heinrikson, R. L.; Schneider, J.; Kent, S. B.; Wlodawer, A. *Biochemistry* **1991**, *30*, 1600.
15. Charfield, D. C.; Broks, B. R. *J. Am. Chem. Soc.* **1995**, *117*, 5561.
16. Swaminathan, S.; Harte, W. E.; Beveridge, D. L. *J. Am. Chem. Soc.* **1991**, *113*, 2717.
17. Harte, W. E.; Berveridge, D. L. *J. Am. Chem. Soc.* **1993**, *115*, 3883.
18. York, D. M.; Darden, T. A.; Pedersen, L. G.; Anderson, M. W. *Biochemistry* **1993**, *32*, 1443.
19. Dryer, G. B.; Lambert, D. M.; Meek, T. D.; Carr, T. J.; Tomaszek Jr., T. A.; Fernamdez, A. V.; Bartus, H.; Cacciavillani, E.; Hassell, A. M.; Minnich, M.; Petteway Jr., S. R.; Metcalf, B. W. *Biochemistry* **1992**, *31*, 6646.
20. Brooks, S. R.; Bruccoleir, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
21. MacKerell Jr., A. D.; Brooks, B.; Brooks III, C. L.; Nilsson, N.; Roux, B.; Won, Y.; Karplus, M. In *Encyclopedia of Computational Chemistry*; van Ragu Schleyer, P., Allinger, N. L., Kollman, P. A., Clark, T., Schaefer III, H. F., Gasteiger, J., Eds.; John Wiley & Sons, Inc.: New York, 1998; p 271.
22. McQuarrie, D. A. *Statistical Mechanics*; Harper & Row Publishers Inc.: New York, 1976; p 63.
23. Voet, D.; Voet, J. G. *Biochemistry*; John Wiley & Sons, Inc.: New York, 1985; p 58.