JCTC Journal of Chemical Theory and Computation

Calculation of the Relative Binding Affinity of Enzyme Inhibitors Using the Generalized Linear Response Method

Xin Chen[†,‡] and Alexander Tropsha*[,†]

Laboratory for Molecular Modeling, Division of Medicinal Chemistry and Natural Products, School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, and Computer Assisted Drug Discovery, Research and Early Development, Johnson & Johnson Pharmaceutical Research and Development, L.L.C., 920 Route 202, Raritan, New Jersey 08869

Received February 20, 2006

Abstract: The generalized linear response (GLR) method initially developed for *hydration* free energy calculations has been adapted for *binding* free energy calculations. The calculations employ the concept of thermodynamic cycle. To obtain the value of the relative binding free energy between two ligands, we run molecular dynamics simulations at only four "midpoint" states along the thermodynamic pathways connecting the two ligands in the unbound and bound states, respectively. This approach significantly simplifies and accelerates the calculations as compared to the traditional free energy simulations where significantly more intermediate states are usually sampled. We show that each of these "midpoint" states can be approximately defined by a modified force field function in which both the van der Waals and electrostatic interactions between the variant part of a ligand and its environment, either binding site or aqueous solution, are scaled by a factor of 0.5. We tested this new approach to relative binding affinity calculations on the HIV-1 protease complex with its inhibitor JG365 as a starting point for the following two structural transformations: (a) the critical chiral center on the central residue was changed from (*S*) to (*R*) configuration, and (b) the C terminal valine residue was deleted. In both cases, the GLR method afforded calculated values that were in good agreement with the experimental data.

Introduction

The success of structure-based drug design approach relies on our ability to predict, rapidly and accurately, the binding modes and affinities of ligands that interact with their macromolecular receptors.[1] In general, current computational approaches to modeling ligand−receptor interaction can be divided into two categories: qualitative (empirically based) and quantitative (first-principle based) methods.[2] Qualitative modeling methods typically rely on empirical measures of geometric and/or chemical complementarities between ligands and receptors.[3] These methods have not only the advantages of speed and simplicity but also the limited ability to differentiate between structurally similar ligands despite some recent efforts to improve their accuracy by taking into account the conformational flexibility[4−6] and the solvation effect.[7,8] As a result, they may be only capable of identifying weak active hits from database screening.

Quantitative modeling methods are usually based on molecular simulations, such as free energy perturbation (FEP) and thermodynamic integration (TI) that use molecular dynamics or Monte Carlo simulations as the sampling technique.[9−12] These methods are theoretically sound and in principle able to afford accurate calculation of relative

* Corresponding author fax: (919)966-0204; e-mail: alex_tropsha@unc.edu.

† University of North Carolina at Chapel Hill.

‡ Johnson & Johnson Pharmaceutical Research and Development.

10.1021/ct600071z CCC: $33.50
Published on Web 07/04/2006

binding affinity of structurally similar ligands. Typically, they can provide calculated values within 1.5 kcal/mol range of the experimental data.[12] Adequate sampling is a key problem for these methods due to the extremely high computational cost usually required to sample all microscopic states accessible to the system to obtain a convergent calculated value. Another problem is the "force field problem", which implies an insufficiently accurate description of van der Waals and electrostatic interactions in ligand−receptor systems, although certain progress has been made to circumvent these problems as well.[13,14]

Several years ago, Aqvist et al.[15] and Jorgensen et al.[16] proposed a new approach to accelerate free energy calculations based on the linear response approximation. This approach avoids the exhaustive sampling at multiple intermediate states along the transformation path from the initial to final state as in traditional free energy simulations and therefore provides a computationally more efficient route for free energy calculations. According to Aqvist et al., the binding free energy $\Delta G_{\text{binding}}$ can be calculated from the change of electrostatic and van der Waals interaction energies between a ligand and its environment as it is transferred from aqueous solution to binding site, as follows:[15]

$$\Delta G_{\text{binding}} = 1/2 \cdot \Delta\langle V_{\text{elec}}\rangle + \alpha \cdot \Delta\langle V_{\text{vdw}}\rangle \quad (1)$$

Here, $\langle V_{\text{elec}}\rangle$ and $\langle V_{\text{vdw}}\rangle$ are the ensemble averages of the electrostatic and van der Waals interaction energies, respectively, between the ligand and its environment, and Δ refers to the difference between the ensemble averages in the aqueous solution and in the binding site. The 1/2 coefficient before the electrostatic contribution term can be derived from the standard theory behind the linear response approximation, whereas the α coefficient before the van der Waals contribution term is determined by empirical calibration against the experimental data. Equation 1 was first applied to calculate the binding affinities of five structurally different inhibitors of endothiapepsin, and it was found that the α value of 0.162 gave the best agreement with the experimental data with a mean unsigned error of ca. 0.3 kcal/mol.[15] Subsequently, Aqvist's group has successfully applied this approach to several other ligand−receptor systems.[17−22]

Carlson and Jorgensen[16] proposed a slightly different version as an extended linear response equation for the hydration free energy calculations. They introduced an additional cavity contribution term and included adjustable coefficients before all terms. Thus, the hydration free energy ΔG_{hyd} is calculated as[16]

$$\Delta G_{\text{hyd}} = \beta \cdot \langle V_{\text{elec}}\rangle + \alpha \cdot \langle V_{\text{vdw}}\rangle + \gamma \cdot (\text{cavity\_term}) \quad (2)$$

Here, $\langle V_{\text{elec}}\rangle$ and $\langle V_{\text{vdw}}\rangle$ are the ensemble averages of the electrostatic and van der Waals interaction energies between a solute and aqueous solvents. The coefficients, α, β, and γ, are adjustable parameters, which are obtained by fitting to the experimental hydration free energies. The cavity contribution term could represent one of three different molecular properties: molecular surface area, solvent-accessible surface area, or molecular volume. Equation 2 was first applied to 14 diverse organic compounds. It was found that when the cavity term represented the solvent accessible surface area, the best agreement with the experimental data was obtained with RMS deviations of ca. 0.8 kcal/mol for the all-atom model and 0.9 kcal/mol for the united-atom model. Jorgensen et al. then expanded the test data set to 35 organic compounds and obtained an RMS deviation of ca. 1.0 kcal/mol from the experimental data.[23] They have also successfully applied their approach to several ligand−receptor binding systems.[24−27]

The published work on extending the standard linear response approximation to free energy calculations has demonstrated the efficiency of this approach. However, the empirical parameters, such as α, β, and γ used in eqs 1 and 2, were derived from the analysis of specific training sets. Consequently, as with any empirically derived parameters, there is always a question concerning their generality and applicability to other systems. We proposed a more generalized linear response (GLR) method[28] for free energy calculations, which does not include empirically derived parameters such as α, β, and γ used in eqs 1 and 2 and, consequently, has an inherently greater potential to be applied to diverse molecular systems. Previously, we have successfully applied this method to calculate hydration free energies for a set of organic compounds.[28] In this report, we extend the GLR method to include relative binding free energy calculations, using HIV-1 protease complexes with several of its inhibitors as the test system. We show that the calculated relative binding free energies of the inhibitor JG-365 vs two of its analogues agree closely with the experimental binding free energies yet computed in a much more efficient way than using traditional free energy simulation approaches.

Methodology

Linear Response Approximation (LRA). Linear response approximation for treating electrostatic interaction has been considered in many cases including the Marcus theory of electron-transfer reactions.[29] According to LRA, a system may respond linearly to the change of its electrostatic field, yielding a free energy curve with parabolic form as illustrated in Figure 1.[15,30] The reaction coordinate X describes the transformation of the whole system between the initial state A and the final state B, which are defined by their potential functions V_A and V_B, respectively. It is assumed that the free energy functions of the system at states A and B can be approximately described by two parabolic curves, $\Delta g_A(x)$ and $\Delta g_B(x)$, respectively, and that these two parabolic curves have the same force constant, i.e., the whole system responds linearly to the change of electrostatic field with the same dielectric constant along the path. From these postulates, it can be easily inferred that the change of the free energy between states A and B, ΔG_{AB}, can be expressed as[15,30]

$$\Delta G_{AB} = 1/2 \cdot (\langle V_B - V_A\rangle_A - \langle V_A - V_B\rangle_B) \quad (3a)$$

and[31,32]

$$\Delta G_{AB} = \langle V_B - V_A\rangle_{\text{mid}} \quad (3b)$$

Here, $\langle \cdots \rangle_A$ and $\langle \cdots \rangle_B$ denote the ensemble averages at states A and B, respectively, while $<...>_{\text{mid}}$ stands for the ensemble average at the "midpoint" state between states A

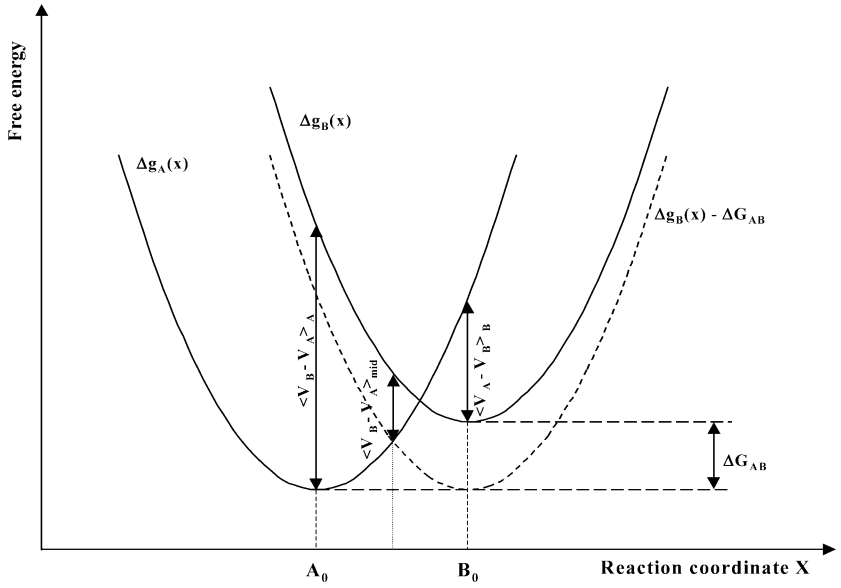

Figure 1. Illustration of the free energy curves for a system that obeys the linear response approximation (LRA). The vertical dotted line indicates the "midpoint" state between the equilibrium coordinates A_0 and B_0 of the initial and final states. The dashed curve $\Delta g_B(x)$-ΔG_{AB} is a copy of the $\Delta g_B(x)$ curve obtained by shifting it downward by ΔG_{AB}. Since $\Delta g_A(x)$ and $\Delta g_B(x)$-ΔG_{AB} have the equal curvature, it is easy to infer that $\langle V_B - V_A \rangle_{mid}$ is equal to ΔG_{AB}. See text for the additional discussion.

and B. The geometric descriptions of all the terms in eqs 3a and 3b are given in Figure 1.

LRA and Free Energy Simulations. The free energy calculations as applied to studying intermolecular interaction are often associated with the physicochemical processes characterized by the transfer of a compound between two phases, e.g., from vacuum to aqueous solution (hydration process) or from aqueous solution to binding site (binding process). These processes can be theoretically perceived and described as the disappearance of a compound from one phase followed by its appearance in the other phase. In a physical sense, it means that both van der Waals and electrostatic interactions between a compound and its environment are eliminated in the initial phase and then re-established in the final phase. Although LRA has been successfully applied to calculate the free energy change resulting from the change of electrostatic interaction, to the best of our knowledge prior to our previous publication[28] there were no reports on the direct (i.e., without introducing any additional empirically derived constants) application of LRA to describe the contribution of van der Waals interaction to the free energy change.

From the mathematical point of view, both the van der Waals and electrostatic terms of a molecular force field have a singularity point at the interaction distance of zero, which may lead to the failure of LRA as illustrated by the following example. Assume that the state A in Figure 1 represents a system consisting of a solute atom surrounded by water molecules, and the state B represents the system after the solute atom has been removed from the aqueous solution. After both van der Waals and electrostatic interactions between the solute atom and the surrounding water molecules disappear, the water molecules have a chance to occupy the former position of the solute atom. If that happens, the value of V_A and therefore $V_A - V_B$ in eq 3a will approach infinity and obviously break the LRA assumption (cf. Figure 1). In fact, this problem has been encountered in traditional free energy simulations, where it is often referred to as the "singularity problem". However, if one only allows the electrostatic interaction between the solute atom and the water solvents disappear, the presence of the remaining van der Waals interaction will prevent the water molecules from approaching the solute atom too closely (or overlapping with it) and thereby avoid the "singularity problem". This is exactly the case in the electron transfer processes and explains why LRA has been successfully applied to these processes.

Generalized Linear Response (GLR) Method. In the previous publication,[28] we suggested that the linear response approximation could also be applied to the change of van der Waals interaction in the course of free energy simulations provided that one could find a proper way to circumvent the "singularity problem". In the case of hydration of small molecules, we have demonstrated that the hydration free energy ΔG can be expressed as[28]

$$\Delta G = \langle V \rangle_{mid} + \Delta G^c \qquad (4)$$

Here, V is the interaction energy between the hydrated solute molecule and water molecules; $\langle \cdot \cdot \cdot \rangle_{mid}$ stands for the ensemble average at the "midpoint" state of the hydration process, which can be approximately defined by a modified molecular force field in which the interaction between the hydrated molecule and the surrounding water molecules are decreased by a half from the original one. The ΔG^c term is similar to a cavity contribution term, representing the free energy required for introducing n independent point particles into the aqueous solution, where n is the number of atoms in the hydrated molecule. Based on the scaled particle theory,[33] we have shown[28] that this term can be approximated as $1.49 k_B n T$, if the SPC model[34] is used to describe water molecules (k_B is the Boltzmann constant; T is the system

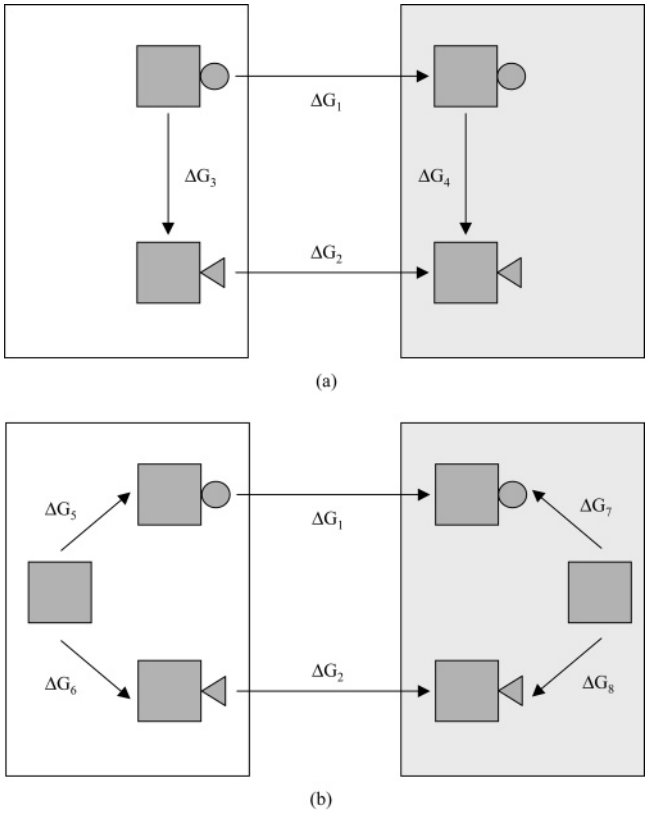

Figure 2. (a) Thermodynamics cycle used in the traditional free energy simulation methods. (b) New thermodynamics cycle used in the GLR method for relative binding free calculations. The large empty box represents the aqueous solution, while the large gray box represents the binding site

temperature). Equation 4 was tested on a data set of 14 chemically diverse organic compounds, and the results were in a good agreement with those obtained from the traditional free energy simulations with an average deviation of only 0.4−0.5 kcal/mol.[28]

GLR and Relative Binding Free Energy Calculations. In drug design, we are often concerned with estimating the relative binding free energy between two different ligands rather than with their absolute binding constants. Free energy simulations in the context of thermodynamic cycle, as shown in Figure 2(a), are frequently used to compute relative binding free energies of two ligands. According to this cycle, the difference between two binding free energies, $\Delta\Delta G$, is defined as $\Delta G_2 - \Delta G_1$ and can be expressed as $\Delta G_4 - \Delta G_3$ (cf. Figure 2(a)) as well. ΔG_3 and ΔG_4 are the free energy changes associated with the transformation of one ligand into another, in solution and in binding site, respectively; these can be relatively easily calculated by free energy simulations if the structural difference between these two ligands is relatively small. To extend eq 4 to relative binding free energy calculations, we introduce two intermediate states into the simple thermodynamics cycle, in which the pseudoligands are composed of the common structural part of the two original ligands. This modified thermodynamic cycle is shown in Figure 2(b). ΔG_5, ΔG_6, ΔG_7, and ΔG_8 represent the free energies required for changing the pseudoligands into the original ligands, respectively. Then, we assume that eq 4 can be used for calculating each of these four values,

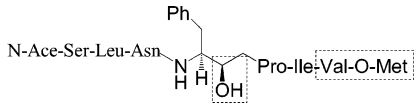

Figure 3. Structure of HIV-PR inhibitor, JG365(S). The dashed boxes indicate the part of the structure subjected to transformation in the course of simulations.

and therefore the relative binding free energy can be expressed as

$$\Delta\Delta G = \Delta G_5 - \Delta G_6 - \Delta G_7 + \Delta G_8$$
$$= (\langle V_5\rangle_{\text{mid}} + \Delta G_5^{\text{c}}) - (\langle V_6\rangle_{\text{mid}} + \Delta G_6^{\text{c}}) - (\langle V_7\rangle_{\text{mid}} + \Delta G_7^{\text{c}}) + (\langle V_8\rangle_{\text{mid}} + \Delta G_8^{\text{c}}) \quad (5)$$

Here, V_5, V_6, V_7, and V_8 are the interaction energies between the variant parts of the ligands and their environments, either the aqueous solution or the binding site. $\langle \cdot\cdot\cdot\rangle_{\text{mid}}$ stands for the ensemble average at the "midpoint" state of each of the four transformation paths. This state can be approximately defined by a modified molecular force field in which the interactions between the variant part of a ligand and its environment are decreased by a half from the original ones. ΔG_5^{c}, ΔG_6^{c}, ΔG_7^{c}, and ΔG_8^{c} are the cavity contribution terms for each transformation step, and we speculate that their net contribution is close to zero due to the following considerations. First, the water density in the binding site should be similar to that in the bulk water, if not identical. Since water density is the major variable to determine the free energy required to introduce a point particle into water,[28,33] this will lead to similar free energy contributions from the cavity terms in the aqueous solution and in the binding site, and therefore they may cancel each other. Second, ΔG_7^{c} and ΔG_8^{c} very probably occur in the same local environment inside the binding pocket. If they are not structurally highly dissimilar to each other, these two terms may also cancel each other. Consequently, eq 5 can be simplified as

$$\Delta\Delta G = \langle V_5\rangle_{\text{mid}} - \langle V_6\rangle_{\text{mid}} - \langle V_7\rangle_{\text{mid}} + \langle V_8\rangle_{\text{mid}} \quad (6)$$

This is the master equation that was used in this work for the relative binding free energy calculations.

HIV-1 Protease and Its Inhibitors as a Test System. HIV-1 protease (HIV-PR) is one of the key enzymes that activate the human immunodeficiency virus (HIV). Many efforts went into the design of specific inhibitors based on the crystallographic structure of this enzyme. One of the well-known specific inhibitors of HIV-PR, JG365, is a non-hydrolyzable heptapeptide analogue of the protease substrate. It incorporates a modified phenylanaline residue named Phs, in which a backbone carbonyl group is substituted with $-CH(OH)CH_2-$ as illustrated in Figure 3. Tropsha and Hermans[35] and Ferguson et al.[36] used molecular simulations independently to calculate the relative binding affinity of (*S*) vs (*R*) isomers of JG365. Reddy et al.[37] used a thermodynamic cycle approach to calculate the effect of deleting the C terminal valine residue of JG365 on the binding affinity of the modified inhibitor. Our GLR method was tested on these two model systems.

In the first case, the configuration of the Phs residue of JG365 is changed from (*S*) to (*R*). Thus, the small circle in

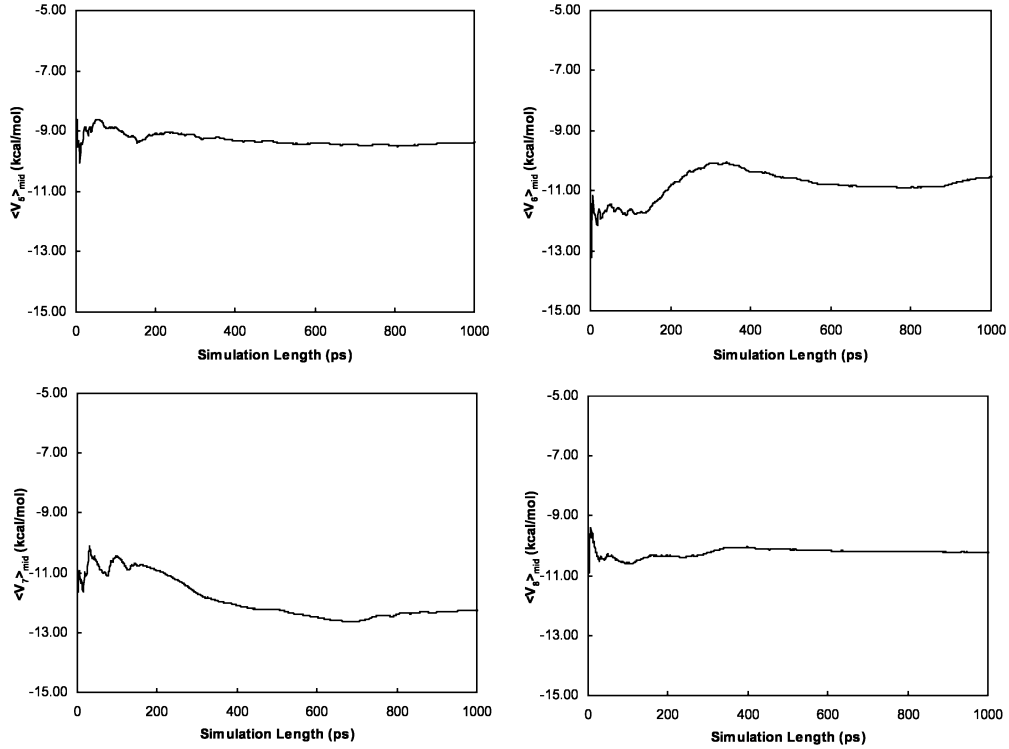

Figure 4. Values of the $\langle V_5 \rangle_{mid}$, $\langle V_6 \rangle_{mid}$, $\langle V_7 \rangle_{mid}$, and $\langle V_8 \rangle_{mid}$ terms in eq 6 as the function of simulation time, for calculating the relative binding free energy between JG365(S) and JG365(R).

Figure 2(b) represents the −CH(OH)− group in the (*S*) configuration, while the small triangle in Figure 2(b) represents the same group in the (*R*) configuration. In the second case, the valine residue of JG365 is removed so the terminal methoxy group is directly linked to the preceding isoleucine residue at the end of transformation. In this case, the small circle in Figure 2(b) represents the valine residue and the terminal methoxy group, while the small triangle in Figure 2(b) represents the terminal methoxy group only. For each case, four independent molecular dynamics simulations were run at the "midpoint" states of each transformation path, in which the interactions between the variant parts of the ligands and their environments, either binding site or aqueous solution, were decreased by one-half. The ensemble average values of the original interaction energies between the variant parts of the ligands and their environments were recorded, and the final $\Delta\Delta G$ value was calculated according to eq 6.

Computational Details

The X-ray structure of HIV-PR in complex with JG365(S) (PDB code 7HVP[38]) was used as the reference structure, from which the structures of HIV-PR and JG365 analogues were modeled. The charge distributions for JG365(S), JG365(R), and their hexapeptide analogues were the same as used in the earlier work.[35] Similar to the procedure reported earlier,[35] JG365 and its analogues were considered in the protonated states, both in the aqueous solution and in the binding site of HIV-PR. For the HIV-PR/ligand complexes, a sufficiently large box of water was defined to guarantee that there was at least 10 Å from each side of the box to the nearest protein atom. Thus, the dimensions of the water box were 77 × 56 × 61 Å, and it contained over 7000 water molecules. The same criterion was used to determine the size of the water box for the ligands alone in aqueous solution. The box dimensions were 44 × 27 × 29 Å for JG365(S) and JG365(R) and 42 × 27 × 29 Å for their hexapeptide analogues.

The molecular dynamics program package Sigma[39] was used for the all simulations. Sigma uses the same description of the nonbonded interactions as GROMOS molecular force field[40] and the bonded interaction parameters that were developed independently.[41] The cutoff radius for the nonbonded interactions was set to 10 Å. The Shake algorithm was used to maintain the bond lengths close to the standard values.[42] The SPC model was used to describe water molecules,[34] and periodic boundary conditions were used. All the simulations were done at 298 K.

For JG365(S), JG365(R), and their hexapeptide analogues in the HIV-PR binding sites, 50 steps of energy minimization followed by 20 ps molecular dynamics simulations were used to equilibrate the whole systems. Then, atoms with an unconstraint motion were limited to those within a 12 Å sphere around the C_α atoms of the Phs residue, its two adjacent residues on the ligand and the critical residues Asp25 and Asp125 on the protein. One nanosecond molecular dynamics simulation was run for calculating each of the four ensemble average terms in eq 6, to guarantee the convergence of each calculated value. For JG365(S), JG365(R), and their hexapeptide analogues in the aqueous solution, the same protocol of simulation was followed, except that all the water molecules were allowed to move.

Results

In the case of the transformation from JG365(S) to JG365(R), the values of the four ensemble average terms in eq 6 as a function of simulation time are shown in Figure 4. With 1

Table 1. Calculated and Experimental Relative Binding Free Energies (kcal/mol)

ligand structures	$\Delta\Delta G$ (calc)	$\Delta\Delta G$ (expt)
JG365(S) vs JG365(R)	3.1	2.6[a]
JG365(S) vs its hexapeptide analog	4.0	
JG365(R) vs its hexapeptide analog	2.0	
mixture of JG365(S) and JG365(R) vs their hexapeptide analogues	3.0[b]	3.8[c]

[a] Reference 43. [b] Average calculated value, assuming the equal-molar mixture of (*S*) and (*R*) configurations of JG365 and their correspondent hexapeptide analogues. [c] Reference 37. Experimental value for the equilibrium mixture of (*S*) and (*R*) configurations of JG365 and their hexapeptide analogues, whose ratio was undetermined.

ns molecular dynamics simulation, the ensemble average values are −9.4 kcal/mol for $\langle V_5\rangle_{\mathrm{mid}}$, −10.5 kcal/mol for $\langle V_6\rangle_{\mathrm{mid}}$, −12.2 kcal/mol for $\langle V_7\rangle_{\mathrm{mid}}$, and −10.2 kcal/mol for $\langle V_8\rangle_{\mathrm{mid}}$, respectively. Thus, according to eq 6, the value of $\Delta\Delta G$ is 3.1 kcal/mol (cf. Table 1), meaning that the (*S*) configuration of JG365 binds to HIV-PR stronger than the (*R*) configuration by about 3.1 kcal/mol. This result is in a reasonable agreement with the experimental data of the difference of the binding energies of these two isomers of JG365, ∼2.6 kcal/mol.[43]

Since the available experimental data for the relative biding affinity between JG365 and its hexapeptide analogue were obtained from the equilibrium mixture of both (*S*) and (*R*) configurations, we did two independent series of simulations to determine the relative binding free energies for both (*S*) and (*R*) configurations. In the case of JG365(S) and its hexapeptide analogue, the values of the four ensemble average terms in eq 6 are plotted as a function of simulation time in Figure 5. Based on 1 ns molecular dynamics simulation, the ensemble average values are −6.0 kcal/mol for $\langle V_5\rangle_{\mathrm{mid}}$, −4.9 kcal/mol for $\langle V_6\rangle_{\mathrm{mid}}$, −10.4 kcal/mol for $\langle V_7\rangle_{\mathrm{mid}}$, and −5.3 kcal/mol for $\langle V_8\rangle_{\mathrm{mid}}$. According to eq 6, we obtain the $\Delta\Delta G$ value of 4.0 kcal/mol for the (*S*) configuration (cf. Table 1). In the case of JG365(R) and its hexapeptide analogue, the values of the four ensemble average terms in eq 6 are shown as a function of the simulation time in Figure 6. Based on 1 ns molecular dynamics simulation, the ensemble average values are −6.8 kcal/mol for $\langle V_5\rangle_{\mathrm{mid}}$, −4.8 kcal/mol for $\langle V_6\rangle_{\mathrm{mid}}$, −9.4 kcal/mol for $\langle V_7\rangle_{\mathrm{mid}}$, and −5.4 kcal/mol for $\langle V_8\rangle_{\mathrm{mid}}$. According to eq 6, we obtain a value of $\Delta\Delta G$ as 2.0 kcal/mol for the (*R*) configuration (cf. Table 1). Due to the fact that the ratio of (*S*) and (*R*) configurations in the mixture has not been experimentally determined, we arbitrarily assume a 50:50 ratio of these two configurations. It leads to an average $\Delta\Delta G$ value of ∼3.0 kcal/mol, indicating that the binding free energy of the racemic mixture of JG365 isomers to HIV-PR is more favorable than that of their hexapeptide analogues by about 3.0 kcal/mol. This result is still in a reasonable agreement with the experimental result of 3.8 kcal/mol.[37] Actually, some evidences indicated that the (*S*) configuration might be the dominant component in the mixture,[37] so if we assume a higher proportion of the (*S*) configuration, we will get a calculated value even closer to the experimental data.

1 ns molecular dynamics simulation was conducted for each of the ensemble average terms in eq 6 to monitor the convergence behavior. We can see from Figures 4−6 that in most cases the ensemble average value is practically converged after 400−600 ps. This gives us the confidence

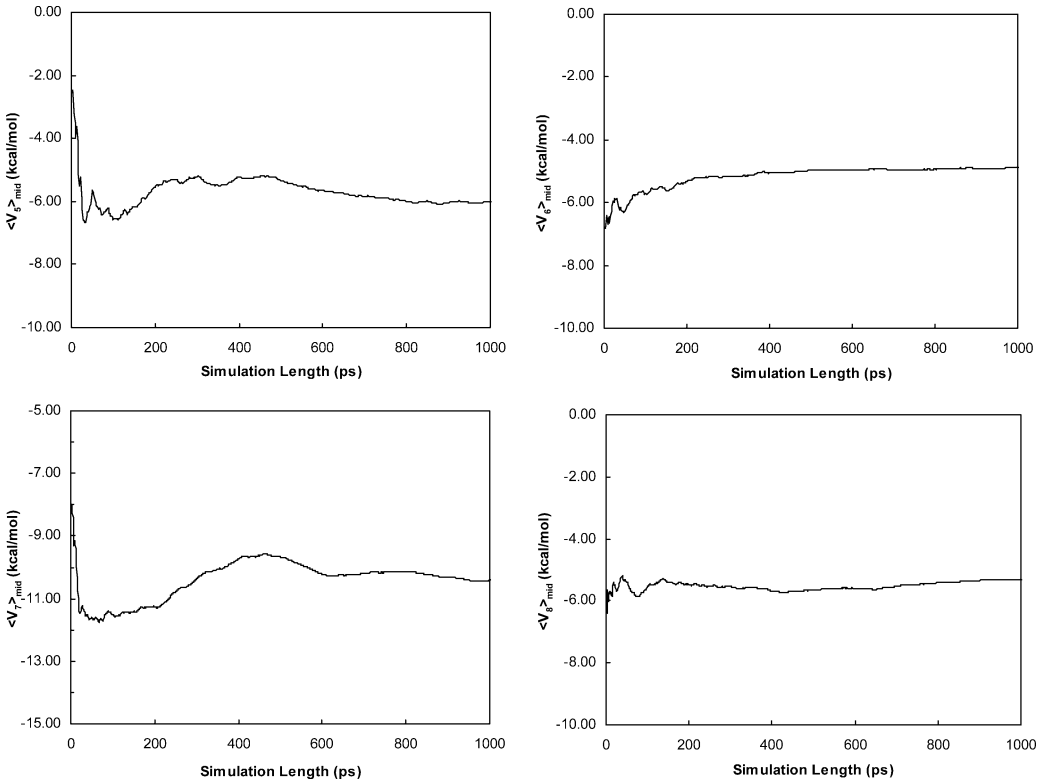

Figure 5. Values of the $\langle V_5\rangle_{\mathrm{mid}}$, $\langle V_6\rangle_{\mathrm{mid}}$, $\langle V_7\rangle_{\mathrm{mid}}$, and $\langle V_8\rangle_{\mathrm{mid}}$ terms in eq 6 as the function of simulation time, for calculating the relative binding free energy between JG365(S) and its hexapeptide analogue.

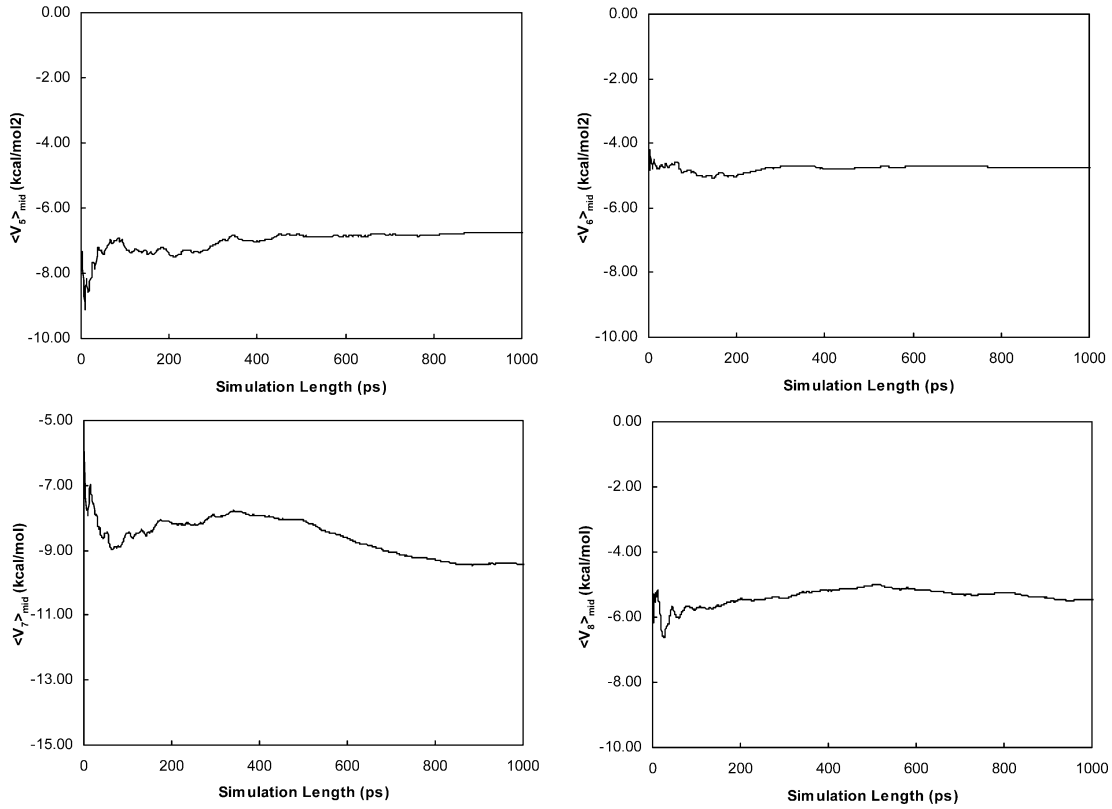

Figure 6. Values of the $\langle V_5 \rangle_{mid}$, $\langle V_6 \rangle_{mid}$, $\langle V_7 \rangle_{mid}$, and $\langle V_8 \rangle_{mid}$ terms in eq 6 as the function of simulation time, for calculating the relative binding free energy between JG365(R) and its hexapeptide analogue.

in the reliability of these calculated values. This is an important observation because the convergence of calculated values is not always guaranteed in the traditional free energy simulations where more than several nanoseconds simulation time is usually required.[44,45]

Discussion

The GLR method is a novel and simple approach for relative binding free energy calculations. It can be regarded as a method somewhere between the traditional qualitative and quantitative modeling methods. It differs from the qualitative methods in that it utilizes molecular dynamics or Monte Carlo simulation technique to sample the complex system, thereby directly taking into account the effects such as flexibility and entropy. On the other hand, it is also different from the quantitative methods, mainly the traditional free energy simulation methods, in that it only requires sampling at a particular "midpoint" state, instead of multiple (usually, dozens) intermediate states along the transformation pathway. Consequently, the GLR method is computationally much more efficient as compared with the traditional free energy simulation methods. Compared with other methods in the same class,[19,46] our GLR method does not require the calibration parameters that are usually used to fit the known experimental data and therefore may be more general.

As reported here, we have tested the GLR method on the HIV-1 protease complex with its inhibitor JG365. Two independent series of calculations were done: one for the relative binding free energy between the (*S*) and (*R*) configurations of JG365 and the other for the relative binding free energy between JG365 and its hexapeptide analogues. These two cases provide us with two representative examples that may well reflect the application scope of the current free energy simulation techniques. In the first case, the structural change is very small, only involving the conversion of a single configuration center that is critical for the binding of HIV-PR inhibitors. In the second one, the structural change is much larger: the entire valine residue disappears during the simulation. Such a structural change probably represents the upper limit of the current capability of free energy simulations. In both cases, our GLR method affords the results in good agreement with the experimental data, with the deviation values of 0.5 and 0.8 kcal/mol, respectively. This accuracy is also comparable with those in the studies conducted by Aqvist et al.[17−22] and Jorgensen et al.;[24−27] however, it should be emphasized that unlike our approach, their studies used several specially derived empirical parameters to fit the experimental data for each compound series.

At the present level, the GLR method requires about several hundred picoseconds of molecular dynamics simulation to guarantee a convergent calculated value, and therefore its application is still limited to the lead optimization rather than lead identification stage. More efficient sampling techniques may further increase the computational efficiency of this method. However, the GLR method does provide a novel and efficient alternative to the traditional free energy simulation techniques for calculating relative binding free energies. We suggest that it can be effectively used in structure-based lead optimization projects.

References

(1) Blaney, J. M.; Dixon, J. S. A Good Ligand Is Hard to Find: Automated Docking Methods. *Perspect. Drug Discovery Des.* **1993**, *1*, 301−319.

(2) Kollman, P. A. Theory of Macromolecule-ligand Interactions. *Curr. Opin. Struct. Biol.* **1994**, *4*, 240−245.

(3) Ajay; Murcko, M. A. Computational Methods to Predict Binding Free Energy in Ligand−receptor Complexes. *J. Med. Chem.* **1995**, *38*, 4953−4967.

(4) 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.

(5) Taylor, J. S.; Burnett, R. M. DARWIN: a Program for Docking Flexible Molecules. *Proteins: Struct., Funct., Genet.* **2000**, *41*, 173−191.

(6) Diller, D. J.; Merz, K. M., Jr. High Throughput Docking for Library Design and Library Prioritization. *Proteins: Struct., Funct., Genet.* **2001**, *43*, 113−124.

(7) Zou, X.; Sun, Y.; Kuntz, I. D. Inclusion of Solvation in Ligand Binding Free Energy Calculations using the Generalized-born Model. *J. Am. Chem. Soc.* **1999**, *121*, 8033−8043.

(8) Majeux, N.; Scarsi, M.; Caflisch, A. Efficient Electrostatic Solvation Model for Protein-fragment Docking. *Proteins: Struct., Funct., Genet.* **2001**, *42*, 256−268.

(9) Karplus, M.; Petsko, G. A. Molecular Dynamics Simulations in Biology. *Nature* **1990**, *347*, 631−639.

(10) Van Gunsteren, W. F.; Berendsen, H. J. C. Molecular Dynamics Computer Simulation. Method, Application and Perspectives in Chemistry. *Angew. Chem.* **1990**, *102*, 1020−1055.

(11) Jorgensen, W. L. Free Energy Calculations: a Breakthrough for Modeling Organic Chemistry in Solution. *Acc. Chem. Res.* **1989**, *22*, 184−189.

(12) Kollman, P. A. Free Energy Calculations: Application to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395−2417.

(13) York, D. M.; Darden, T. A.; Pedersen, L. G. The Effect of Long-Range Electrostatic Interactions in Simulations of Macromolecular Crystals: a Comparison of Ewald and Truncated List Methods. *J. Phys. Chem.* **1993**, *99*, 8345−8348.

(14) Caldwell, J. W.; Kollman, P. A. Cation-pi Interactions - Nonadditive Effects are Critical in Their Accurate Representation. *J. Am. Chem. Soc.* **1995**, *117*, 4177−4178.

(15) Aqvist, J.; Medina, C.; Samuelsson, J.-E. A New Method for Predicting Binding Affinity in Computer-aided Drug Design. *Protein Eng.* **1994**, *7*, 385−391.

(16) Carlson, H. A.; Jorgensen, W. L. An Extended Linear Response Method for Determining Free Energies of Hydration. *J. Phys. Chem.* **1995**, *99*, 10667-.

(17) Hansson, T.; Aqvist, J. Estimation of Binding Free Energies for HIV Protease Inhibitors by Molecular Dynamics Simulations. *Protein Eng.* **1995**, *8*, 1137−1144.

(18) Hulten, J.; Bonham, N. M.; Nillroth, U.; Hansson, T.; Zuccarello, G.; Bouzide, A.; Aqvist, J.; Classon, B.; Danielson, U. H.; Karlen, A.; Kvarnstrom, I.; Samuelsson, B.; Hallberg, A. Cyclic HIV-1 Protease Inhibitors Derived from Mannitol: Synthesis, Inhibitory Potencies, and Computational Predictions of Binding Affinities. *J. Med. Chem.* **1997**, *40*, 885−97.

(19) Hansson, T.; Marelius, J.; Aqvist, J. Ligand Binding Affinity Prediction by Linear Interaction Energy Methods. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 27−35.

(20) Marelius, J.; Graffner-Nordberg, M.; Hansson, T.; Hallberg, A.; Aqvist, J.Computation of Affinity and Selectivity: Binding of 2,4-diaminopteridine and 2,4-diaminoquinazoline Inhibitors to Dihydrofolate Reductases. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 119−131.

(21) Ljungberg, K. B.; Marelius, J.; Musil, D.; Svensson, P.; Norden, B.; Aqvist, J. Computational Modelling of Inhibitor Binding to Human Thrombin. *Eur. J. Pharm. Sci.* **2001**, *12*, 441−446.

(22) Brandsdal, B. O.; Aqvist, J.; Smalas, A. O. Computational Analysis of Binding of P1 Variants to Trypsin. *Protein Sci.* **2001**, *10*, 1584−1595.

(23) McDonald, H.; Carlson, A.; Jorgensen, W. L. Free Energies of Solvation in Chloroform and Water from a Linear Response Approach. *J. Phys. Org. Chem.* **1997**, *10*, 563−576.

(24) Jones-Hertzog, D. K.; Jorgensen, W. L.Binding Affinities for Sulfonamide Inhibitors with Human Thrombin using Monte Carlo Simulations with a Linear Response Method. *J. Med. Chem.* **1997**, *40*, 1539−1549.

(25) Smith, R. H., Jr.; Jorgensen, W. L.; Tirado-Rives, J.; Lamb, M. L.; Janssen, P. A. J.; Michejda, C. J.; Smith, M. B. K. Prediction of Binding Affinities for TIBO Inhibitors of HIV-1 Reverse Transcriptase using Monte Carlo Simulations in a Linear Response Method. *J. Med. Chem.* **1998**, *41*, 5272−5286.

(26) Lamb, M. L.; Tirado-Rives, J.; Jorgensen, W. L. Estimation of the Binding Affinities of FKBP12 Inhibitors using a Linear Response Method. *Bioorg. Med. Chem.* **1999**, *7*, 851−860.

(27) Zhou, R.; Friesner, R. A.; Ghosh, A.; Rizzo, R. C.; Jorgensen, W. L.; Levy, R. M. New Linear Interaction Method for Binding Affinity Calculations using a Continuum Solvent Model. *J. Phys. Chem. B* **2001**, *105*, 10388−10397.

(28) Chen, X.; Tropsha, A. A Generalized Linear Response Method: Application to the Hydration Free Energy Calculations. *J. Comput. Chem.* **1999**, *20*, 749−759.

(29) Marcus, R. A. Chemical and Electrochemical Electron-transfer Theory. *Annu. Rev. Phys. Chem.* **1964**, *15*, 155−196.

(30) Lee, F. S.; Chu, Z.-T.; Bolger, M. B.; Warshel, A. Calculations of Antibody−antigen Interactions: Microscopic and Semi-microscopic Evaluation of the Free Energies of Binding of Phosphorylcholine Analogs to McPC603. *Protein Eng.* **1992**, *5*, 215−228.

(31) King, G.; Barford, R. A. Calculation of Electrostatic Free Energy Differences with a Time-saving Approximate Method. *J. Phys. Chem.* **1993**, *97*, 8798−8802.

(32) Dejaegere, A.; Karplus, M. Analysis of Coupling Schemes in Free energy Simulations: a Unified Description of Nonbonded Contribution to Solvation Free Energies. *J. Phys. Chem.* **1996**, *100*, 11148−11164.

(33) Pierotti, R. A. A Scaled Particle Theory of Aqueous and Nonaqueous Solutions. *Chem. Rev.* **1976**, *76*, 717−726.

(34) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. *Intermolecular Forces*; Reidel: Dordrecht, Holland, 1981.

(35) Tropsha, A.; Hermans, J. Application of Free Energy Simulations to the Binding of a Transition-state-analogue Inhibitor to HIV Protease. *Protein Eng.* **1992**, *5*, 29−33.

(36) Ferguson, D. M.; Radmer, R. J.; Kollman, P. A. Determination of the Relative Binding Free Energies of Peptide Inhibitors to the HIV-1 Protease. *J. Med. Chem.* **1991**, *34*, 2654−2659.

(37) Reddy, M. R.; Viswanadhan, V. N.; Weinstein, J. N. Relative Differences in the Binding Free Energies of Human Immunodeficiency Virus 1 Protease Inhibitors: a Thermodynamic Cycle-perturbation Approach. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 10287−10291.

(38) Jaskolski, M.; Tomasselli, A. G.; Sawyer, T. K.; Staples, D. G.; Heinrikson, R. L.; Schneider, J.; Kent, S. B. H.; Wlodawer, A. Structure at 2.5-.ANG. Resolution of Chemically Synthesized Human Immunodeficiency Virus Type 1 Protease Complexed with a Hydroxyethylene-based Inhibitor. *Biochemistry* **1991**, *30*, 1600−1609.

(39) Hermans, J. *Sigma Manual*; University of North Carolina at Chapel Hill, 1994.

(40) Hermans, J.; Berendsen, H. J. C.; van Gunsteren, W. F.; Postma, J. P. M. A Consistent Empirical Potential for Water-protein Interactions. *Biopolymers* **1984**, *23*, 1513−1518.

(41) Ferro, D. R.; McQueen, J. E.; McCown, J. T.; Hermans, J. Energy Minimizations of Rubredoxin. *J. Mol. Biol.* **1980**, *136*, 1−18.

(42) Van Gunsteren, W. F.; Karplus, M. A Method for Constrained Energy Minimization of Macromolecules. *J. Comput. Chem.* **1980**, *1*, 266−274.

(43) Rich, D. H.; Sun, C. Q.; Vara Prasad, J. V. N.; Pathiasseril, A.; Toth, M. V.; Marshall, G. R.; Clare, M.; Mueller, R. A.; Houseman, K. Effect of Hydroxyl Group Configuration in Hydroxyethylamine Dipeptide Isosteres on HIV Protease Inhibition - Evidence for Multiple Binding Modes. *J. Med. Chem.* **1991**, *34*, 1222−1225.

(44) Pearlman, D. A. A Comparison of Alternative Approaches to Free Energy Calculations. *J. Phy. Chem.* **1994**, *98*, 1487−1493.

(45) Reddy, M. R.; Appelt, K. HIV-1 Protease: Structure-based Drug Design Using the Free Energy Perturbation Approach. *Free Energy Calculations Rational Drug Des.* **2001**, 317−334.

(46) Sham, Y. Y.; Chu, Z. T.; Tao, H.; Warshel, A. Examining Methods for Calculations of Binding Free Energies: LRA, LIE, PDLD-LRA, and PDLD/S-LRA Calculations of Ligands Binding to an HIV Protease. *Proteins: Struct., Funct., Genet.* **2000**, *39*, 393−407.

CT600071Z