Rapid Screening of Binding Affinities: Application of the λ-Dynamics Method to a Trypsin-Inhibitor System

Zhuyan Guo and Charles L. Brooks, III*

Department of Molecular Biology
The Scripps Research Institute
10550 North Torrey Pines Road
La Jolla, California 92037

Received October 1, 1997

The rapid assessment of potential ligands as putative leads in the design of new therapeutics is a growing interest given the recent emergence of combinatorial approaches to compound synthesis. A computational approach that rapidly evaluates the relative binding affinities of a set of ligands to a common protein receptor is a key component in such design. While the methodology for calculations of free energy differences between two molecules is well established,[1−4] the existing methods are computationally intensive and therefore not appropriate for evaluating large numbers of ligands. Methods based on protein−ligand interaction energy may be fast, but they lack accuracy.

Recently, we have developed a free energy based approach to rapidly assess ligand binding affinity.[5,6] This methodology is built on the idea that multiple ligands will compete for a common receptor on the basis of their relative free energies and that one can examine multiple ligands in a common receptor environment using multiple copy simultaneous search approaches.[7] The method is expected to be much more efficient than conventional free energy calculation methods in evaluating multiple ligands, primarily because of the use of the simultaneous "search" component of the technique. Furthermore, since it screens on the basis of the binding free energy of the ligands instead of energy, it provides the potential for accurate assessment of ligand binding affinity. In terms of speed and accuracy, it is a compromise between the conventional free energy calculation methods and energy-based methods.

In this paper, we describe an application of the method to evaluate a set of benzamidine derivatives binding to trypsin. The particular inhibitors studied are benzamidine (*p*−H), *p*-aminobenzamidine (*p*-NH_2), *p*-methylbenzamidine (*p*-CH_3), and *p*-chlorobenzamidine (*p*-Cl) (the structures and charge models used in our calculations can be found in Figure 1s of the Supporting Information. The method gives the correct ranking of binding affinities and requires less than 100 ps of simulation, even though the binding affinity between some of the inhibitors differs by ∼0.5 kcal/mol. We validate our results by comparison to conventional free energy perturbation (FEP) calculations.

The potential function formulated to perform such "competitive binding experiments" is[5,6]

$$V(\{\lambda\},(x)) = \sum_{i=1}^{L} \lambda_i^2 (V_i(x) - F_i) + V_{\text{env}}(x) \qquad (\sum_{i=1}^{L} \lambda_i^2 = 1) \tag{1}$$

where L is the total number of ligands, $V_{\text{env}}(x)$ is the interaction involving the environmental atoms only (e.g., solvent, protein, and the invariant atoms of the ligands), $V_i(x)$ is the interaction involving any of the atoms in the distinct group of molecule i, λ_i is the coupling parameter, and F_i is the reference free energy. There is no interaction among atoms in distinct groups, i.e., ligands are invisible to one another. Each λ_i is treated as a fictitious particle with mass m_i. The dynamics of the system is described by an extended Hamiltonian[5,6,8]

$$H(\{\lambda\},x) = T + T_{\{\lambda\}} + \sum_{i=1} \lambda_i^2 (V_i(x) - F_i) + V_{\text{env}}(x) \tag{2}$$

We note that a straightforward implementation of Monte Carlo methods can also be used here to "evolve" the chemical (λ) variables.[9] The free energy difference between molecules i and j, with reference to free energies F_i and F_j, respectively, can be obtained from

$$\Delta\Delta A_{ij} = -\frac{1}{\beta} \ln \frac{P(\lambda_i = 1, \{\lambda_{m\neq i}\} = 0)}{P(\lambda_j = 1, \{\lambda_{l\neq j}\} = 0)} \tag{3}$$

where $P(\lambda_i = 1, \{\lambda_{m\neq i}\} = 0)$ is the probability that the hybrid system is in a state dominated by ligand i.

Two components pertain to the calculation of the binding free energy of ligands: the free energy of the solvated ligand and the free energy of the complexed ligand−receptor bound state. The solvation free energy can be calculated using conventional free energy methods, or it can be rapidly evaluated by methods based on continuum solvation models such as the Poisson−Boltzmann[10] and generalized Born methods.[11−14] In this work, the free energy of the ligands in solution is taken from previous calculations[6] (column 2 of Table 1). This free energy is taken as the reference free energy F_i in eqs 1 and 2. Because the reference free energy appears in the Hamiltonian of the protein−ligand complexed state, it becomes incorporated into the dynamics simulations. Therefore, the resulting free energy from eq 3 corresponds to the binding free energy of the ligands.

Straightforward λ-dynamics calculations were carried out for 110 ps for the solvated protein−ligand complex, including trypsin and the four noninteracting ligands (see Figure 1s and text of the Supporting Information for details of models and simulation protocol). The values of λ were initialized to the same "position" 1/L, and their velocity was zeroed. These initial conditions place all λ values on an equal footing prior to competition. The

Table 1. Summary of Free Energy Calculations[a,b]

R	ΔG(free)[c]	ΔG(bound)[d]	$\Delta\Delta G$(bind)	$\Delta\Delta G$(solv)
H	0.0	0.0	0.0	0.0
NH_2	−10.6 (−9.5)	−10.2 (−8.9)	0.4 (0.7)	−3.5
CH_3	−4.2 (−3.5)	−1.9 (−1.6)	2.3 (1.9)	−1.0
Cl	−1.1 (−0.8)	1.2 (1.8)	2.2 (2.6)	−8.5

[a] All free energy values referenced to benzamidine and statistical uncertainties are ∼±0.5 kcal/mol for all values reported. [b] Free energy changes from λ-dynamics listed in parentheses. [c] Free energy half-cycle with ligand free in solution. [d] Free energy half-cycle with ligand bound to trypsin.

* Author to whom correspondence should be addressed: http://www.scripps.edu/brooks.

(1) Beveridge, D. L.; DiCapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *92*, 18−431.
(2) Fleischman, S. H.; Brooks, III, C. L., III *Proteins: Struct., Funct., Genet.* **1990**, *7*, 52−61.
(3) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395−2417.
(4) Wong, C. F.; McCammon, J. A. *J. Am. Chem. Soc.* **1986**, *108*, 3830.
(5) Kong, X.; Brooks, C. L., III *J. Chem. Phys.* **1996**, *105*, 2414−2423.
(6) Guo, Z.; Brooks, C. L., III; Kong, X. *J. Phys. Chem.* In press.
(7) Elber, R.; Karplus, M. *J. Am. Chem. Soc.* **1990**, *112*, 9161−9175.
(8) Anderson, H. C. *J. Chem. Phys.* **1980**, *72*, 2384.
(9) Tidor, B. *J. Phys. Chem.* **1993**, *97*, 1069−1073.
(10) Klapper, I.; Hagstrom, R.; Fine, R.; Sharp, K.; Honig, B. *Proteins: Struct. Funct., Genet.* **1986**, *1*, 47−59.
(11) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(12) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122−129.
(13) Schaefer, M.; Karplus, M. *J. Phys. Chem.* **1996**, *100*, 1578−1599.
(14) Edinger, S. R.; Cortis, C.; Shenkin, P. S.; Friesner, R. A. *J. Phys. Chem. B* **1997**, *101*, 1190−1197.

S0002-7863(97)03418-5 CCC: $15.00
Published on Web 02/14/1998

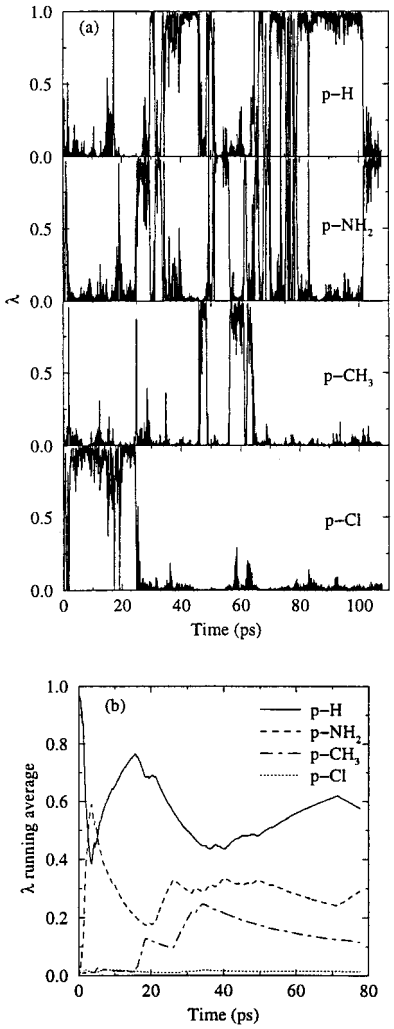

Figure 1. (a) Dynamics trajectory of λ values for (a) *p*-H, (b) *p*-NH_2, (c) *p*-CH_3, and (d) *p*-Cl. The larger the λ value, the stronger the ligand interaction with the protein at that instant. (b) Running average of each λ as a function of simulation time. An initial 30 ps equilibration period of the λ-trajectory was not included in the running averages.

resulting λ trajectories are shown in Figure 1a. Since the overall binding affinity among the species differs by only about 2 kcal/mol (see Table 1), they all compete reasonably well for binding, as indicated by the nonzero population of each species in the $\lambda \approx 1$ state. Figure 1a shows that during the first 30 ps, while the system is still equilibrating, the *p*-Cl derivative clearly dominates, but afterward its contribution dies out completely. The *p*-CH_3 derivative competes with $\lambda \approx 1$ for only a small fraction of time. Benzamidine and the *p*-NH_2 derivative possess the highest binding affinities, as evident by the large population of both species in the $\lambda \approx 1$ state.

To provide a qualitative measure of the binding affinity of the ligands, we constructed the running average of each λ (Figure 1b). From the λ-trajectories and the running averages shown in Figure 1, the relative ranking in binding affinity, *p*-H > *p*-NH_2 > *p*-CH_3 > *p*-Cl, clearly emerges. Since we are interested in ligands that bind favorably, the ranking of *p*-CH_3 and *p*-Cl is less significant. For a more rigorous measure of binding affinity, one can evaluate the population (or fraction of time) of a ligand in the $\lambda = 1$ state. Higher population reflects more favorable binding free energy. However, because each ligand populates the regions of λ-space with λ near 1 or 0, the running average shown in Figure 1b also well reflects the relative binding affinities. To check that the ligands were not trapped in local minima, we performed additional simulations with different initial λ "coordinates" and velocities, and we obtained the same ranking.

To evaluate the performance of the λ-dynamics method, we performed standard FEP calculations on the same system using the same force field. In these calculations, 90 ps of sampling at $\lambda = 0.125$, 0.5, and 0.875 was performed to map the reactant state to the product state using double-wide sampling. Experimental results show that the binding affinities of the ligands under study are quite close.[15] The maximum difference is only about 1 kcal/mol. Although our model parametrization is not intended to reproduce experimental values, these measurements can still give us some insight into the range of the results we expect. In the relatively long FEP calculations, three transformations—*p*-Cl to *p*-H, *p*-NH_2 to *p*-Cl, and *p*-CH_3 to *p*-NH_2—were considered, each involved a 270 ps dynamics simulation. The binding free energies of the ligands from these calculations are listed in Table 1. The binding affinity ranking is *p*-H > *p*-NH_2 > (*p*-CH_3,*p*-Cl). Benzamidine and the *p*-NH_2 derivative differ by only 0.4 kcal/mol, while the *p*-CH_3 derivative and the *p*-Cl derivative are equal within the statistical uncertainties inherent in these calculations. Results from both methods are in good agreement. Examination of the various free energy terms in Table 1 reveals that both desolvation and protein−ligand interactions contribute to the binding free energies.

We also list in the table the relative binding free energy of the ligands from a 260 ps λ-dynamics simulation of the hybrid multiligand system. In these calculations, a set of biasing potentials, each corresponding to the estimated free energy of a ligand, was used as the reference free energy in eqs 1 and 2.[6] The results are consistent with those from FEP calculations but the computation time was significantly reduced. The same ordering as in the screening calculations was obtained, which further validates the λ-dynamics approach for screening calculations.

In summary, the λ-dynamics methodology, designed to rapidly identify ligands with favorable binding free energy, has been applied to a protein system. The method clearly identified and correctly ordered the two species possessing the strongest binding affinities, benzamidine and its *p*-NH_2 derivative, although the binding free energy of the two differ by only 0.4 kcal/mol. Because the binding affinities of the ligands studied are fairly close, competition for binding was quite high and consequently required relatively long simulations to resolve the binding free energies (~100 ps). In general, species whose binding affinities differ by more than 3 kcal/mol from the most favorable binder can be easily screened out within tens of picoseconds of simulation because they cannot compete, i.e., they never reach the $\lambda \approx 1$ state. Longer simulations provide the correct ranking of the favorable binders. Although only four ligands were examined in this study, the simulation time should in general be nearly independent of the total number of ligands because only the favorable binders are able to compete for the $\lambda = 1$ state. In contrast, the simulation time to perform FEP calculations on a pair of ligands is on the order of hundreds of picoseconds and increases with the number of ligands. If more quantitative assessment of binding affinities is desired after initial screening, an iterative procedure can be employed with the same λ-dynamics formalism.[6] The λ-dynamics method has great potential for the study of ligand binding and molecular design.

Acknowledgment. Funding of this work from the National Institutes of Health (NIH GM37554) and the National Science Foundation (NSF ASC9503998) is highly appreciated. Z.G. thanks Drs. J. Hirst and Z. Zhou for helpful comments during manuscript preparation.

Supporting Information Available: Listing of the molecular structures and simulation protocol used in the present calculations (2 pages). See any current masthead page for ordering information and Web access instructions.

JA973418E

(15) Mares-Guia, M.; Nelson, D. L.; Rogana, E. *J. Am. Chem. Soc.* **1977**, *99*, 2331−2336.