Calculation of protein–ligand binding free energy by using a polarizable potential

Dian Jiao[†], Pavel A. Golubkov[†], Thomas A. Darden[‡], and Pengyu Ren[†§]

[†]Department of Biomedical Engineering, University of Texas, Austin, TX 78712; and [‡]National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709

Edited by Arieh Warshel, University of Southern California, Los Angeles, CA, and accepted by the Editorial Board February 20, 2008 (received for review December 13, 2007)

The binding of charged ligands benzamidine and diazamidine to trypsin was investigated by using a polarizable potential energy function and explicit-water molecular dynamics simulations. The binding free energies were computed from the difference between the free energies of decoupling the ligand from water and protein environments. Both the absolute and the relative free energies from the perturbation simulations agree with experimental measurements to within 0.5 kcal·mol^{-1}. Comparison of free-energy components sampled from different thermodynamic paths indicates that electrostatics is the main driving force behind benzamidine recognition of trypsin. The contribution of electronic polarization to binding appears to be crucial. By computing the free-energy contribution caused by the polarization between the ligand and its surroundings, we found that polarization has the opposite effect in dissimilar environments. Although polarization favors ligand solvation in water, it weakens the protein–ligand attraction by screening the electrostatic interaction between trypsin and benzamidine. We also examined the relative binding free energies of a benzamidine analog diazamidine to trypsin. The changes in free energy on benzamidine-diazamidine substitution were tens of kilocalories in both water and trypsin environments; however, the change in the total binding free energy is <2 kcal·mol^{-1} because of cancellation, consistent with the experimental results. Overall, our results suggest that the use of a polarizable force field, given adequate sampling, is capable of achieving chemical accuracy in molecular simulations of protein–ligand recognition.

simulation | molecular dynamics | trypsin | benzamidine | force field

Specific recognition of ligands by proteins is at the core of many crucial biological functions and systems such as enzyme catalysis and intracellular signaling. Binding affinity characterizes the strength of such recognition. With the recent advancements in computing, prediction of the binding affinity based on physical principles of molecular interaction has come to the forefront of active research and has been the subject of regular reviews (1–5). All-atom molecular dynamics (MD) simulation with explicit solvent, coupled with efficient free-energy sampling algorithms, can potentially offer accurate prediction of binding free energies of ligands to proteins (5). Common free-energy simulation algorithms include the double-decoupling method (DDM) and potential of mean force approach (PMF). Free-energy perturbation (FEP), thermodynamic integration (TI), or umbrella sampling can be used to compute free-energy differences in either DDM or PMF. It has been argued that DDM is problematic for charged systems, because the binding free energy is computed as a small difference between two large solvation energies in water and in protein (6). However, the PMF approach does not quantify absolute solvation energies of ligand, which makes it difficult to detect potential problems in treatment of long-range effect and boundary conditions (7, 8). A comparison of PMF and FEP in ion channel study indicated the former suffered more seriously from hysteresis (9). Alternatives to free-energy pathway calculations include linear response analysis (LRA) (10) and linear interaction energy (LIE) (11), where only the ligand-bound and unbound states are simulated. A semimacroscopic approach based on Protein Dipoles Langevin Dipole (PDLD/S) was applied previously in LRA to further reduce the computational cost (12). Recent reviews have summarized some of the advantages and drawbacks of LRA and LIE (5, 13).

MD/FEP methods have been used to calculate the absolute binding free energies of different protein–ligand systems, such as L99A mutant of T4 lysozyme with benzene (14–16), tyrosyl-tRNA-synthetase with tyrosine (14), FKBP with several ligands (17, 18), and human Lck SH2 domain with phosphotyrosine peptide (6), to name a few. Strong correlation between computed binding free energies and experimental values has been reported for a series of ligands binding to FKBP and lysozyme. Nonetheless, the calculated absolute binding free energies can still deviate from experimental measurement by several kilocalories. There has been a limited number of simulation studies of highly charged systems. Recently, the PMF approach was used successfully in calculating the binding free energies of a charged peptide binding to the SH2 domain (6).

It has been recognized that the bottlenecks to achieving chemical accuracy in molecular simulations are the underlying physical models and the sampling convergence (5). The current-generation common force fields employ fixed atomic charges and therefore lack the ability to respond to the actual local electrostatic environment. Explicit treatment of polarization to provide realistic electrostatic representation dates back to Warshel and Levitt's use of atomic induced dipoles in the enzyme reaction study (19). Polarized force field (PFF) was later applied to estimating binding free energies in systems such as trypsin, antibody–antigen, and DNA polymerase (10, 20, 21). History and development of PFF have been covered in recent reviews (13, 22). Aside from the physical potential, sampling convergence remains an enormous challenge in binding simulations with atomic force fields, especially when a large number of conformational and other degrees of freedom are involved (23).

In this study, we report the calculation of the absolute and relative binding free energies of trypsin with two charged ligands from molecular dynamics simulations with a polarizable force field, in which the electrostatics was represented by atomic multipoles and induced dipoles. Trypsin is a serine protease that has been well studied, both experimentally and computationally. Benzamidine and its derivatives are well characterized inhibitors of trypsin that form a salt bridge with the D189 aspartic acid in the S1 site. Benzamidines carry net charges and are relatively small and rigid. This allows us to achieve adequate sampling and focus on the application of the polarizable potential in the calculation of binding free energies. Besides comparing calculated free energy with ex-

Author contributions: P.R. designed research; D.J., T.A.D., and P.R. performed research; D.J., P.A.G., and P.R. analyzed data; and D.J., P.A.G., T.A.D., and P.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. A.W. is a guest editor invited by the Editorial Board.

[§]To whom correspondence should be addressed. E-mail: pren@mail.utexas.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0711686105/DCSupplemental.

© 2008 by The National Academy of Sciences of the USA

Downloaded from https://www.pnas.org by 38.88.236.210 on October 11, 2024 from IP address 38.88.236.210.

BeiGene Exhibit 1127.001
BeiGene v. Pharmacyclics
PGR2024-00009

Downloaded from https://www.pnas.org by 38.88.236.210 on October 11, 2024 from IP address 38.88.236.210.

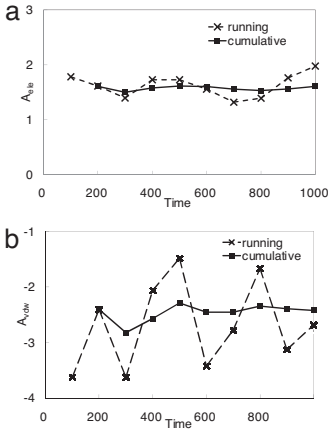

Fig. 1. Electrostatic (*a*) and van der Waals (*b*) decoupling free energies (kcal·mol^{-1}) of ligand–water system. The dashed line with cross markers is the running average of every 100-ps block. The solid line with square markers is the cumulative average. Time is in picoseconds.

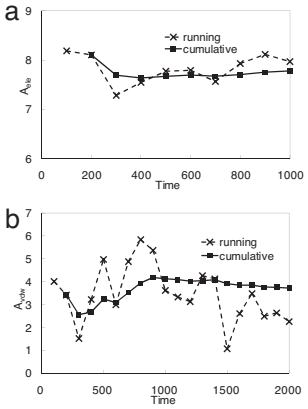

Fig. 2. Electrostatic (*a*) and van der Waals (*b*) decoupling free energies of ligand–protein system (kcal·mol^{-1}). The dashed lines with cross markers are the running average of every 100-ps block. The solid lines with square markers are the cumulative average. Time is in picoseconds.

perimental literature, we also examined the role of electrostatics and polarization in ligand–protein binding.

Results and Discussion

Absolute Binding Free Energy. To evaluate the absolute binding free energy of benzamidine to trypsin, the free energies of decoupling benzamidine from water and trypsin-water were computed from MD simulations, respectively. The decoupling free energies were evaluated from a path in which the electrostatic and then the van der Waals (vdW) interactions between benzamidine and its environment were turned off in steps. A soft-core version of buffered-14-7 potential with $n = 5/\alpha = 0.7$ (see Eq. **2**) was used in the vdW decoupling. A harmonic potential ($k = 20$ kcal·mol^{-1}·Å^{-2}) was used to restrain the benzamidine to the trypsin.

Figs. 1 and 2 show the decoupling free energies of the ligand–water system and the ligand–protein system, respectively. The first 100 ps of MD trajectories in all simulations were considered as system equilibration and subsequently ignored in the free-energy analysis. In addition to calculating cumulative averages, running averages were computed from 100-ps blocks of trajectories to illustrate the fluctuation. Note that the running averages do not reflect the statistical error in the final free energy. The free energies of the ligand–water system are reasonably converged in 1 ns (Fig. 1), whereas the ligand–protein free energy took longer to settle down, even with a restraining bond between the ligand and protein (Fig. 2*b*). This is expected because trypsin–water is much more complex than bulk water and thus requires longer simulations. For both ligand–water and ligand–protein, the electrostatic free energy fluctuates much less than the vdW component. This is because full van der Waals interactions are present during electrostatic decoupling, which confines benzamidine to the pocket with low mobility. In contrast, as the vdW interactions between benzamidine and its environment were gradually decoupled, the benzamidine molecule sampled greater and greater regions of space, coming in close contact with and eventually penetrating the surrounding water and protein atoms. Additionally, water molecules were found to occupy the pocket after ≈500 ps as benzamidine was annihilated from the S1 pocket of trypsin. All of these factors may contribute to the wild fluctuations in the vdW free energy.

The electrostatic and the van der Waals decoupling free energies were determined to be 1.27 ± 0.2 kcal·mol^{-1} and −2.42 ± 0.4 kcal·mol^{-1}, respectively, for benzamidine–water, and 7.78 ± 0.2 kcal·mol^{-1} and 3.72 ± 0.3 kcal·mol^{-1} for benzamidine–trypsin (Table 1). Thus, the free energy of benzamidine binding to trypsin is −6.78 kcal·mol^{-1} (see Eq. **4**), which includes a correction of 5.87 kcal·mol^{-1} for the bias due to the restraint. Several experimental binding free energies were reported, ranging from 6.3 to 7.3 kcal·mol^{-1} (24–27, 53). It is not uncommon for the experimental binding affinity to vary up to a factor of 10 (1.3 kcal·mol^{-1}), depending on the assay conditions (27), or the specific experimental method such as spectrophotometry (27) or crystallography (53).

The free energy of decoupling benzamidine from water appears to be much lower than the expected solvation free energy of a charged molecule, which is typically several tens of kilocalories per mole. As noted in *Methods*, we decoupled the electrostatic interaction by zeroing out the atomic multipoles and polarizabilities of the benzamidine; and the end states of the ligand–water and ligand–protein simulations feature a "ghost" benzamidine molecule without intramolecular electrostatic interactions. We have determined that the recharging free energy of benzamidine in vacuum is 46.92 kcal·mol^{-1}. Combining this value with the decoupling energy above, the electrostatic solvation free energy of benzamidine in water becomes −48.19 kcal·mol^{-1} (Table 2). Thus, a large portion of solvation free energy of charged benzamidine (both in water and trypsin environments) is actually not responsible for driving its binding to trypsin.

Effect of Soft-Core vdW Potential. Free energy is a state function, and therefore independent of the sampling path. We investigated two soft-core modifications of the buffered-14-7 function (see Eq. **2**), $n = 5$ and $\alpha = 0.7$, in the calculation of vdW decoupling free energy in ligand–water and ligand–protein. As shown in Fig. 3, the free energies computed by using the two potentials converge toward each other after ≈1 ns of simulation. By using $n = 4/\alpha = 0.5$, the van der Waals decoupling free energies were found to be −2.27 ± 0.4 and 3.42 ± 0.3 kcal·mol^{-1} for ligand–protein and ligand–water, respectively, compared with −2.42 ± 0.4 and 3.72 ± 0.3 kcal·mol^{-1} from simulations with $n = 5/\alpha = 0.7$ (Table 1). The differences between the two sets of values are comparable to the statistical error.

Free Energy as Driving Force for Binding. It is possible to decode the physical driving force behind benzamidine binding to trypsin because our calculations offer detailed information on atomic interactions that are not easily measurable by experimental procedures.

BIOPHYSICS

BeiGene Exhibit 1127.002
BeiGene v. Pharmacyclics
PGR2024-00009

Table 1. Absolute free energy of benzamidine binding to trypsin computed by using different force constants and soft-core coefficients

Restraint constant, kcal·mol^{-1}·Å^{-2}	Soft core	ΔA_{wat} ($L \rightarrow 0$)*		ΔA_{pro} ($L \rightarrow 0$)*		Restraint correction	ΔA_{calc}	ΔA_{exp}
		ΔA_{ele}	ΔA_{vdw}	ΔA_{ele}	ΔA_{vdw}			
20	0.5/4	1.27 (0.2)	−2.27(0.4)	7.78 (0.2)	3.42 (0.3)	5.87	−6.33	−6.3[†] −7.3[‡] −6.4[§] −6.7[¶]
20	0.7/5	1.27[‖]	−2.42(0.4)	7.78[‖]	3.72 (0.3)	5.87	−6.78	–
40	0.7/5	1.27[‖]	−2.35**	7.57 (0.2)	4.56 (0.2)	6.48	−6.73	

*The numbers in parentheses are the standard errors.
[†]Ref. 27.
[‡]Ref. 25.
[§]Ref. 53.
[¶]Ref. 26.
[‖]The value is taken from the row above.
**Averaged from the two ligand–water vdW decoupling free energies in the rows above.

However, this is complicated because of the presence of the restraint between benzamidine and trypsin. We overcame this complication by comparing simulation results from different restraints.

Besides using a force constant of 20 kcal·mol^{-1}·Å^{-2} for the restraint, we performed another set of simulations of trypsin–benzamidine with doubled restraint strength (k = 40 kcal·mol^{-1}·Å^{-2}). The electrostatic and the vdW decoupling free energies from the two sets of simulations were compared in Fig. 4. The electrostatic decoupling free energy is 7.57 ± 0.2 kcal·mol^{-1} from the simulation with k = 40, compared with 7.78 ± 0.2 kcal·mol^{-1} when using the weaker restraint (Table 1). This similarity indicates that the restraint strength has little effect on the electrostatic decoupling. This is not surprising because benzamidine is likely to be confined within the trypsin-binding pocket by vdW interactions at this stage. The restraint does not truly come into effect until the vdW interactions are turned off. Indeed, the van der Waals free energies differ by 0.8 kcal·mol^{-1} depending on the restraint strength. However, the correction to the binding free energy caused by the restraint also varies as the restraint strength increases from 20 to 40 by a similar amount (0.6 kcal·mol^{-1}). Based on the preceding observation, we argue that the restraint affects mostly the vdW decoupling such that the correction should be applied mainly to the vdW decoupling free energy of benzamidine–trypsin. After taking these corrections into account, the van der Waals free energies of decoupling benzamidine from trypsin become −2.3 kcal·mol^{-1} (averaged over values from two soft-core vdW potential at k = 20) and −1.9 kcal·mol^{-1} (k = 40). These quantities are fairly close to the vdW decoupling free energy of ligand–water (−2.27 and −2.42 kcal·mol^{-1}, depending on the soft-core potential). In contrast, the electrostatic decoupling free energy of ligand–water and ligand–protein differ by −6.4 kcal·mol^{-1} on average, which amounts to almost all of the binding free energy. Thus, we conclude that the electrostatic interaction is responsible for the binding of benzamidine to trypsin.

Table 2. Absolute solvation free energies of benzamidine and diazamidine in water and trypsin

	Protein		Water		Binding		
	ele*	vdw[†]	ele*	vdw	ele	vdw	Total
Benzamidine	−54.60[‡]	2.17[‡]	−48.19	2.35[‡]	−6.41	−0.18	−6.59
Diazamidine[§]	−75.56	−2.17	−69.38	−2.97	−6.18	0.80	−5.38

*Intramolecular contribution of −46.92 kcal·mol^{-1} is included.
[†]Restraint correction is included in the vdW component.
[‡]Averaged from values in Table 1.
[§]Computed based on the relative binding free energy.

As shown in Fig. 4*b*, the vdW decoupling free energy with stronger restraint (k = 40) seems to stabilize much quicker than that with k = 20. Nonetheless, it still drifts slightly over the 2-ns simulation timeframe. We extended the simulation to 3 ns, during which the free-energy value changed by −0.31 kcal·mol^{-1}. Therefore, the stronger force constant may offer a quicker estimation of the binding free energy, although long simulations (≈3 ns) are still necessary to obtain accurate results.

In Table 1, we summarized the three sets of the binding free energies computed by using different soft-core vdW and restraint strengths. Consistency among the simulation results supports the premise that the sampling is adequate and the results are well converged, owing to the presence of the restraint and to the fact that benzamidine is small and rigid. Our best estimate of the absolute binding free energy of benzamidine–trypsin is therefore 6.6 kcal·mol^{-1}, averaged over all simulation results. The agreement between the calculated and experimental values is well within chemical accuracy.

Polarization Effect. A unique feature of the present model is the explicit treatment of dipole polarization, which allows the electrostatics to respond to the environment, be it water or protein. It was

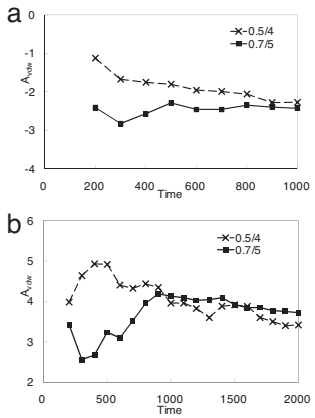

Fig. 3. Comparison of vdW decoupling free energies (kcal·mol^{-1}) of ligand–water (*a*) and ligand–protein (*b*) systems with different soft-core pathways. All lines are cumulative averages. The dashed lines with cross markers are from $n = 4/\alpha = 0.5$ simulations. The solid lines with square markers are from $n = 5/\alpha = 0.7$ simulations (the same data as in Fig. 2*b*). Time is in picoseconds.

Downloaded from https://www.pnas.org by 38.88.236.210 on October 11, 2024 from IP address 38.88.236.210.

BeiGene Exhibit 1127.003
BeiGene v. Pharmacyclics
PGR2024-00009

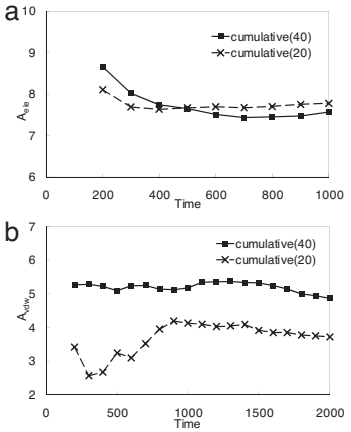

Fig. 4. Comparison of electrostatic (*a*) and van der Waals (*b*) decoupling free energies (kcal·mol^{-1}) of benzamidine–trypsin system by using different force constants for the restraint that confines the benzamidine to trypsin. The dashed lines with cross markers are the cumulative averages from a force constant $k = 20$ kcal·mol^{-1}·Å^{-2}, whereas the solid lines with square markers are from a force constant $k = 40$ kcal·mol^{-1}·Å^{-2}. Time is in picoseconds.

suggested that accounting for polarization improves the transferability of a force field (28), which would be critical for transferring ligand from bulk water into the protein. It is therefore of interest to examine the effect of polarization on the thermodynamics of benzamidine binding to trypsin. We turned off the dipole induction between benzamidine and its environment by using free-energy perturbation to compute the "polarization free-energy" in both bulk water and trypsin. Note that there is still polarization present between water–water and trypsin–water, although both water and trypsin are unable to feel the electric field due to benzamidine and vice versa. The free-energy change caused by the removal of the polarization between benzamidine and water is 4.49 kcal·mol^{-1}, and -22.37 kcal·mol^{-1} between benzamidine and trypsin. Not only are the magnitudes dramatically different, the sign is also opposite. Although polarization seems to enhance the solvation of benzamidine in water, it weakens the association between benzamidine and trypsin. Overall, the polarization works to diminish the effect of permanent electrostatics in driving the binding of benzamidine to trypsin.

Turning on polarization leading to an increase in the energy of the protein–ligand complex may appear counterintuitive. Indeed, polarization energy always lowers the "total" system energy by making a negative contribution. However, the polarization energy of the complex became less negative when the polarization between benzamidine and trypsin–water was present. In other words, the gain in electrostatic energy because of permanent electrostatic interactions, e.g., salt bridges, is counterbalanced by the loss in the polarization energy. The local polarization response to the association of two charged entities is to screen the electrostatic interactions, similar to the dielectric effect of water screening charge interactions. To verify this phenomenon, we computed the total dipole moment of the carboxyl group (CO_2^-) of trypsin's aspartic acid D189, which forms a salt bridge with benzamidine, before and after polarization was turned on. Consistent with our observation of energy, the dipole moment did decrease by 0.1 D when polarization was present.

In our model, polarization energy between benzamidine–water and benzamidine–trypsin differs by 27 kcal·mol^{-1}. Our results agree with earlier findings that electrostatics is sensitive to local environment. Calculations by Lee *et al.* (10) showed that the polarization energy in water and antibody–antigen complex varied by as much as 8 kcal·mol^{-1}. A quantum mechanics/molecular mechanics (QM/MM) study of HIV-1 protease-inhibitor binding (29) suggested that polarization contributed to approximately one-third of the total electrostatic interaction energy.

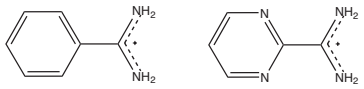

Fig. 5. Chemical structure of two charged ligands of trypsin: benzamidine (*Left*) and diazamidine (*Right*).

However, the artificial model we created by turning off polarization is not equivalent to fixed-charge-based force fields. Previous LIE studies of trypsin–benzamidine using fixed-charge potentials reported electrostatic contribution to the binding free energy to be between -4.9 and -6.4 kcal·mol^{-1} (30–32). The values are close to -6.5 kcal·mol^{-1} obtained in this study, although the electrostatic solvation energies of benzamidine in water differ from our values by -4 to -17 kcal·mol^{-1} (Table 2). This suggests that it is possible for fixed-charge models to implicitly include the overall polarization effect in the binding equilibrium. Direct comparisons of polarizable and nonpolarizable force fields in the free-energy pathway calculations will perhaps offer further insight. It has also been discussed previously that unless divalent ions are involved (13) or binding occurs at the protein interior (19), polarization plays a secondary role and may be absorbed by effective parameterization of fixed charges.

Relative Binding Free Energy. In addition to simulating benzamidine, we also calculated the relative binding free energy of another ligand, diazamidine (Fig. 5). Diazamidine is a benzamidine derivative with the phenyl ring replaced by a 1,3-diazine (pyrimidine). In benzamidine, there is essentially one internal rotatable bond linking the aromatic ring and the amidine group (C_a–C_i). In Fig. 6, the gas-phase QM (MP2/6–311++G(2d,2p)) conformational energy profiles about this bond have been plotted for both benzamidine and diazamidine. The minimum energy occurs at 45° for benzamidine and at 0° for diazamidine. Most interestingly, the rotational barrier increases by a factor of 4 from benzamidine to diazamidine, implying a double-bond nature of the C_a–C_i bond in the latter. Although the conformational flexibilities of the two ligands are dramatically different, we have adopted a generic torsional energy parameter, $E_{tor} = 2.7*(1 - \cos 2\varphi)$ kcal·mol^{-1}, for both molecules. The same parameter is also used for the C_ε–C_ζ–O–H torsion in AMOEBA tyrosine. As shown in Fig. 6, the conformational energies computed by AMOEBA are consistent with the QM results. The locations of minima and maxima are well reproduced. The rotational energy barrier of diazamidine is replicated almost exactly, whereas the barrier for benzamidine is lower in our model.

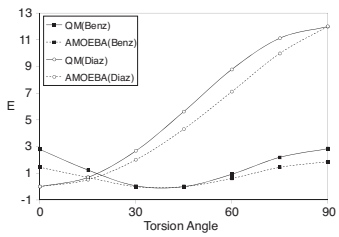

Fig. 6. AMOEBA and QM potential energy (kcal·mol^{-1}) with respect to the rotation about the C_a–C_i bond (degrees) for the two ligands of trypsin: benzamidine (Benz) and diazamidine (Diaz). The C_a–C_i bond connects the aromatic ring and the amidine group.

BIOPHYSICS

Downloaded from https://www.pnas.org by 38.88.236.210 on October 11, 2024 from IP address 38.88.236.210.

BeiGene Exhibit 1127.004
BeiGene v. Pharmacyclics
PGR2024-00009

In AMOEBA energy profile, the torsion potential contributes to <30% of the rotational barrier in diazamidine. The majority of the conformational energy arises from electrostatic and vdW interactions. This illustrates that transferability of torsion potentials requires a physical treatment of intramolecular nonbonded interactions.

We calculated the binding free energy of diazamidine by perturbation from benzamidine. Two sets of MD simulations were performed in ligand–water and ligand–protein systems where benzamidine was mutated into diazamidine. No restraint was used between trypsin and ligands. The perturbation was done by changing the parameters involved in the C→N mutation in several steps (electrostatic, vdW, two bonds, and one angle). A new set of atomic multipoles was derived for diazamidine from QM. The hydrogen atoms bonded to the mutated phenyl carbon atoms were turned into dummies (no nonbonded interactions with all other atoms). The free-energy change on mutating benzamidine into diazamidine in bulk water was -26.51 ± 0.5 kcal·mol^{-1}, whereas the free-energy change in trypsin was computed to be -25.30 ± 0.3 kcal·mol^{-1}. Approximately 80% of the free-energy changes are due to electrostatics. We repeated the calculation where polarization between ligand and its environment was turned off. We observed a similar difference in the diazamidine polarization free energies in water (5.56) and trypsin (-19.80), compared with 4.49 and -22.37 kcal·mol^{-1} obtained for benzamidine. In addition, the molecular polarizability of benzamidine, 13.8 Å^3, is only slightly greater than that of diazamidine, 12.6 Å^3. These results imply that it is primarily the permanent electrostatics that differentiates between benzamidine and diazamidine in binding to trypsin.

The calculated binding energy of diazamidine is thus weaker than that of benzamidine by 1.21 kcal·mol^{-1}, compared with the experimental value of 1.59 kcal·mol^{-1} (33). Note that diazamidine was solvated much more favorably than benzamidine in both water and trypsin. The total solvation free energy of diazamidine in water is -72.35 kcal·mol^{-1} (Table 2), similar to that of a potassium ion (34). The 1.21 kcal·mol^{-1} difference was further decomposed into its electrostatic (0.11 kcal·mol^{-1}), vdW (0.98 kcal·mol^{-1}), and geometric (0.12 kcal·mol^{-1}) components. These results indicate that vdW is actually the main cause for the drop in the binding affinity, because the large free-energy changes due to electrostatics cancel out between water and trypsin–water environments.

Conclusion

A polarizable force field was applied to compute the binding affinity of a positively charged ligand and its analog to protein. Parameters were either directly derived from QM or transferred from the protein force field without modification or recalibration. Molecular dynamics simulations were performed with double decoupling of benzamidine from both water and trypsin binding site with free-energy perturbation. Different thermodynamic sampling paths were used by varying the softness of vdW potential and the restraint strength, and the resulting free energies were consistent. The computed absolute binding free energy is well within experimental accuracy.

Our results indicate that the electrostatics is the driving force for benzamidine binding to trypsin. We have further evaluated the role of polarization in binding by "turning off" the dipole induction between the ligand and its environment. It was found that polarization response varies drastically depending on the nature of the environment, and its contribution to the decoupling free energy does not simply cancel between water and protein. As a result of this finding, we believe that it is critical to treat polarization explicitly to achieve chemical accuracy in predicting the binding affinity of charged systems.

We also evaluated the relative binding free energy of a second ligand, diazamidine, with a pyrimidine group in place of the phenyl ring of benzamidine. On mutation from benzamidine to diazamidine, the binding weakens by 1.21 kcal·mol^{-1}, consistent with 1.59 kcal·mol^{-1} measured by experiment. Nevertheless, our calculations show that the changes in the solvation free energy in both water and trypsin complex are actually tens of kilocalories. It is the cancellation between the water and trypsin that leads to the subtle change in the binding energy.

In summary, electrostatics and polarization play important roles in molecular recognition and need to be accounted for in quantitative modeling. Our study demonstrates that chemical accuracy in predicting protein–ligand binding free energy can be achieved with a polarizable potential energy function when adequate sampling is possible.

Methods

Potential Energy Model and Ligand Parameterization. The potential energy of the system (protein, water, and ligand) is given by

$$E = E_{\text{ele}} + E_{\text{vdW}} + E_{\text{bond}} + E_{\text{angle}} + E_{\text{torsion}} + E_{\text{oop}} \quad [1]$$

The electrostatic contribution consists of permanent and polarized (by atomic-induced dipole) components (22, 35, 36). The van der Waals interaction is described by a buffered-14-7 function (37). The valence terms consist of typical harmonic function for bond stretching, angle bending, three-term Fourier torsional potential, and out-of-plane term for trigonal centers, as in the MM3 potential by Allinger and coworkers (38).

Previously, a potential, AMOEBA, based on this model has been developed for water (36, 39), ions (34, 40, 41), and organic molecules and peptides (22, 35). Independent studies using AMOEBA have also been reported (42–45). Parameters for proteins are available with the TINKER modeling package (46). In the current work, the AMOEBA potential was used for trypsin without modification. For the ligand, the electrostatic parameters were derived from QM, whereas vdW and other parameters were transferred from AMOEBA protein potential. The details of parameterization and ligands parameters are provided in the supporting information (SI) Dataset S1.

Absolute Binding Free Energy from Double-Decoupling Simulations. The absolute free energy of benzamidine binding to trypsin was calculated by using the double-decoupling method by "disappearing" the ligand in both water and solvated protein complex in two sets of simulations. The electrostatic interactions were decoupled in 10 steps by scaling down ligand atomic multipoles and polarizabilities linearly. Similar thermodynamic cycles were adopted in previous studies where atomic charges were zeroed out (10, 21). It is worth noting that scaling the ligand electrostatic parameters turns off electrostatic interactions between the ligand and its environment, and within the ligand, resulting in seemingly small decoupling energies (Table 1). To evaluate the complete ligand solvation energy (Table 2), the ligand was recharged in vacuum after being decoupled from its environment. However, the recharging energy is identical for both ligand–water and ligand–protein, and hence makes no contribution to the binding free energy. The recharging in vacuum was done by scaling the multipoles and atomic polarizabilities back to their full values in six steps linearly. A time step of 0.1 fs was used, with polarization convergence set to 1×10^{-5} D per atom.

VdW interactions between the ligand and environment were decoupled by directly modifying the pairwise interactions. To avoid singularity at small-vdW interaction distances, a situation where ligand atoms are in close contact and vdW energy approaches infinity, we replaced the buffered-14-7 vdW function with a soft-core modification (47):

$$U_{ij} = \lambda^n \varepsilon_{ij} \frac{1.07^7}{[\alpha(1-\lambda)^2 + (\rho + 0.07)^7]} \cdot \left(\frac{1.12}{\alpha(1-\lambda)^2 + \rho^7 + 0.12} - 2 \right) \quad [2]$$

At $\lambda = 1$, Eq. **2** reduces to the original buffered-14-7 function. By varying λ from 1.0 to 0.0, vdW interactions between ligand and its environment were turned off linearly in 10 uniform steps.

MD simulations were performed in parallel for all steps along the decoupling pathway. The SANDER executable from the AMBER package (v. 9) was used (48). The benzamidine–trypsin complex (1BTY) was placed in a periodic octahedral water box of 2,222 water molecules. The initial dimensions of the cube that encloses the octahedron were 50 Å on each side. Particle Mesh Ewald (49) was used to compute electrostatic interactions, with a real-space cutoff of 7.0 Å. Further details of MD simulation are given in the *SI Text*.

Downloaded from https://www.pnas.org by 38.88.236.210 on October 11, 2024 from IP address 38.88.236.210.

BeiGene Exhibit 1127.005
BeiGene v. Pharmacyclics
PGR2024-00009

A harmonic virtual bond was used to restrain the ligand to the protein pocket during the decoupling process (14, 50):

$$U(r) = \frac{k}{2}(r - r_0)^2 \qquad [3]$$

where k is the force constant. The positional fluctuation of benzamidine in the binding pocket of trypsin was measured from 100-ps MD simulations without restraint. A force constant of 20 kcal·mol^{-1}·Å^{-2} was subsequently obtained from $k = 3RT/\langle \delta r^2 \rangle$, where δr is the atomic position fluctuation.

The final binding free energy was calculated from the difference between the decoupling free energies in both water and protein–water environments (50, 51):

$$\Delta A_{\mathrm{calc}} = \Delta A_{\mathrm{wat}}(L \rightarrow 0) - \Delta A_{\mathrm{pro}}(L \rightarrow 0) + \Delta A_{\mathrm{correction}} \qquad [4]$$

A correction was made to the final binding free energy to remove the bias introduced by the harmonic restraint, $-RT \ln(C^{\circ}V)$ where C° is the standard concentration and V is the sampling volume of the ligand under the restraint. The correction term amounts to 5.87 kcal·mol^{-1} for $k = 20$ kcal·mol^{-1}·Å^{-2}, and 6.48 kcal·mol^{-1} for $k = 40$ kcal·mol^{-1}·Å^{-2}.

Free-Energy Analysis. The free-energy changes between adjacent steps were computed by using the Bennett acceptance ratio method (BAR) (52). The free-energy change between simulations λ_i and λ_{i+1} was computed iteratively by using equation 12 from ref. 52. The statistical error of estimated free energy between successive steps was computed from equation 10 in ref. 52. The total statistical error in decoupling free energy is computed as the sum of the errors from individual steps.

Polarization Free Energy. We computed the free energy caused by the induction between the ligand and its environment. The polarization between ligand and surrounding atoms was turned off by scaling the damping coefficient from the original 0.39 (36) to 0 in 5 steps (0.39, 0.039, 0.0039, 0.00039, 0). The polarization within water or protein–water remained unchanged. A 500-ps MD simulation run was carried out at each step. Because dipole induction is short-ranged, a cutoff of 14 Å and 8 Å was used for damping in protein and in bulk water, respectively.

ACKNOWLEDGMENTS. This work was supported by National Institute of General Medical Sciences Grant R01GM079686. T.D. was supported by the Intramural Research Program of the National Institutes of Health, National Institute of Environmental Health Sciences.

1. Lamb ML, Jorgensen WL (1997) Computational approaches to molecular recognition. *Curr Opin Chem Biol* 1:449–457.
2. Kollman PA, *et al.* (2000) Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc Chem Res* 33:889–897.
3. Brandsdal BO, *et al.* (2003) Free energy calculations and ligand binding. *Protein Simul* 66:123–158.
4. Jorgensen WL (2004) The many roles of computation in drug discovery. *Science* 303:1813–1818.
5. Gilson MK, Zhou HX (2007) Calculation of protein-ligand binding affinities. *Annu Rev Biophys Biomol Struct* 36:21–42.
6. Woo HJ, Roux B (2005) Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc Natl Acad Sci USA* 102:6825–6830.
7. Burykin A, Schutz CN, Villa J, Warshel A (2002) Simulations of Ion current in realistic models of ion channels: The Kcsa potassium channel. *Proteins* 47:265–280.
8. Warshel A, Sharma PK, Kato M, Parson WW (2006) Modeling Electrostatic effects in proteins. *Biochim Biophys Acta Proteins Proteom* 1764:1647–1676.
9. Kato M, Warshel A (2005) Through the channel and around the channel: validating and comparing microscopic approaches for the evaluation of free energy profiles for ion penetration through ion channels. *J Phys Chem B* 109:19516–19522.
10. Lee FS, Chu ZT, Bolger MB, Warshel A (1992) Calculations of antibody Antigen interactions—Microscopic and semimicroscopic evaluation of the free-energies of binding of phosphorylcholine analogs to Mcpc603. *Protein Eng* 5:215–228.
11. Aqvist J, Medina C, Samuelsson JE (1994) New method for predicting binding-affinity in computer-aided drug design. *Protein Eng* 7:385–391.
12. Sham YY, Chu ZT, Tao H, Warshel A (2000) 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* 39:393–407.
13. Warshel A, Kato M, Pisliakov AV (2007) Polarizable force fields: History, test cases, and prospects. *J Chem Theory Comput* 3:2034–2045.
14. Boresch S, Tettinger F, Leitgeb M, Karplus M (2003) Absolute binding free energies: A quantitative approach for their calculation. *J Phys Chem B* 107:9535–9551.
15. Deng YQ, Roux B (2006) Calculation of standard binding free energies: Aromatic molecules in the T4 lysozyme L99a mutant. *J Chem Theory Comput* 2:1255–1273.
16. Hermans J, Wang L (1997) Inclusion of loss of translational and rotational freedom in theoretical estimates of free energies of binding. Application to a complex of benzene and mutant T4 lysozyme. *J Am Chem Soc* 119:2707–2714.
17. Fujitani H, *et al.* (2005) Direct calculation of the binding free energies of Fkbp ligands. *J Chem Phys* 123:084108.
18. Wang JY, Deng YQ, Roux B (2006) Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophys J* 91:2798–2814.
19. Warshel A, Levitt M (1976) Theoretical studies of enzymic reactions—Dielectric, electrostatic and steric stabilization of carbonium-ion in reaction of lysozyme. *J Mol Biol* 103:227–249.
20. Warshel A, Sussman F, Hwang JK (1988) Evaluation of catalytic free-energies in genetically modified proteins. *J Mol Biol* 201:139–159.
21. Florian J, Goodman MF, Warshel A (2002) Theoretical Investigation of the binding free energies and key substrate-recognition components of the replication fidelity of human DNA polymerase beta. *J Phys Chem B* 106:5739–5753.
22. Ponder JW, Case DA (2003) Force fields for protein simulations. *Adv Protein Chem* 66:27–85.
23. Kato M, Braun-Sand S, Warshel A (2008) *Computational and Structural Approaches to Drug Discovery: Ligand-Protein Interactions*, eds Stroud R, Finer-Moore J (Royal Society of Chemistry, Cambridge, UK), pp 268–290.
24. Katz BA, *et al.* (2000) Structural basis for selectivity of a small molecule, S1-binding, submicromolar inhibitor of urokinase-type plasminogen activator. *Chem Biol* 7:299–312.
25. Katz BA, *et al.* (2001) A novel serine protease inhibition motif involving a multi-centered short hydrogen bonding network at the active site. *J Mol Biol* 307:1451–1486.
26. Mares-Guia M (1965) Studies on active center of trypsin—Binding of amidines and guanidines as models of substrate side chain. *J Biol Chem* 240:1579.
27. Schwarzl SM, Tschopp TB, Smith JC, Fischer S (2002) Can the calculation of ligand binding free energies be improved with continuum solvent electrostatics and an ideal-gas entropy correction? *J Comput Chem* 23:1143–1149.
28. Geerke DP, van Gunsteren WF (2007) Calculation of the free energy of polarization: Quantifying the effect of explicitly treating electronic polarization on the transferability of force-field parameters. *J Phys Chem B* 111:6425–6436.
29. Hensen C, *et al.* (2004) A combined Qm/Mm approach to protein-ligand interactions: Polarization effects of the Hiv-1 protease on selected high affinity inhibitors. *J Med Chem* 47:6673–6680.
30. Aqvist J (1996) Calculation of absolute binding free energies for charged ligands and effects of long-range electrostatic interactions. *J Comput Chem* 17:1587–1597.
31. Wang W, Wang J, Kollman PA (1999) What determines the van der Waals coefficient beta in the LIE (linear interaction energy) method to estimate binding free energies using molecular dynamics simulations? *Proteins* 34:395–402.
32. Leiros HKS, *et al.* (2004) Trypsin specificity as elucidated by lie calculations, x-ray structures, and association constant measurements. *Protein Sci* 13:1056–1070.
33. Grater F, Schwarzl SM, Dejaegere A, Fischer S, Smith JC (2005) Protein/ligand binding free energies calculated with quantum mechanics/molecular mechanics. *J Phys Chem B* 109:10474–10483.
34. Grossfield A, Ren PY, Ponder JW (2003) Ion solvation thermodynamics from simulation with a polarizable force field. *J Am Chem Soc* 125:15671–15682.
35. Ren PY, Ponder JW (2002) Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. *J Comput Chem* 23:1497–1506.
36. Ren PY, Ponder JW (2003) Polarizable atomic multipole water model for molecular mechanics simulation. *J Phys Chem B* 107:5933–5947.
37. Halgren TA (1992) Representation of Van Der Waals (Vdw) interactions in molecular mechanics force fields: Potential form, combination rules, and Vdw parameters. *J Am Chem Soc* 114:7827–7843.
38. Allinger NL, Yuh YH, Lii J-H (1989) Molecular mechanics. The Mm3 force Field for hydrocarbons. *J Am Chem Soc* 111:8551–8566.
39. Ren PY, Ponder JW (2004) Temperature and pressure dependence of AMOEBA water model. *J Phys Chem B* 108:13427–13437.
40. Jiao D, King C, Grossfield A, Darden TA, Ren PY (2006) Simulation of Ca^{2+} and Mg^{2+} solvation using polarizable atomic multipole potential. *J Phys Chem B* 110:18553–18559.
41. Piquemal JP, *et al.* (2006) Towards accurate solvation dynamics of divalent cations in water using the polarizable amoeba force field: From energetics to structure. *J Chem Phys* 125:054511.
42. Tuma L, Jenicek D, Jungwirth P (2005) Propensity of heavier halides for the water/vapor interface revisited using the amoeba force field. *Chem Phys Lett* 411:70–74.
43. Liang T, Walsh TR (2006) Molecular dynamics simulations of peptide carboxylate hydration. *Phys Chem Chem Phys* 8:4410–4419.
44. Liang T, Walsh TR (2007) Simulation of the hydration structure of glycyl-alanine. *Mol Simul* 33:337–342.
45. Jiang H, Jordan KD, Taylor CE (2007) Molecular dynamics simulations of methane hydrate using polarizable force fields. *J Phys Chem B* 111:6486–6492.
46. Ponder JW (2004) Tinker: Software Tools for Molecular Design (Washington Univ, Saint Louis, MO), Version 4.2. Available at: http://dasher.wustl.edu/tinker/. Accessed on April 5, 2008.
47. Beutler TC, Mark AE, Vanschaik RC, Gerber PR, Vangunsteren WF (1994) Avoiding singularities and numerical instabilities in free-energy calculations based on mol simuls. *Chem Phys Lett* 222:529–539.
48. Case DA, *et al.* (2005) The amber biomolecular simulation programs. *J Comput Chem* 26:1668–1688.
49. Sagui C, Pedersen LG, Darden TA (2004) Towards an accurate representation of electrostatics in classical force fields: Efficient implementation of multipolar interactions in biomolecular simulations. *J Chem Phys* 120:73–87.
50. Hamelberg D, McCammon JA (2004) Standard free energy of releasing a localized water molecule from the binding pockets of proteins: Double-decoupling method. *J Am Chem Soc* 126:7683–7689.
51. Gilson MK, Given JA, Bush BL, McCammon JA (1997) The statistical-thermodynamic basis for computation of binding affinities: A critical review. *Biophys J* 72:1047–1069.
52. Bennett CH (1976) Efficient estimation of free-energy differences from monte-carlo data. *J Comput Phys* 22:245–268.
53. Maresguia M, Nelson DL, Rogana E (1977) Electronic effects in interaction of para-substituted benzamidines with trypsin—Involvement of pi-electronic density at central atom of substituent in binding. *J Am Chem Soc* 99:2331–2336.

Downloaded from https://www.pnas.org by 38.88.236.210 on October 11, 2024 from IP address 38.88.236.210.

BeiGene Exhibit 1127.006
BeiGene v. Pharmacyclics
PGR2024-00009