Article

# Distinguishing Binders from False Positives by Free Energy Calculations: Fragment Screening Against the Flap Site of HIV Protease

Nanjie Deng,*[†,‡] Stefano Forli,[§] Peng He,[†,‡] Alex Perryman,[§] Lauren Wickstrom,[∥] R. S. K. Vijayan,[†,‡] Theresa Tiefenbrunn,[§] David Stout,[§] Emilio Gallicchio,[⊥] Arthur J. Olson,[§] and Ronald M. Levy*[†,‡]

[†]Center for Biophysics & Computational Biology/ICMS, [‡]Department of Chemistry, Temple University, Philadelphia, Pennsylvania 19122, United States

[§]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, California 92037, United States

[∥]Borough of Manhattan Community College, The City University of New York, Department of Science, New York, New York 10007, United States

[⊥]Department of Chemistry, Brooklyn College, the City University of New York, Brooklyn, New York, United States

ABSTRACT: Molecular docking is a powerful tool used in drug discovery and structural biology for predicting the structures of ligand−receptor complexes. However, the accuracy of docking calculations can be limited by factors such as the neglect of protein reorganization in the scoring function; as a result, ligand screening can produce a high rate of false positive hits. Although absolute binding free energy methods still have difficulty in accurately rank-ordering binders, we believe that they can be fruitfully employed to distinguish binders from nonbinders and reduce the false positive rate. Here we study a set of ligands that dock favorably to a newly discovered, potentially allosteric site on the flap of HIV-1 protease. Fragment binding to this site stabilizes a closed form of protease, which could be exploited for the design of allosteric inhibitors. Twenty-three top-ranked protein−ligand complexes from AutoDock were subject to the free energy screening using two methods, the recently developed binding energy analysis method (BEDAM) and the standard double decoupling method (DDM). Free energy calculations correctly identified most of the false positives (≥83%) and recovered all the confirmed binders. The results show a gap averaging ≥3.7 kcal/mol, separating the binders and the false positives. We present a formula that decomposes the binding free energy into contributions from the receptor conformational macrostates, which provides insights into the roles of different binding modes. Our binding free energy component analysis further suggests that improving the treatment for the desolvation penalty associated with the unfulfilled polar groups could reduce the rate of false positive hits in docking. The current study demonstrates that the combination of docking with free energy methods can be very useful for more accurate ligand screening against valuable drug targets.
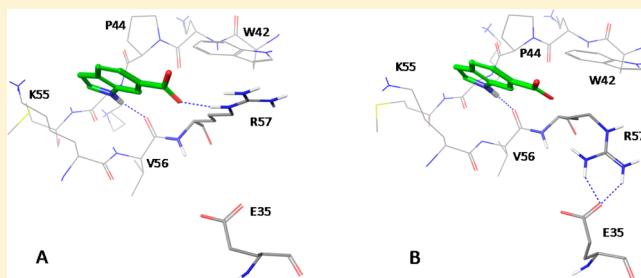
INTRODUCTION

Molecular docking is widely used in rational drug discovery and structural biology for predicting the most favorable pose and for estimating the strength of ligand−receptor binding.[1,2] In a typical virtual screening application, a large library of compounds is docked against a receptor target site to generate plausible poses ranked by scoring functions. Such functions are typically designed to have a simple form for computational efficiency. While docking has matured into a powerful tool for pharmaceutical research after decades of development,[1−7] the accuracy of docking calculations continues to be limited by these relatively simple scoring functions which lack a complete treatment of desolvation and receptor reorganization.[8,9] Additionally, entropic factors are generally not captured well by scoring based on a single structure.[8,10] As a result, structure-based ligand screening by docking often generates a large number of false positive hits. As a recent example, Shoichet et al.[11] conducted a parallel study of docking and HTS to screen 197861 compounds against cruzain, a thiol protease with a relatively rigid binding pocket. Among the top 0.1% of the docking-ranked library, 97.5% of the hits were found to be false positives.[11]

Binding free energy methods are based on statistical mechanics and atomistic simulations and, in principle, can

capture the desolvation, receptor reorganization, and entropic effects, which provides a less empirical route to the calculation of ligand-binding affinities.[12,13] The methodology of free energy perturbation for computing the relative and absolute free energy of molecular association was pioneered by Jorgensen,[14,15] McCammon,[16−18] and Kollman[19−21] 30 years ago. Since then, advances in the methodology, energy functions, and computer hardware have enabled free energy calculations to play an increasingly important role in the study of biomolecular recognition.[22−38] Significant progress has been made in recent years in applying the free energy perturbation method (FEP) to the discovery of highly potent drug molecules for pharmaceutical research.[31,39−41] While the accuracy of absolute binding free energy methods is still constrained by the quality of the current force fields and the extent of sampling, which makes it challenging to rank-order binders with similar binding affinities, we believe that free energy methods can be fruitfully applied as additional filters for docking to separate binders from nonbinders. Because of the high computational cost associated with free energy calculations, currently it is only practical to perform such simulations on a relatively small set of top ligands obtained from docking. While the calculation of the absolute binding free energies of libraries containing thousands of ligands is beyond the current technology, the use of free energy methods to score the binding affinity of hundreds of ligands to their target receptor is now possible.[8] In a recent study, we have shown that the combined application of docking and free energy methods resulted in a large improvement[8] over docking alone,[42] which helps researchers to focus on the true binders and prioritize synthetic chemistry efforts more effectively.

HIV-1 protease (PR) has been a major drug target for antiviral therapy against AIDS.[43] To date, a total of nine protease inhibitors have been FDA-approved, all of which are active site binders; they constitute a key component in the highly active antiretroviral therapy (HAART) cocktails. However, the efficacy of these active-site inhibitors has been reduced by the emergence of drug-resistance mutations. Novel inhibition strategies targeting alternative, allosteric sites of PR are needed to overcome the drug resistance mutations affecting active site inhibition. The active site of PR is covered by two flexible β-hairpin flaps that control substrate access to the active site cavity, and the mobility of the flaps is crucial to the enzymatic activity of PR.[44−46] The flaps of PR can exist in different conformations: most apo structures adopt a semiopen conformation, whereas a closed conformation is favored when the active site is occupied by a ligand.[47,48] Using fragment screening by X-ray crystallography, Perryman and Tiefenbrunn and co-workers recently discovered a new binding site, named the flap binding site, located on top of the flaps of PR (Figure 1).[49−51] Fragment 1F1 (indole-6-carboxylic acid) binding to this flap site was found to stabilize the closed conformation of PR, even in the absence of an active site ligand.[50] This suggests that ligand binding to the flap site could potentially enhance the binding affinity of an active site inhibitor by reducing the protein reorganization free energy cost associated with the semiopen-to-closed transition.[46] Tiefenbrunn et al. performed docking to virtually screen a focused library of 2518 compounds selected for similar structural features against the 1F1 flap binding site.[50] 41 top hits were selected for further experimental screening, including cocrystallization. Out of these 41 top hits, only a small fraction of the fragments showed any binding signal in various assays.
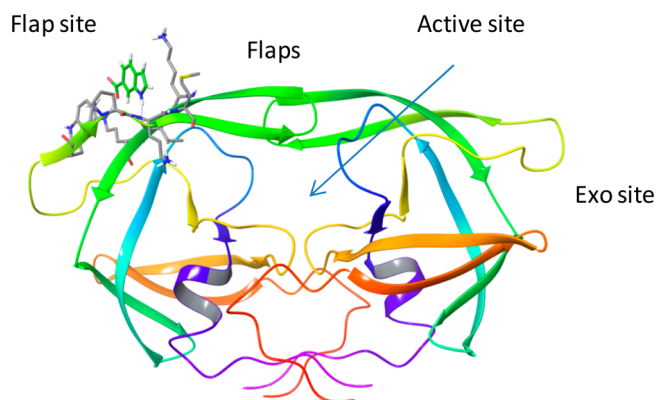

Figure 1. Crystal structure of HIV PR (pdb id: 3kfr) with its flap site occupied by the ligand 1F1 shown in green stick. The active site ligand is removed from the figure for clarity.

In this work, we perform binding free energy calculations on a set of ligands that dock favorably to the flap site of PR. The structures of the protein−ligand complexes from AutoDock[1,2] were used as the starting point for the free energy calculation. Two free energy methods, the binding energy distribution analysis method (BEDAM) and the double decoupling method (DDM) were employed. While DDM[25,52,53] is the standard method for computing absolute binding free energy in explicit solvent, the recently developed BEDAM[54] method employs Halmiltonian replica exchange in an implicit solvent model to accelerate the sampling of the phase space. The calculations are performed for 23 ligands, including 3 confirmed actives, 8 likely binders, and 12 false positives. A majority of the false positive ligands chosen have more favorable docking score than that of the true binders, which increases the challenge for the free energy methods. Our calculations using each of the methods correctly identified ≥83% of the false positives and recovered all of the confirmed binders. A gap averaging ≥3.7 kcal/mol in the computed binding free energy is found between the binders and false positives. Six out of eight likely binders are predicted to bind at the flap site, a prediction that requires further confirmation by NMR screening with site labeling. The free energy simulations also revealed the different binding modes for certain binders carrying a carboxylate group. One of the binding modes contains an intermolecular salt bridge between Arg57 and the ligand carboxylate ion. In the other binding mode, this intermolecular salt bridge is replaced by the intramolecular salt bridge between Arg57 and Glu35. Free energy calculations suggest that the intermolecular salt bridge provides the largest contribution to the binding affinity for 1F1. Analysis of the free energy decomposition results revealed that the main reason for the high false positive rate in docking could be due to the presence of partially buried, unfulfilled polar groups for which the high desolvation penalty is not adequately captured by the scoring function. The free energy calculation also provided insights into fragment optimization for potency enhancement. This study provides further confirmation of the power of combining docking with free energy methods for accurate ligand screening against pharmaceutical targets. The main challenge in successfully applying this approach to ligand screening is to be able to carry out a relatively large number of absolute binding free energy calculations in a reliable, automated, and rapid fashion.[8,55]

B

dx.doi.org/10.1021/jp506376z | J. Phys. Chem. B XXXX, XXX, XXX−XXX

■ COMPUTATIONAL METHODS

Binding Energy Distribution Analysis Method (BEDAM). In the BEDAM approach,[54] which uses the OPLS force field[56,57] and an implicit solvation model AGBNP2,[58] the standard binding free energy (ΔG_b^0) is computed using the following hybrid effective potential connecting the unbound ($\lambda = 0$) and bound state ($\lambda = 1$)

$$U_\lambda(r) = U_0(r) + \lambda u(r) \tag{1}$$

Here, the unbound state Hamiltonian $U_0(r)$ is the sum of the effective energies of a receptor A and a ligand B when the two are fully decoupled from each other

$$U_0(r) = U(r_A) + U(r_B) \tag{2}$$

And $u(r)$ is the binding energy function, i.e. for each conformation $r = (r_A, r_B)$, $u(r)$ is the difference in the effective energy between the complex and the two dissociated molecules A and B:

$$u(r) = U(r_A, r_B) - U(r_A) - U(r_B) \tag{3}$$

From eqs 1−3, it can be seen that, when λ goes from 0 to 1, the Hamiltonian of the system changes from that of the unbound state $U_0(r)$ to that of the fully coupled state $U_1(r) = U(r_A, r_B)$. The binding free energy ΔG_b is the free energy difference between the two end states

$$\Delta G_{0 \to 1} = -kT \ln <e^{-u/kT}>_0 = -kT \ln \int P_0(u) e^{-u/kT} \, du \tag{4}$$

The integrand is a product of a steeply increasing function of u, $P_0(u)$, and a steeply decreasing function $e^{-u/kT}$. Therefore, it is necessary to calculate the favorable energy tail of the distribution $P_0(u)$ accurately. To accomplish this, BEDAM employs a Hamiltonian Replica Exchange λ-hopping strategy (H-REM), in which the simulated systems at different λ periodically attempt to exchange their configurations through MC moves. The use of H-REM has been shown to yield superior conformational sampling and more rapid convergence rates by allowing conformational transitions to occur at values of λ at which they are most likely to occur and to be then propagated to other states.[54,59]

In BEDAM, $\Delta G_{0 \to 1}$ is computed using the multistate Bennett acceptance ratio estimator (MBAR)[60] $\Delta G_{0 \to 1} = kT(\hat{f}_1 - \hat{f}_0)$, from values of binding energy (u) sampled at a series of intermediate λ values using molecular dynamics simulations. Here $\hat{f}_\lambda$ is the MBAR dimensionless free energy defined as the negative of the logarithm of the partition function at an intermediate λ, $\hat{f}_\lambda = -\ln Z_\lambda$. The values of $\hat{f}_\lambda$ are obtained from the self-consistent solution of the set of MBAR equations[60]

$$\hat{f}_i = -\ln \sum_{j=1}^{K} \sum_{n=1}^{N_j} \frac{\exp(-\lambda_i u_{jn}/kT)}{\sum_{k=1}^{K} N_k \exp(\hat{f}_k - \lambda_k u_{jn}/kT)} \tag{5}$$

Here $\hat{f}_i = \hat{f}_{\lambda_i}$, u_{jn} is the nth binding energy sampled at intermediate λ_j replica, K is the number of replicas, and N_j is the number of samples in replica j. For the MBAR analysis, we employed the code provided by John Chodera and Michael Shirts (http://alchemistry.org).[60]

The standard binding free energy (ΔG_b^0) is then obtained using

$$\Delta G_b^0 = -k_B T \ln C^0 V_{site} + \Delta G_b \tag{6}$$

In this work, BEDAM simulations of the 23 ligand−PR complexes were performed for 3 ns per replica (48 ns of total simulation time per ligand). The systems were first energy-minimized and then gradually heated to 300 K in 75 ps and equilibrated at 300 K for an additional 75 ps before the production run.

Hamiltonian replica exchange simulations were conducted using 16 lambdas: 0.0, 0.001, 0.002, 0.004, 0.006, 0.008, 0.01, 0.02, 0.04, 0.07, 0.1, 0.25, 0.4, 0.55, 0.75, and 1.0. The binding site is defined as a sphere with a radius of 2 Å from the center of mass of residues 44, 46, 55, and 57 of chain B of the HIV-PR dimer. The formation of the complex is defined by the center of the mass of the ligand being within the binding site sphere. There are two equivalent binding sites on the HIV PR dimer. In this study, the site on chain B is used in the simulation. The Cα atoms of the receptor were restrained with a force constant of 2.0 kcal/mol/Å^2. These atomic restraints are applied in the BEDAM simulations in order to prevent the slow drift of the structure which can occur over long times. During the simulations, the structures were saved every picosecond. The last 2 ns of data were used for analysis. Statistical uncertainties were obtained by comparing the results computed using the first and second halves for the last 3 ns of each BEDAM simulation.

The binding free energy can be expressed as the sum of the reorganization free energy and the average binding energy,

$$\Delta G_b^0 = \Delta E_b + \Delta G_{reorg}^0 \tag{7}$$

As described in an earlier paper,[61] this decomposition corresponds to a hypothetical thermodynamic cycle in which the unbound ligand and protein in solution are first reorganized to match those of the complex, and in a subsequent step, interactions between the ligand and protein are turned on. The first step is associated with the reorganization free energy (ΔG_{reorg}^0). The second step is accompanied by the change in the effective potential energy ΔE_{bind}, which includes direct noncovalent interactions (electrostatic and van der Waals) as well as the net desolvation of the binding partners.[54,61] ΔE_{bind} is computed from the average, $\langle u \rangle_1$, of the binding energy function in the ensemble of conformations of the complex in the coupled state ($\lambda = 1$). ΔG_{reorg}^0 is computed from the difference of the computed binding free energy and the average binding energy:

$$\Delta G_{reorg}^0 = \Delta G_b^0 - <u>_1 \tag{8}$$

DDM Calculation. The double decoupling[25,52,53] calculations in explicit solvent (TIP3P water model[62] plus counterions) were performed for the 23 ligands at 300 K to estimate the binding free energies. The protein molecule is modeled by the Amber ff99sb-ILDN force field,[63] and the ligands are described by the Amber GAFF[64] parameters set. The partial charges of the ligands are obtained using the AM1-BCC method.[65] A DDM calculation involves two legs of simulation, in which a restrained ligand is gradually decoupled from the receptor binding pocket or from the aqueous solution. In each leg of the decoupling simulations, the Coulomb interaction is turned off first using 11 lambda windows, $\lambda = 0.0$, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0; the Lennard-Jones interactions are then turned off in 17 lambda windows, $\lambda = 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.94, 0.985,$ and 1.0. The two decoupling free energies $\Delta G^{gas* \to complex}$ and $\Delta G^{gas \to water}$ associated with the two legs of

C

dx.doi.org/10.1021/jp506376z | J. Phys. Chem. B XXXX, XXX, XXX−XXX

the DDM cycle were determined using thermodynamic integration (TI). The Hamiltonian derivative $\langle \partial U / \partial \lambda \rangle_\lambda$ at a series of λ from 0 to 1 were collected and integrated to obtain the free energy difference. The MD sampling at each λ were performed using the GROMACS[66,67] version 4.6.4 for 15 ns; the last 10 ns was used for the calculation of binding free energy using TI.

Umbrella Sampling. To calculate the PMF of the intramolecular distance Glu35-Arg57 using umbrella sampling,[68] a series of MD simulations is performed using GROMACS[66,67] (version 4.6.4) on the apo PR with the harmonic distance restraint between the CD atom of Glu35 and CA atom of Arg57. The biasing potential in the ith simulation window is $W_i(r) = (1/2)k_i(r - r_i^0)^2$, where k_i is the force constant, r is the reaction coordinate, and r_i^0 is the reference distance for the ith sampling window. The full range of the reaction coordinate space is covered using 20 windows centered at: $r_i^0 = 3.7, 4.0, 4.3, 4.6, 4.9, 5.2, 5.5, 5.8, 6.1, 6.4, 6.7, 7.0, 7.3, 7.6, 7.9, 8.2, 8.5, 8.8, 9.1,$ and 9.4 Å. A single force constant $k = 2.4$ kcal mol^{-1} Å^{-2} is used for all the sampling windows. In each sampling window, an 11 ns MD simulation is performed, starting from the last conformation of the previous sampling window. The first 1 ns is treated as equilibration, which allows the system to adjust to the current umbrella potential. The last 10 ns of sampling data are used for the calculation of PMF. The biased probability distributions of the end-to-end distance $P'(r_i)$ accumulated in these sampling windows are unbiased and combined using the weighted histogram analysis method (WHAM) method to yield the unbiased distribution $P(r_i)$ and the associated potential of mean force.[69]

■ RESULTS AND DISCUSSIONS

Performance of Ligand Screening by Free Energy. Experimentally, ligand binding to the flap site was determined using a range of techniques, including cocrystallization, backscattering interferometry (BSI), differential scanning fluorimetry (DSF), and surface plasmon resonance (SPR).[50] Among the 23 ligands studied in this work, three are crystallographically confirmed binders: 1F1, 1F1−N, and AK2097. Binding of these fragments was also independently confirmed in BSI and DSF measurements. Eight other ligands show a BSI-measured $K_d \leq 1$ mM and/or induce a sizable increase in the T_m of apo-PR as measured by DSF. Some of these eight ligands also inhibit the nucleation of PR crystals.[50] On the other hand, the cocrystallization with these eight ligands proved to be unsuccessful. Here we consider these as likely binders for the flap site because of the binding signals they present and because they can be readily docked into the binding cavity with good shape complementarity. It should be noted that since the binding signals from BSI and DSF do not specify the binding site for these ligands, the possibility that these likely ligands bind at a different site on HIV-PR cannot be ruled out. The 12 remaining ligands show no signal in any of the assays and are labeled as nonbinders.

Figure 2 shows the computed binding free energies ranked from low (favorable) to high (unfavorable) for all the ligands. Both free energy models (BEDAM and DDM) score binders (colored in blue in Figure 2) more favorably than the nonbinders (red). Table 1 shows the free energies and their uncertainties together with the information on binding from experiments. It can be seen that all the binders and many likely binders exhibit more favorable ΔG_b^0 compared with that of the nonbinders. The average binding free energies computed by
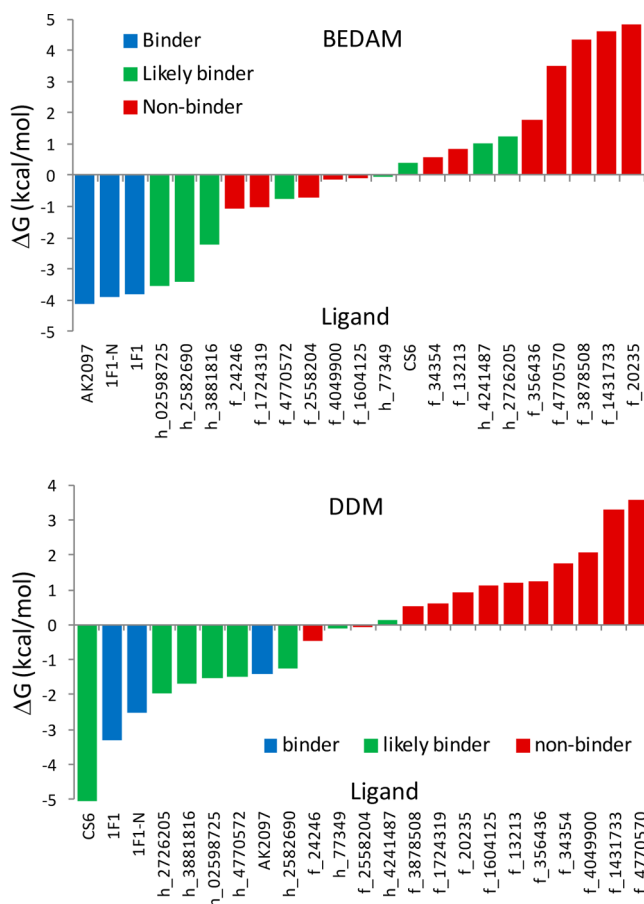

Figure 2. Computed binding free energies ranked from low-to-high from left-to-right. Upper: BEDAM; Lower: DDM.

BEDAM for these two ligand groups are −3.9 kcal/mol, and 1.4 kcal/mol, respectively. The corresponding values computed using DDM are −2.4 and 1.3 kcal/mol. Therefore, free energy calculations achieve a ≥ 3.7 kcal/mol binding free energy gap separating the binders from nonbinders. No comparable separation between the binders and nonbinders were detected using the docking scoring function. The free energy calculations show that the binding affinities for the flap site binders are weak (i.e., in the millimolar range). This is consistent with the fact that the binders are small fragments, with an average molecular weight of 202 Da. In fragment library screening, compounds with molecular weights of around 200 Da often exhibit 100 μM to 1 mM binding affinities.[70] Another reason for the weak affinity is that the flap site cavity is relatively exposed to the solvent. The one ligand that stands out as the strongest flap site binder according to the DDM calculation is CS6, which has a molecular weight of 533 Da and is significantly larger than the rest of the ligands. The structure of CS6 includes aromatic and heterocyclic moieties and sulfonate and amide groups. As seen from Table 1, CS6 is predicted by DDM to have a low micromolar binding affinity compared to the mM affinity for the three crystallographically confirmed binders. There is evidence that supports this computational prediction: the SPR experiments (unpublished results) of ligand binding to HIV PR suggest that CS6 binds significantly more tightly than the binders 1F1 and 1F1−N, in good agreement with the DDM computed binding free energies of these molecules (Table 1).

In addition to retrospectively identifying known binders and false positives, the calculations also provide testable predictions

Table 1. BEDAM and DDM Binding Free Energies, AutoDock Scores and Experimental Information for the 23 Ligands, Unit: kcal/mol[a]
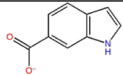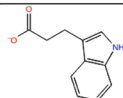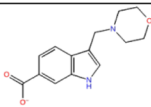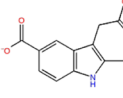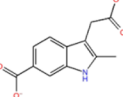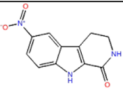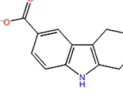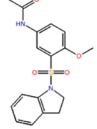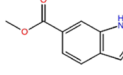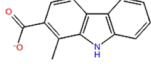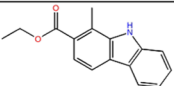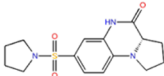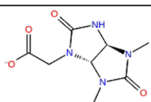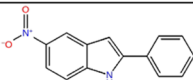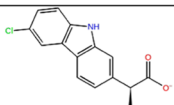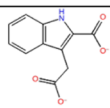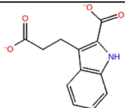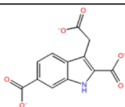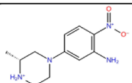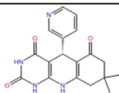
Ligand	Ligand Structure	ΔG_b^0 : BEDAM	ΔG_b^0 : DDM	Autodock score	Experimental results
1F1		-3.8 ± 0.3	-3.0 ± 0.2	-6.3	binder
1F1-N		-3.9 ± 0.1	-2.5 ± 0.2	-5.6	binder
AK2097		-4.1 ± 0.4	-1.4 ± 0.03	n/c	binder
CS6		0.38 ± 0.1	-7.5 ± 0.4	-8.2	likely binder
h_2582690		-3.4 ± 0.5	-1.2 ± 0.1	-6.5	likely binder
h_02598725		-3.6 ± 0.3	-1.5 ± 0.2	-6.7	likely binder
h_3881816		-2.2 ± 0.6	-1.7 ± 0.3	-7.6	likely binder
h_4770572		-0.8 ± 0.1	-1.5 ± 0.1	-7.5	likely binder
h_77349		-0.06 ± 0.5	-0.1 ± 0.2	-7.1	likely binder
h_2726205		1.2 ± 1.1	-2.0 ± 0.2	-6.9	likely binder
h_4241487		1.0 ± 0.7	0.2 ± 0.3	-7.3	likely binder
f_2558204		-0.7 ± 0.4	-0.05 ± 0.1	-5.1	non-binder
f_1604125		-0.1 ± 0.3	1.1 ± 0.1	-6.6	non-binder

Table 1. continued

Ligand	Ligand Structure	ΔG_b^0: BEDAM	ΔG_b^0 : DDM	Autodock score	Experimental results
f_1724319		-1.0 ± 0.6	0.6 ± 0.1	-5.9	non-binder
f_4049900		-0.1 ± 0.1	2.14 ± 0.4	-6.9	non-binder
f_356436		1.8 ± 0.4	1.2 ± 0.04	-6.2	non-binder
f_24246		-1.1 ± 0.6	-0.5 ± 0.2	-6.7	non-binder
f_20235		4.8 ± 0.5	0.9 ± 0.2	-6.9	non-binder
f_34354		0.6 ± 0.7	1.7 ± 0.2	-6.4	non-binder
f_13213		0.9 ± 0.5	1.2 ± 0.1	-6.5	non-binder
f_4770570		3.5 ± 0.4	3.6 ± 0.1	-7.3	non-binder
f_3878508		4.3 ± 0.7	0.5 ± 0.2	-7.7	non-binder
f_1431733		4.6 ± 0.7	3.3 ± 0.05	-7.2	non-binder

[a]The ligands shown in bold face are confirmed binders and likely binders. The entries of incorrectly predicted free energies are marked in red. The cutoff in the computed ΔG_b^0 for separating binders and nonbinders is chosen to be $\Delta G_b^0 \leq -1.0$ kcal/mol. The error bars are estimated by comparing the free energy results obtained from using the first and second halves of the simulation trajectories. n/c: docking was not performed for the ligand; the crystal structure was used in the free energy calculation. Due to the proprietary nature of the compound, the structure of CS6 has been withheld upon request.

for other likely binders: both BEDAM and DDM predict that three ligands in this category h_2582690, h_02598725, and h_3881816 bind at the flap site. In addition, DDM also predicts that CS6, h_2726205, and h_4770572 are also binders (Table 1). Experimental studies using NMR screening with site labeling are currently underway to test these computational predictions.

To assign binders and nonbinders, we use a cutoff $\Delta G_b^0 <$ −1.0 kcal/mol. This is chosen to match approximately the weakest computed binding free energy among all three confirmed binders (AK2097, $\Delta G_b^0 = -1.4$ kcal/mol, computed by DDM). Using this criterion, we analyze the performance of the free energy calculations in discriminating between binders and nonbinders. The result is summarized in Table 2. BEDAM

Table 2. Binders and Non-Binders Correctly Identified by Free Energy Calculations

	BEDAM	DDM
ligand	number of correct predictions	number of correct predictions
binders	3 out of 3 (100%)	3 out of 3 (100%)
nonbinders	10/12 (83.3%)	11/12 (91.6%)

and DDM calculations recovered all of the confirmed binders. Ten out of the 12 or 83.3% of the false positives are correctly identified by the BEDAM method. The DDM calculations have identified 91.6% the nonbinders. If we use a different cutoff, $\Delta G_b^0 < -0.5$ kcal/mol to assign the binders then the BEDAM and DDM would still correctly identify 75% and 83.3% of the false positives, respectively.

It is of interest to examine the correlation between results of the two free energy methods BEDAM and DDM. The correlation between the two sets of computed binding free energies is given in Figure 3. It can be seen that except for the
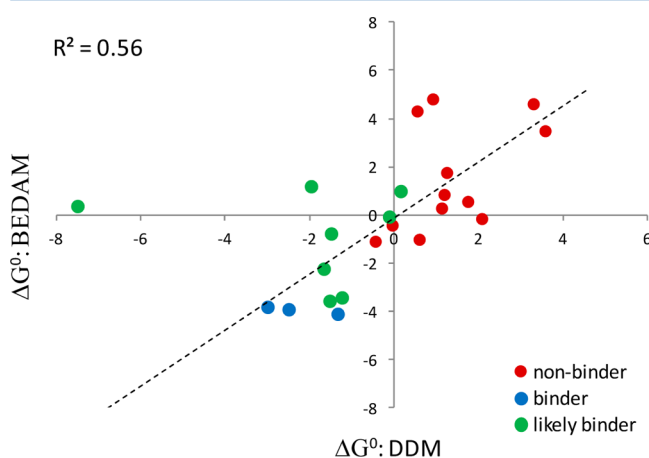
Figure 3. Correspondence between the binding free energies computed using BEDAM and DDM. Unit in kcal/mol. The ligand CS6 is excluded from the linear regression.

single outlier CS6, which is a likely binder, the majority of the points generally follow the similar trend. For CS6, its binding affinity was underestimated by the BEDAM calculation. If we exclude this one outlier, the correlation between the results from BEDAM and DDM is $R^2 = 0.56$, showing reasonably good agreement between the two methods despite the use of the different solvent models and force fields. It would be interesting in future studies to compare the two approaches using exactly the same force field in order to better assess the potential sources of errors.

Analysis of an Incorrectly Predicted Free Energy. While our free energy calculations are successful overall in distinguishing binders from false positives, there are also several incorrect predictions (entries shown in red in Table 1). Some of the erroneous predictions are likely to stem from limitations of the force field parameters for the ligands involved. One such example is with the likely binder CS6, for which the SPR assay shows that the molecule binds much more strongly than the crystallographic binders 1F1 and 1F1−N (unpublished data). This experimental result is consistent with the DDM calculated ΔG_b^0 of −7.5 kcal/mol for CS6, while the BEDAM calculation for this ligand yields an unfavorable ΔG_b^0 of 0.38 kcal/mol. The CS6 molecule contains a charged sulfonate group and a

nonpolar heterocyclic moiety. It is likely that the AGBNP2 solvation parameter (used by BEDAM) for the sulfonate group does not fully capture the desolvation penalty. In accordance with DDM simulations in explicit solvent, the sulfonate group of the ligand is solvated by water, but in the AGBNP2 implicit solvent model, the sulfonate group forms an intermolecular salt bridge between with the amine group of Lys55, which prevents the burial of the nonpolar heterocyclic ring of the ligand inside the binding cavity defined by Pro44, Met46, and Lys55 side chains. In addition, the AGBNP2 solvation parameter for the sulfur atom in the heterocyclic ring was not optimized to capture the crucial hydrophobic enclosure of this nonpolar moiety in the binding cavity. As a result, in the BEDAM simulated structure, the heterocyclic ring resides outside the binding cavity, while in the DDM simulated structure it is hydrophobically enclosed in the cavity, which significantly enhances the binding. We are currently working to improve the AGBNP2 solvation parameters to optimize binding free energy estimation.

Structural and Energetic Insights from Free Energy Simulations. Free energy simulations not only give estimates for the binding affinities but also provide structural and thermodynamic insights into ligand binding to the flap site. For the crystallographically confirmed binders, the free energy simulations reproduced the crystallographic binding modes (Figure 4). In addition, the simulations predicted the binding mode for the likely binder CS6 for which the crystal structures are not yet available.
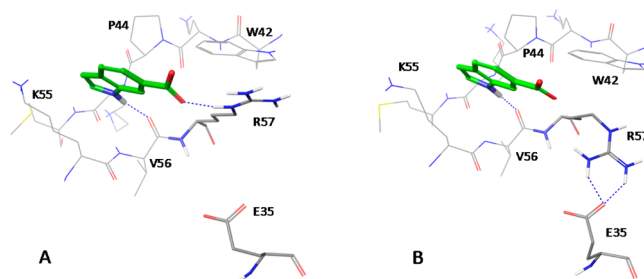

Figure 4. Two binding modes observed for the binder 1F1. (A) The dominant binding mode observed in the binders 1F1 and AK2097. (B) An alternative binding mode adopted by 1F1. The hydrogen bonds are shown in dotted blue.

Figure 4 illustrates the representative binding modes for binder 1F1 in the flap site of PR observed in both the DDM and BEDAM free energy simulations. The binding mode A (Figure 4A), which is also found in binder AK2097, contains three key ligand−receptor interactions: (1) the hydrophobic indole ring is enclosed in the nonploar pocket formed by Trp42, Pro44, Met46, and the Lys55 side chain; (2) the indole N−H group forms a buried hydrogen bond with the Val56 backbone carbonyl oxygen; and (3) the carboxylate group in the ligand forms an intermolecular salt bridge with Arg57. 1F1 can also bind in an alternative mode denoted mode B (Figure 4B), in which the intermolecular salt bridge between the ligand carboxylate and Arg57 is replaced by an intramolecular salt bridge Arg57-Glu35. The roles of the two binding modes A and B for 1F1 binding are discussed in a later section.

The DDM free energy simulation predicted structure of the likely binder CS6 reveals significant differences with the binding mode predicted by docking. The former is stabilized entirely by nonpolar interactions: (a) the enclosure of the heterocyclic ring

in the hydrophobic pocket formed by the Trp42, Pro44, Met46, and the Lys55 side chain and (b) the nonpolar interaction between the phenyl group in the ligand and Val56, and with the nonpolar atoms in the side chains of Lys55 and Arg57. There are no intermolecular hydrogen bonds in the simulation predicted complex of CS6-PR. This is in contrast to the docked structure, which contains several ligand−protein intermolecular hydrogen bonds, such as the sulfonate group in CS6 forming a salt bridge with Arg57, and the amino carbonyl group in CS6 forming hydrogen bond with the backbone carbonyl atom of Pro44. In the free energy simulation, which starts from the docked structure, such intermolecular hydrogen bonds on the protein surface become unstable; both the sulfonate and the amino carbonyl groups in the ligand moves away from the initial hydrogen-bonded position and toward the solvent. This is because both the sulfonate group and the amino carbonyl group are highly polar, carrying very favorable solvation free energy. By design, the docking method tries to maximize the number of ligand−protein intermolecular interactions; in solution, however, such interactions near the protein surface have to compete with the solvation forces that always act to weaken solute−solute interaction. This example with CS6 suggests that absolute binding free energy simulations in explicit solvent (DDM) could be used as an aid in improving scoring functions for docking.

Role of the Different Binding Modes. As shown in Figure 4 and also indicated by the different crystal structures of HIV PR, the side chains of Arg57 and Glu35 are quite mobile and can adopt different orientations. In Figure 4A, the Glu35 is solvated while Arg57 forms an intermolecular salt bridge with the carboxylate in the ligand (binding mode A); in Figure 4B, Arg57 forms a salt-bridge with Glu35 while the ligand carboxylate group is solvated (binding mode B). As shown below, our free energy simulations suggest that binding mode A (Figure 4A) is the more dominant conformation in solution, although both binding modes were observed in the crystal structures of 1F1. The transition from binding mode B to the more dominant mode A, which involves replacing the intramolecular salt bridge Glu35-Arg57 by the intermolecular salt bridge, has been observed in the free energy simulation of a similar binder 1F1−N, which also carries a carboxylate group: see Figure 5.

For ligand binding to multiple receptor conformations grouped into macrostates, the overall equilibrium binding constant is the weighted sum of the binding constant associated with the binding to each of the macrostates, as long as the macrostate conformations are defined in such a way that they include all possible conformations of the receptor.[12,54] Here we partition the receptor conformations into two macrostates A and B, as shown in Figure 6. The binding free energy can be expressed as

$$\Delta G_b = -kT \ln(P_A e^{-\Delta G_A/kT} + P_B e^{-\Delta G_B/kT}) \tag{9}$$

Here P_A and P_B are the populations of the receptor macrostates in the unbound state. ΔG_A and ΔG_B are the binding free energies restricted to the respective receptor macrostates. To derive eq 9, we start from the expression for the total binding free energy, $\Delta G_b = -kT \ln(Z_{RL})/(Z_R Z_L)$, where $Z_X = \int e^{-H_X(r)/kT} dr$ is the configuration integral for species X (RL: receptor−ligand complex; R and L, receptor and ligand, respectively, free in solution.) For a receptor having two
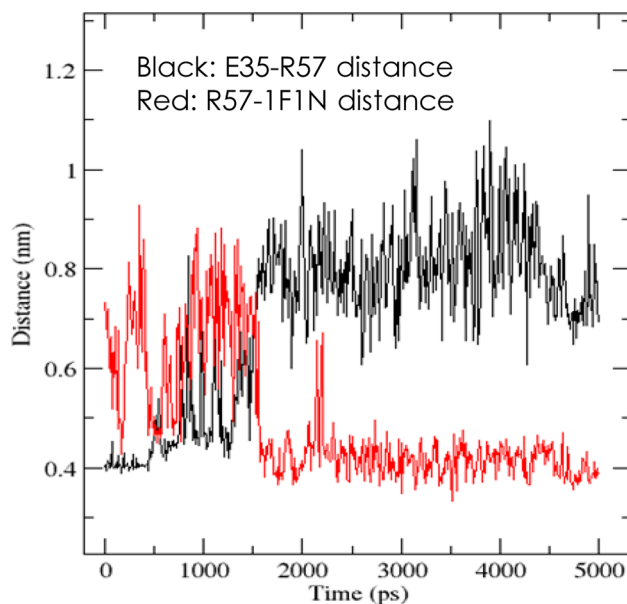

Figure 5. Conversion of the intramolecular salt bridge E35-R57 into the intermolecular salt bridge between R57 and the ligand 1F1−N, observed in the free energy simulation at $\lambda = 1$.
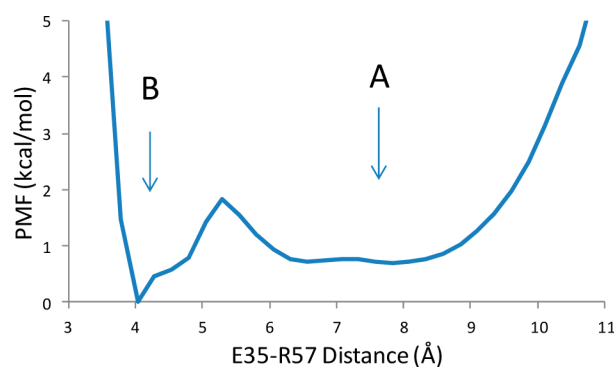

Figure 6. PMF along the Glu35-Arg57 distance in the apo PR computed by umbrella sampling MD simulations in explicit solvent.

conformational macrostates A and B, the total binding free energy can be written as

$$\Delta G_b = -kT \ln \frac{Z_{RL}}{Z_R Z_L}$$

$$= -kT \ln \frac{\int_A e^{-H_{RL}(r)/kT} dr + \int_B e^{-H_{RL}(r)/kT} dr}{Z_P Z_L}$$

$$= -kT \ln\left(\frac{Z_{R_A L}}{Z_R Z_L} + \frac{Z_{R_B L}}{Z_R Z_L}\right) \tag{10}$$

Note that

$$\frac{Z_{R_A L}}{Z_R Z_L} = \frac{Z_{R_A}}{Z_R} \times \frac{Z_{R_A L}}{Z_{R_A} Z_L} = P_A e^{-\Delta G_A/kT}$$

$$\frac{Z_{R_B L}}{Z_R Z_L} = \frac{Z_{R_B}}{Z_R} \times \frac{Z_{R_B L}}{Z_{R_B} Z_L} = P_B e^{-\Delta G_B/kT} \tag{11}$$

where Z_{R_A} and Z_{R_B} are the configuration integrals of the unbound receptor macrostates A and B respectively, and $Z_R =$

$Z_{R_A} + Z_{R_B}$. Substituting eq 11 into eq 10, we obtain the overall binding free energy as in eq 9.

An alternative approach to computing the overall binding free energy is to use the confine−release thermodynamic cycle,[30] which requires computing the potentials of mean force (PMF) for both the apo and holo receptor. In this work, we use eq 9 to estimate the contribution of two binding modes to the binding free energy of 1F1, which requires the calculation of the PMF for just the apo receptor. The unbound populations P_i were estimated from the unbound receptor PMF computed using umbrella sampling in which the Glu35-Arg57 distance is the reaction coordinate. The computed PMF along this reaction coordinate is shown in Figure 6, in which the two basins corresponding to the conformations A and B are separated by a free energy barrier of ∼2 kcal/mol. The unbound receptor populations P_A and P_B are found to be 0.61 and 0.39, respectively. The fact that the two basins have comparable occupancies and are separated by a moderate barrier is consistent with the observation that the Glu35 and Arg57 side chains are mobile and both conformations A and B are observed in different crystal structures.

The conditional binding free energies for each mode A and B are computed by performing two separate free energy simulations, in which the intramolecular distance E35-R57 is restrained to the respective conformational basin. The resulting binding free energies for the two binding modes are $\Delta G_A = -3.3$ kcal/mol and $\Delta G_b = -0.5$ kcal/mol. This shows that binding mode A, which features the intermolecular salt bridge between Arg57-ligand, makes the dominant contribution to ligand 1F1 binding. Substituting these values into eq 9, we obtain the overall $\Delta G_b^0 = -3.0$ kcal/mol, which is close to the binding free energy for mode A. [The error bars in the overall binding free energy (ΔG_b^0) and individual binding free energy (ΔG_A) are 0.2 kcal/mol, which is slightly smaller than their difference.] The presence of significant population of conformation A separated from conformation B by an appreciable free energy barrier in the unbound state (Figure 6) suggests that the 1F1 binding is likely to follow a conformational selection mechanism, in which the ligand specifically binds and stabilizes a preformed unbound conformation of the receptor. This example also shows that when there are multiple binding modes present, the overall affinity is usually dominated by the binding mode with the strongest individual binding affinity.

Thermodynamic Determining Factor Separating Binders from Nonbinders. To gain a deeper understanding of the determining factors and the thermodynamic driving forces for binding, we examine the free energy components $\Delta\Delta G$(elec) and $\Delta\Delta G$(vdw) that are associated with the charge-decoupling and vdw-decoupling stages of the DDM calculation, respectively (Figure 7, panels A and B). A DDM calculation has two simulation legs, in which a ligand is decoupled from the solvent environment and separately from the protein binding site. In both decoupling legs, the Coulomb intermolecular interaction is turned off first, and then the van der Waals intermolecular interaction is turned off. The electrostatic component of the binding free energy $\Delta\Delta G$(elec) is the difference between the free energy of turning off the Coulomb interaction in the solvent environment and that in the binding site environment i.e., $\Delta\Delta G$(elec) = $\Delta G_{\text{Coulomb}}$(water) − $\Delta G_{\text{Coulomb}}$(binding_site). The van der Waals component of the binding free energy $\Delta\Delta G$(vdw) is
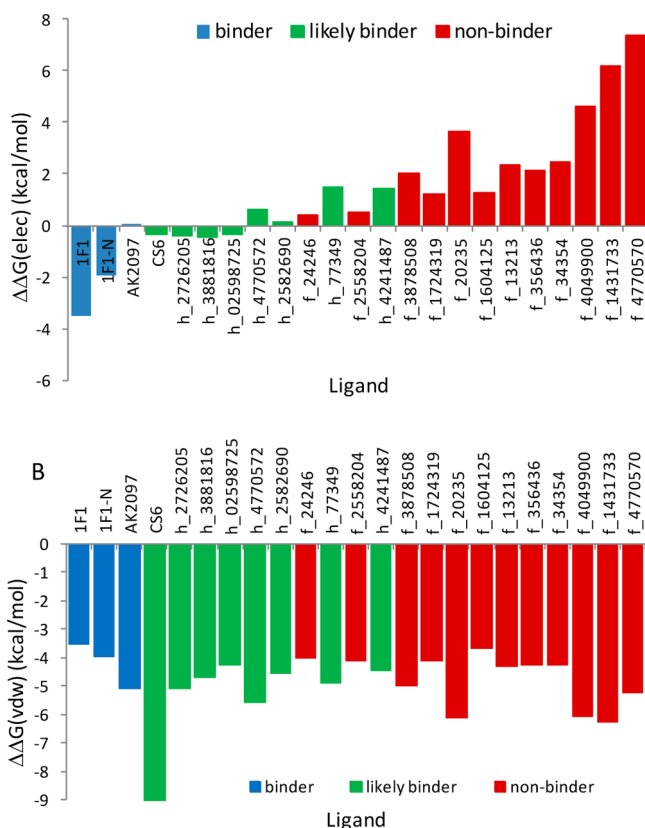

Figure 7. (A) Polar and (B) nonpolar components of the binding free energy computed by DDM.

computed in the analogous way, $\Delta\Delta G$(vdw) = ΔG_{vdw}(water) − ΔG_{vdw}(binding_site). These two components generally reflect the contributions from polar and nonpolar interactions to the overall binding free energy.[10] Here we note that the decomposition of the binding free energy is path dependent; therefore, the information drawn from such component free energies is only qualitative, yet it can facilitate insights into factors that promote binding. Figure 7 shows that on average, the $\Delta\Delta G$(elec) for the binders are either favorable or neutral, while the nonbinders have significantly more unfavorable $\Delta\Delta G$(elec). On the other hand, as seen from Figure 7, the binders and nonbinders have a similarly large, favorable nonpolar component $\Delta\Delta G$(vdw). Indeed, while the average $\Delta\Delta G$(elec) for the binders and nonbinders are separated by a gap of −4.7 kcal/mol (−1.8 and 2.9 kcal/mol, respectively) favoring the binders, the difference between the average $\Delta\Delta G$(vdw) for the binders and nonbinders is only +0.6 kcal/mol (at −4.2 and −4.8 kcal/mol, respectively), favoring the nonbinders. The result suggests that for a ligand to bind at the flap site, it cannot have a large unfavorable electrostatic free energy component. While nonpolar interactions drive binding in general, for fragment binding at the flap site of HIV PR, it appears that it is the ligand−protein polar interaction that separates binders from nonbinders.

To understand the physical origin for the electrostatic component which distinguishes binders from nonbinders, we examine the modeled structures of the binders and nonbinders for clues. Most binders (except for CS6, see discussion earlier) benefit from intermolecular hydrogen bonds between the ligand and the receptor; none of the structures of the binders have polar atoms that are not hydrogen bonded to solute or
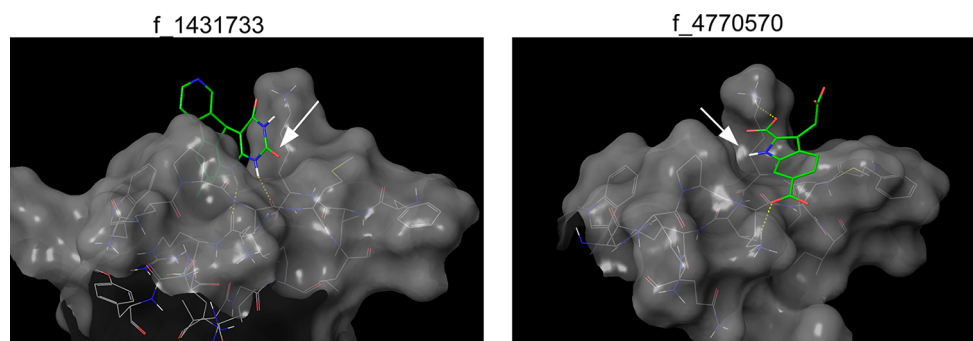
Figure 8. Examples of nonbinders which contain partially buried, unfulfilled polar groups, as indicated by the white arrow in each panel. The ligand−receptor hydrogen bonds are shown in dotted yellow.

solvent. In contrast, the simulated structures of those nonbinders with relatively large unfavorable $\Delta\Delta G$(elec) contain partially buried polar atoms that are not hydrogen bonded with either the receptor or the solvent. Such polar groups suffer from high desolvation penalty. These unfulfilled polar atoms can exist either in the ligand or in the protein. Figure 8 shows examples of the nonbinders with large unfavorable $\Delta\Delta G$(elec). In both cases, in order to form intermolecular hydrogen bonds with the receptor, two polar atoms, an NH group, and/or a carboxylate oxygen are forced to make contact with the nonpolar surface of the receptor.

As seen from Figure 7A, 1F1 and 1F1−N have the most favorable electrostatic contributions to binding among all binders and likely binders. This is attributable to the fact that a carboxylate group in both of them forms a salt bridge with the Arg57 side chain of the protease. Other binders lack such an intermolecular salt bridge. CS6 does not form any hydrogen bonds with the protease, and its binding was driven by nonpolar interactions. Therefore, the favorable electrostatic component of the binding of 1F1 and 1F1−N is due to a direct ligand−receptor hydrogen bonding effect, rather than a desolvation effect.

It is worth noting that the computed nonpolar components $\Delta\Delta G$(vdw) roughly correlate with the docking score [i.e., ligands with large, favorable $\Delta\Delta G$(vdw) tend to have more favorable docking score]. This suggests that while the ligand−receptor nonpolar interaction is well-described by the scoring function for docking, the treatment for the desolvation penalty associated with the unfulfilled polar groups may be inadequate, which has contributed to the high rate of false positives.

Insights into Ligand Optimization. The computed binding free energies of the three crystallographically confirmed binders in this study are rather weak, which appear to explain the lack of significant inhibition against PR from 1F1−N in biochemical assay, even though the binding of 1F1 and 1F1−N were found to preferentially stabilize the closed form of HIV-PR.[50] In this work, both the free energy calculation and the SPR measurements suggest that the likely binder CS6 exhibits a much stronger binding affinity compared to the known binders (Table 1). The simulation predicted binding mode of CS6 revealed a possible structural basis for its stronger binding affinity. In addition to the binding cavity P1 formed by Met42, Lys55, and Pro44, which is filled by all of the binders and also utilized by CS6, there is an adjacent pocket P2, formed by the side chains of Glu35, Lys55, Arg57, and Pro79. In the crystal structures of the three binders (e.g., 1F1 and 1F1−N), the second pocket P2 is largely unoccupied (Figure 9). In the modeled structure of CS6, the P2 pocket is partially occupied
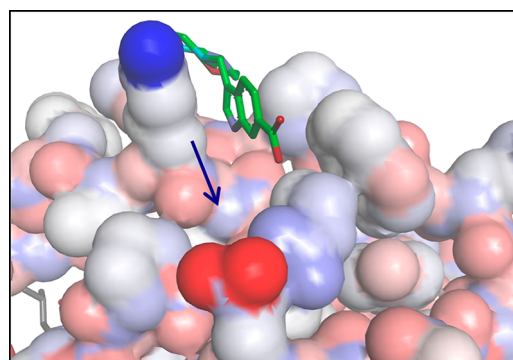

Figure 9. Structure of bound AK2097 in the flap site of PR. The location of second pocket is indicated by the blue arrow.

by a nonpolar aromatic group in the ligand. This is likely the main reason for its exceptionally favorable nonpolar binding free energy component $\Delta\Delta G$(vdw) (see Figure 7). This suggests that for potency enhancement, the future ligand optimization starting from 1F1 and 1F1−N should be directed to further exploit the second pocket P2, while preserving the existing favorable interaction with P1 and the important intermolecular salt bridge with Arg57. Another possible route of optimization is to start from CS6 and add a hydrogen bonding donor group that engages favorably with the Arg57 side chain, as in the case of 1F1 and 1F1−N.

■ CONCLUSION

The main objective of the present study is to evaluate whether absolute binding free energy methods when applied to ligand−protein complexes generated by docking can reduce the false positive rate in docking. We study a set of ligands that dock favorably to a potential allosteric site of HIV-1 protease. Designing potent ligands that bind to this site could stabilize a closed form of the flaps of the enzyme and, when used in combination, could potentially enhance the inhibition activity of an active site ligand. Free energy calculations using the binding energy distribution analysis method (BEDAM) in implicit solvent and the double decoupling method (DDM) in explicit solvent were performed on the 20 three top-ranked protein−ligand complexes taken from AutoDock screening of a library of 2518 compounds to estimate their binding affinities in solution.

The results presented in this work suggest that absolute free energy calculations can eliminate the majority of the false positives from a list of compounds which are top ranked by docking and can also provide important information on the

structural and thermodynamic determinants, separating binders from nonbinders and shed light on how to improve the docking scoring function. The study provides physical insights into ligand optimization against the flap site of PR. Indeed, the ability to discriminate between different thermodynamic components and the desolvation free energy penalty, in particular, is an essential contribution to drive drug design of derivatives and increase efficiency. However, carrying out a large number of absolute binding free energy calculations reliably and rapidly in a semiautomated fashion still remains very challenging, even with the tremendous increase in the availability of computing resources.[55] The automated binding free energy workflow we developed based on the BEDAM method in implicit solvent provides a promising tool in virtually screening up to several hundreds ligands in a practical time frame.[8]

■ AUTHOR INFORMATION

Corresponding Authors
*E-mail: nanjie.deng@gmail.com.
*E-mail: ronlevy@temple.edu.

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30* (16), 2785−2791.

(2) Trott, O.; Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31* (2), 455−461.

(3) Wu, G.; Robertson, D. H.; Brooks, C. L., 3rd; Vieth, M. Detailed analysis of grid-based molecular docking: A case study of CDocker-a Charmm-based MD docking algorithm. *J. Comput. Chem.* **2003**, *24* (13), 1549−1562.

(4) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47* (7), 1739−1749.

(5) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47* (7), 1750−1759.

(6) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein−ligand complexes. *J. Med. Chem.* **2006**, *49* (21), 6177−6196.

(7) Wu, Y.; He, C.; Gao, Y.; He, S.; Liu, Y.; Lai, L. Dynamic modeling of human 5-lipoxygenase-inhibitor interactions helps to discover novel inhibitors. *J. Med. Chem.* **2012**, *55* (6), 2597−2605.

(8) Gallicchio, E.; Deng, N.; He, P.; Wickstrom, L.; Perryman, A. L.; Santiago, D. N.; Forli, S.; Olson, A. J.; Levy, R. M. Virtual screening of integrase inhibitors by large scale binding free energy calculations: The SAMPL4 challenge. *J. Comput.-Aided Mol. Des.* **2014**, *28* (4), 475−490.

(9) Mobley, D. L.; Liu, S.; Lim, N. M.; Wymer, K. L.; Perryman, A. L.; Forli, S.; Deng, N.; Su, J.; Branson, K.; Olson, A. J. Blind prediction of HIV integrase binding from the SAMPL4 challenge. *J. Comput.-Aided Mol. Des.* **2014**, *28* (4), 327−345.

(10) Deng, N. J.; Zhang, P.; Cieplak, P.; Lai, L. Elucidating the energetics of entropically driven protein-ligand association: Calculations of absolute binding free energy and entropy. *J. Phys. Chem. B* **2011**, *115* (41), 11902−11910.

(11) Ferreira, R. S.; Simeonov, A.; Jadhav, A.; Eidam, O.; Mott, B. T.; Keiser, M. J.; McKerrow, J. H.; Maloney, D. J.; Irwin, J. J.; Shoichet, B. K. Complementarity between a docking and a high-throughput screen in discovering new cruzain inhibitors. *J. Med. Chem.* **2010**, *53*, 4891−4905.

(12) Gallicchio, E.; Levy, R. M. Recent theoretical and computational advances for modeling protein-ligand binding affinities. *Adv. Protein. Chem. Struct. Biol.* **2011**, *85*, 27−80.

(13) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical free energy methods for drug discovery: Progress and challenges. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150−160.

(14) Jorgensen, W. L.; Ravimohan, C. Monte Carlo simulation of differences in free energies of hydration. *J. Chem. Phys.* **1985**, *83*, 3050.

(15) Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; Tirado-Rives, J. Efficient computation of absolute free energies of binding by computer simulation. Application to the methane dimer in water. *J. Chem. Phys.* **1988**, *89*, 3742.

(16) Tembe, B. L.; Mc Cammon, J. A. Ligand-receptor interactions. *Comput. Chem.* **1984**, *8*, 81−283.

(17) Lybrand, T. P.; McCammon, J. A.; Wipff, G. Theoretical calculation of relative binding affinity in host-guest systems. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83* (4), 833−835.

(18) Wong, C. F.; McCammon, J. A. Dynamics and design of enzymes and inhibitors. *J. Am. Chem. Soc.* **1986**, *108*, 3830−3832.

(19) Bash, P. A.; Singh, U. C.; Langridge, R.; Kollman, P. A. Free energy calculations by computer simulation. *Science* **1987**, *236* (4801), 564−568.

(20) Rao, S. N.; Singh, U. C.; Bash, P. A.; Kollman, P. A. Free energy perturbation calculations on binding and catalysis after mutating Asn 155 in subtilisin. *Nature* **1987**, *328* (6130), 551−554.

(21) Cieplak, P.; Kollman, P. A. Calculation of the free energy of association of nucleic acid bases in vacuo and water solution. *J. Am. Chem. Soc.* **1988**, *110*, 3734−3739.

(22) Roux, B.; Nina, M.; Pomes, R.; Smith, J. C. Thermodynamic stability of water molecules in the bacteriorhodopsin proton channel: A molecular dynamics free energy perturbation study. *Biophys. J.* **1996**, *71* (2), 670−681.

(23) Kong, X.; Brooks, C. L. Lambda-dynamics: A new approach to free energy calculations. *J. Chem. Phys.* **1996**, *105*, 2414.

(24) Hermans, J.; Wang, L. Inclusion of loss of translational and rotational freedom in theoretical estimates of free energies of binding.

K

dx.doi.org/10.1021/jp506376z | *J. Phys. Chem. B* XXXX, XXX, XXX−XXX

Application to a complex of benzene and mutant t4 lysozyme. *J. Am. Chem. Soc.* **1997**, *119*, 2707−2714.

(25) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The statistical-thermodynamic basis for computation of binding affinities: A critical review. *Biophys. J.* **1997**, *72* (3), 1047−1069.

(26) Guo, Z.; Brooks, C. L.; Kong, X. Efficient and flexible algorithm for free energy calculations using the lambda-dynamics approach. *J. Phys. Chem. B* **1998**, *102*, 2032−2036.

(27) Chang, C. E.; Gilson, M. K. Free energy, entropy, and induced fit in host-guest recognition: Calculations with the second-generation mining minima algorithm. *J. Am. Chem. Soc.* **2004**, *126* (40), 13156−13164.

(28) Fujitani, H.; Tanida, Y.; Ito, M.; Jayachandran, G.; Snow, C. D.; Shirts, M. R.; Sorin, E. J.; Pande, V. S. Direct calculation of the binding free energies of FKBP ligands. *J. Chem. Phys.* **2005**, *123* (8), 084108.

(29) Mobley, D. L.; Chodera, J. D.; Dill, K. A. On the use of orientational restraints and symmetry corrections in alchemical free energy calculations. *J. Chem. Phys.* **2006**, *125* (8), 084902.

(30) Mobley, D. L.; Chodera, J. D.; Dill, K. A. The confine-and-release method: Obtaining correct binding free energies in the presence of protein conformational change. *J. Chem. Theory Comput.* **2007**, *3* (4), 1231−1235.

(31) Jorgensen, W. L.; Thomas, L. L. Perspective on free energy perturbation calculations for chemical equilibria. *J. Chem. Theory Comput.* **2008**, *4* (6), 869−876.

(32) Jorgensen, W. L. Efficient drug lead discovery and optimization. *Acc. Chem. Res.* **2009**, *42* (6), 724−733.

(33) Deng, Y.; Roux, B. Computations of standard binding free energies with molecular dynamics simulations. *J. Phys. Chem. B* **2009**, *113* (8), 2234−2246.

(34) Michel, J.; Essex, J. W. Prediction of protein-ligand binding affinity by free energy simulations: Assumptions, pitfalls and expectations. *J. Comput.-Aided Mol. Des.* **2010**, *24* (8), 639−658.

(35) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* **2006**, *49* (2), 534−553.

(36) Knight, J. L.; Brooks, C. L., 3rd Lambda-dynamics free energy simulation methods. *J. Comput. Chem.* **2009**, *30* (11), 1692−1700.

(37) Huang, Y. M.; Chen, W.; Potter, M. J.; Chang, C. E. Insights from free energy calculations: Protein conformational equilibrium, driving forces, and ligand-binding modes. *Biophys. J.* **2012**, *103* (2), 342−351.

(38) Wang, L.; Deng, Y.; Knight, J. L.; Wu, Y.; Kim, B.; Sherman, W.; Shelley, J. C.; Lin, T.; Abel, R. Modeling local structural rearrangements using fep/rest: Application to relative binding affinity predictions of cdk2 inhibitors. *J. Chem. Theory Comput.* **2013**, *9*, 1282−1293.

(39) Bollini, M.; Domaoal, R. A.; Thakur, V. V.; Gallardo-Macias, R.; Spasov, K. A.; Anderson, K. S.; Jorgensen, W. L. Computationally-guided optimization of a docking hit to yield catechol diethers as potent anti-HIV agents. *J. Med. Chem.* **2011**, *54* (24), 8582−8591.

(40) Buckley, D. L.; Van Molle, I.; Gareiss, P. C.; Tae, H. S.; Michel, J.; Noblin, D. J.; Jorgensen, W. L.; Ciulli, A.; Crews, C. M. Targeting the von Hippel-Lindau E3 ubiquitin ligase using small molecules to disrupt the VHL/HIF-1alpha interaction. *J. Am. Chem. Soc.* **2012**, *134* (10), 4465−4468.

(41) Ravindranathan, K. P.; Mandiyan, V.; Ekkati, A. R.; Bae, J. H.; Schlessinger, J.; Jorgensen, W. L. Discovery of novel fibroblast growth factor receptor 1 kinase inhibitors by structure-based virtual screening. *J. Med. Chem.* **2010**, *53* (4), 1662−1672.

(42) Perryman, A. L.; Santiago, D. N.; Forli, S.; Santos-Martins, D.; Olson, A. J. Virtual screening with AutoDock Vina and the common pharmacophore engine of a low diversity library of fragments and hits against the three allosteric sites of HIV integrase: Participation in the SAMPL4 protein-ligand binding challenge. *J. Comput.-Aided Mol. Des.* **2014**, *28* (4), 429−441.

(43) Wlodawer, A.; Vondrasek, J. Inhibitors of HIV-1 protease: A major success of structure-assisted drug design. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 249−284.

(44) Perryman, A. L.; Lin, J.-H.; McCammon, J. A. HIV-1 protease molecular dynamics of a wild-type and of the v82f/i84v mutant: Possible contributions to drug resistance and a potential new target site for drugs. *Protein Sci.* **2004**, *13*, 1108−1123.

(45) Perryman, A. L.; Lin, J. H.; McCammon, J. A. Restrained molecular dynamics simulations of HIV-1 protease: The first step in validating a new target for drug design. *Biopolymers* **2006**, *82* (3), 272−284.

(46) Deng, N. J.; Zheng, W.; Gallicchio, E.; Levy, R. M. Insights into the dynamics of HIV-1 protease: A kinetic network model constructed from atomistic simulations. *J. Am. Chem. Soc.* **2011**, *133* (24), 9387−9394.

(47) Hornak, V.; Okur, A.; Rizzo, R. C.; Simmerling, C. HIV-1 protease flaps spontaneously close to the correct structure in simulations following manual placement of an inhibitor into the open state. *J. Am. Chem. Soc.* **2006**, *128* (9), 2812−2813.

(48) Fangyu, D.; Layten, M.; Simmerling, C. Solution structure of HIV-1 protease flaps probed by comparison of molecular dynamics simulation ensembles and EPR experiments. *J. Am. Chem. Soc.* **2008**, *130*, 7184−7185.

(49) Perryman, A. L.; Zhang, Q.; Soutter, H. H.; Rosenfeld, R.; McRee, D. E.; Olson, A. J.; Elder, J. E.; Stout, C. D. Fragment-based screen against HIV protease. *Chem. Biol. Drug Des.* **2010**, *75* (3), 257−268.

(50) Tiefenbrunn, T.; Forli, S.; Baksh, M. M.; Chang, M. W.; Happer, M.; Lin, Y. C.; Perryman, A. L.; Rhee, J. K.; Torbett, B. E.; Olson, A. J.; Elder, J. H.; Finn, M. G.; Stout, C. D. Small molecule regulation of protein conformation by binding in the flap of HIV protease. *ACS Chem. Biol.* **2013**, *8* (6), 1223−1231.

(51) Tiefenbrunn, T.; Forli, S.; Happer, M.; Gonzalez, A.; Tsai, Y.; Soltis, M.; Elder, J. H.; Olson, A. J.; Stout, C. D. Crystallographic fragment-based drug discovery: Use of a brominated fragment library targeting HIV protease. *Chem. Biol. Drug Des.* **2014**, *83* (2), 141−148.

(52) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute binding free energies: A quantitative approach for their calculation. *J. Phys. Chem. B* **2003**, *107*, 9535−9551.

(53) Wang, J.; Deng, Y.; Roux, B. Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophys. J.* **2006**, *91* (8), 2798−2814.

(54) Gallicchio, E.; Lapelosa, M.; Levy, R. M. The binding energy distribution analysis method (BEDAM) for the estimation of protein-ligand binding affinities. *J. Chem. Theory Comput.* **2010**, *6* (9), 2961−2977.

(55) Michel, J. Current and emerging opportunities for molecular simulations in structure-based drug design. *Phys. Chem. Chem. Phys.* **2014**, *16*, 4465−4477.

(56) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.

(57) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.

(58) Gallicchio, E.; Paris, K.; Levy, R. M. The AGBNP2 implicit solvation model. *J. Chem. Theory Comput.* **2009**, *5* (9), 2544−2564.

(59) Gallicchio, E.; Levy, R. M. Advances in all atom sampling methods for modeling protein-ligand binding affinities. *Curr. Opin. Struct. Biol.* **2011**, *21* (2), 161−166.

(60) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129*, 124105.

(61) Wickstrom, L.; He, P.; Gallicchio, E.; Levy, R. M. Large scale affinity calculations of cyclodextrin host−guest complexes: Understanding the role of reorganization in the molecular recognition process. *J. Chem. Theory Comput.* **2013**, *9*, 3136−3150.

(62) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926.

L

dx.doi.org/10.1021/jp506376z | *J. Phys. Chem. B* XXXX, XXX, XXX−XXX

(63) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78* (8), 1950−1958.

(64) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25* (9), 1157−1174.

(65) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *23* (16), 1623−1641.

(66) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.

(67) Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29* (7), 845−854.

(68) Deng, N. J.; Cieplak, P. Free energy profile of RNA hairpins: A molecular dynamics simulation study. *Biophys. J.* **2010**, *98* (4), 627−636.

(69) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The weighted histogram analysis method for free energy calculations on biomolecules. 1. The method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.

(70) Rees, D. C.; Congreve, M.; Murray, C. W.; Carr, R. Fragment-based lead discovery. *Nat. Rev. Drug Discovery* **2004**, *3* (8), 660−672.