Published in final edited form as:

Phys Chem Chem Phys. 2018 June 27; 20(25): 17081–17092. doi:10.1039/c8cp01524d.

Comparing Alchemical and Physical Pathway Methods for Computing the Absolute Binding Free Energy of Charged Ligands

Nanjie Deng¹, Di Cui², Bin W. Zhang², Junchao Xia², Jeffrey Cruz¹, and Ronald Levy²

⁽¹⁾Department of Chemistry and Physical Sciences, Pace University, New York, NY 10038

⁽²⁾Center for Biophysics and Computational Biology, Temple University, Philadelphia, PA 19102

Abstract

Accurately predicting absolute binding free energies of protein-ligand complexes is important as a fundamental problem in both computational biophysics and pharmaceutical discovery. Calculating binding free energies for charged ligands is generally considered to be challenging because of the strong electrostatic interactions between the ligand and its environment in the aqueous solution. In this work, we compare the performance of the potential of mean force (PMF) method and double decoupling method (DDM) for computing absolute binding free energies for charged ligands. We first clarify an unresolved issue concerning the use of the binding site volume explicitly to define the complexed state in DDM together with the use of harmonic restraints. We also provide an alternative derivation for the formula for absolute binding free energy using the PMF approach. We use these formulas to compute the binding free energy of charged ligands at an allosteric site of HIV-1 integrase, which has emerged in recent years as a promising target for developing antiviral therapy. As compared with the experimental results, the absolute binding free energies obtained by using the PMF approach show unsigned errors of 1.5–3.4 kcal/mol, which are somewhat better than the results from DDM (unsigned errors of 1.6–4.3 kcal/mol) using the same amount of CPU time. According to the DDM decomposition of the binding free energy, the ligand binding appears to be dominated by nonpolar interactions despite the presence of very large and favorable intermolecular ligand-receptor electrostatic interactions, which are almost completely cancelled out by the equally large free energy cost of desolvation of the charged moiety of the ligands in solution. We discuss the relative strengths of computing absolute binding free energies using the alchemical and physical pathway methods.

INTRODUCTION

The ability to accurately predict protein-ligand binding affinity is important for improving our understanding of molecular recognition and for rational drug design.^{1,2} In recent years, significant progress in methodology, force field and computer speed have been made to

Correspondence to: Nanjie Deng.

nanjie.deng@gmail.com.

SUPPORTING INFORMATION AVAILABLE

Contains two figures on the paths used in the PMF calculation of BI-224436, and a table containing the meaning of the symbols used in the equations in this work.

improve the accuracy of relative binding free energy calculations using the free energy perturbation (FEP)^{3–5} method; for ligands that are chemically not too dissimilar, an accuracy of ~ 1–2 kcal/mol in the relative binding free energy can be expected⁶. However, for ligands that are chemically significantly different and/or bind with very different binding modes, absolute binding free energy methods may be a better approach for computing their binding affinities.

Currently, two methods are primarily used for computing absolute binding free energy in explicit solvent (but also see references that used other approaches^{7–9}), the alchemical double decoupling method DDM^{10–12} and the potential of mean force PMF approach which relies on physical pathways to move the ligand between bound and unbound states.^{13–15} DDM has been applied to a number of protein-ligand complexes in which errors mostly fall in the range 0.5–5 kcal/mol, i.e. somewhat larger than FEP for the relative binding free energy. Compared with the increasing use of DDM in recent years, the PMF method has been employed less widely, with the reported deviations from experiments spanning a wide range of 0.2 kcal/mol to 10 kcal/mol^{13,14,16,17}. However, the PMF approach may have an advantage over DDM for computing binding free energies involving charged ligands.¹³ Such ligands have very large, favorable solvation free energies, which need to be balanced by similarly large and favorable receptor-ligand electrostatic interactions for the binding to occur. The accurate calculation of the two very large, electrostatic decoupling terms can be challenging. In addition, because DDM requires alchemical decoupling the charged ligand from its environment, additional calculations may be necessary to correct the errors caused by the use of a finite sized, periodic solvent box.^{18,19} Such complications are expected to be less important in the PMF approach, where the ligand and receptor remain in the same solvent box throughout the pulling simulation. In addition, PMF may be the only practical method for computing the absolute binding free energy of very large ligands such as those found in the protein-peptide, protein-protein or protein-DNA complexes^{20–22}, where the alchemical decoupling of any of the binding partners would be computationally prohibitive.

In this study, we compare the performance of PMF and DDM in computing the absolute binding free energy of charged ligands with protein in order to understand the important factors that affect the accuracy of the computation in the two methods. The receptor target is an allosteric site in the catalytic core domain dimer interface of HIV-1 Integrase, which has emerged in recent years as a promising drug target for developing antiviral therapy^{23–26}. We also resolve an issue concerning the role of the binding site volume V_{site} in the DDM expression for the absolute binding free energy. The original landmark paper on DDM used a step indicator function to specify a binding site volume V_{site} for the definition of the complexed state.¹⁰ The more recent DDM work discarded the step function and instead use strong harmonic restraints to confine the ligand movement to within a small volume in the binding pocket in order to accelerate sampling^{11,12}; in this approach the effective binding site volume V_{site} is defined implicitly which may cause a problem when treating weakly bound ligand. In addition, it is in principle necessary to explicitly include the V_{site} in the absolute binding free energy expression in order to make direct comparison with experimental binding affinity measurement, even though the numerical value of ΔG_{bind}° is insensitive to the precise definition of V_{site} for reasonably strong binders.^{10,27} We also

include the electrostatics corrections for the effects of the use of finite-size periodic solvent box on the DDM-calculated absolute binding free energies. Lastly we provide an alternative derivation for the formula for computing absolute binding free energy using PMF, compared with the previous formula¹³.

METHODS

A DDM expression that explicitly includes V_{site} in addition to harmonic restraint

In DDM, the absolute binding free energy ΔG_{bind}^o is evaluated by inserting intermediate states into the statistical mechanical expression^{10,11}

$$\Delta G_{bind}^o = -k_B T \ln \frac{V}{V_0} \frac{Z_{RL,N} Z_{0,N}}{Z_{R,N} Z_{L,N}} \quad (1)$$

Here $V = \frac{1}{C}$ is the volume of the simulation system. $V_0 = \frac{1}{C^o} = 1660 \text{ \AA}^3$ is the inverse of the

standard concentration of $C^o = 1 \text{ M}$ solution. $Z_{X,N}$ is the configuration integral

($Z = \int e^{-U(r)/k_B T} dr$) of a single solute X solvated in a box of volume V containing N waters. Thus, $Z_{RL,N}$ represents a system containing one receptor-ligand complex in the bound state solvated by N water molecules. $Z_{0,N}$ represents a system of pure solvent.

To define the receptor-ligand complexed state, the original DDM paper¹⁰ uses a step indicator function $I(\xi)$ acting on the ligand center ξ to specify whether a given configuration is bound or not and this defines an effective binding site volume V_{site}

($V_{site} = \int I(\xi) d\xi$, $I(\xi) = 1$ when ligand is inside and $I(\xi) = 0$ outside the pocket). V_{site} can

be related to the binding signal in an experiment or to the onset of the plateau in the PMF. More recent DDM papers^{11,12} did not explicitly include a geometric criteria to define the receptor-ligand complex; instead, harmonic restraints $U_{restr}(\xi)$ are applied in one or more steps to confine the ligand to within a small volume inside the binding site. This helps accelerate the sampling of configurations. The binding site volume is implicitly defined by the configurational space explored by the fully coupled ligand in the absence of the harmonic restraints. This implicit definition of the complex has since been widely used and works well for reasonably strong binders. However, the binding site volume explored by a bound ligand in unrestrained MD may not be equivalent to the V_{site} that is consistent with experiment. This could make direct comparison between the computed and experimental ΔG_{bind}^o less rigorous. In certain problems such an implicit definition of the bound state can

be problematic: in order to compute the free energy of restraining the ligand-receptor distance using the harmonic restraint $U_{restr}(\xi)$, FEP is typically used, starting with the $\lambda_{restr}=0$ simulation window in which the ligand-receptor distance is unrestrained. For a weakly bound ligand such as those encountered in fragment-based design, the ligand has a finite chance to escape from the binding pocket in the unrestrained simulation in the $\lambda_{restr}=0$ window of the FEP. To avoid such potential complications which arise when the definition of

the bound state is ill defined and to show formally how V_{site} enters into DDM expression, we explicitly define the complexed state using V_{site} by applying a flat-bottomed restraint to implement the step function $I(\xi)$ from the beginning (including $\lambda_{restr} = 0$). Here we provide a modified expression of ΔG_{bind}^o taking into account this explicit definition of the complex.

The configurational integral of the complexed state is

$$Z_{RL,N} = Z_{RL,N}^{V_{site}} = \int I(\xi) e^{-U(r_R, r_L, r_w, \xi)/k_B T} dr_R dr_L dr_w d\xi \quad (2)$$

Where r_R , r_L , and r_w represent the internal coordinates of the receptor atoms, the internal coordinates of the ligand, and the solvent degrees of freedom, respectively. ξ is the position of the ligand center relative to the receptor binding site. Accordingly, Eq. (1) can be evaluated by inserting intermediate states

$$\Delta G_{bind}^o = -k_B T \ln \frac{V}{V_0} \frac{Z_{RL,N}^{V_{site}} Z_{0,N}}{Z_{R,N} Z_{L,N}} = -k_B T \ln \left(\frac{V}{V_0} \frac{Z_{RL,N}^{V_{site}}}{Z_{R \dots L,N}^{V_{site}}} \frac{Z_{R \dots L,N}^{V_{site}}}{Z_{R,N \dots L_{gas}}^{V_{site}}} \frac{Z_{R,N \dots L_{gas}}^{V_{site}}}{Z_{R,N} Z_{L_{gas}}^{V_{site}}} \frac{Z_{L_{gas}}^{V_{site}}}{Z_{L,N}} \frac{Z_{0,N}}{Z_{L,N}} \right) \quad (3)$$

Here, the superscript V_{site} is used to emphasize that the ligand center is confined to within the binding site volume by the flat-bottomed potential which implements $I(\xi)$. $Z_{R \dots L,N}^{V_{site}}$ is the configuration integral of a solvated ligand-receptor complex in the fully coupled state, in which the ligand is harmonically restrained by $U_{restr}(\xi)$, in addition to the step function $I(\xi)$. $Z_{R,N \dots L_{gas}}^{V_{site}}$ is the configuration integral of a bound complex in which the ligand is restrained by both $I(\xi)$ and $U_{restr}(\xi)$ but its nonbonded interactions with its environment, i.e. receptor plus solvent, are turned off. $Z_{R,N+L_{gas}}^{V_{site}}$ represents a bound complex in which the ligand is not interacting with its environment but is confined to the receptor binding site V_{site} only by the step function $I(\xi)$. $Z_{L_{gas}}$ describes an ideal-gas phase ligand and $Z_{0,N}$ represents N pure solvent molecules.

Thus, Eq. (3) allows ΔG_{bind}^o to be computed as

$$\Delta G_{bind}^o = -\Delta G_{restr}^{bound} - \Delta G_{decoupl}^{bound} + \Delta G_{restr}^{gas} + \Delta G_{decoupl}^{bulk} \quad (4)$$

Each term in Eq. (4) can be computed from simulation or analytically:

$$-\Delta G_{restr}^{bound} = -k_B T \ln \frac{Z_{RL,N}^{V_{site}}}{Z_{R \cdots L,N}^{V_{site}}} \quad (4.1)$$

$$-\Delta G_{decoupl}^{bound} = -k_B T \ln \frac{Z_{R \cdots L,N}^{V_{site}}}{Z_{R,N \cdots L_{gas}}^{V_{site}}} \quad (4.2)$$

$$\Delta G_{restr}^{gas} = -k_B T \ln \frac{V}{V_0} \frac{Z_{R,N \cdots L_{gas}}^{V_{site}}}{Z_{R,N+L_{gas}}^{V_{site}}} \frac{Z_{R,N+L_{gas}}^{V_{site}}}{Z_{R,N} Z_{L_{gas}}^{V_{site}}} \quad (4.3)$$

$$\Delta G_{decoupl}^{bulk} = -k_B T \ln \frac{Z_{L_{gas}}^{Z_{0,N}}}{Z_{L,N}} \quad (4.4)$$

In Eq. (4.1), ΔG_{restr}^{bound} is the free energy of switching on the ligand-receptor harmonic distance restraint when the ligand is fully coupled to its environment (receptor and solvent). This term can be readily computed by using FEP (or thermodynamic integration, TI) in a series of λ_{restr} . Since the magnitude of this term is often small, it may be obtained from a single simulation of the bound complex (i.e. $\lambda_{restr}=0$) using the Zwanzig formula³,

$$\Delta G_{restr}^{bound} = k_B T \ln \left\langle e^{\frac{U_{restr}(\xi)/k_B T}{U_{restr}}} \right\rangle_{U_{restr}}. \text{ In Eq. (4.2), } \Delta G_{decoupl}^{bound} \text{ is the free energy of turning off}$$

the nonbonded interactions between the ligand and its environment, with the ligand remaining in the binding site because of the ligand-receptor harmonic distance restraint; this term can be computed using FEP or TI by simulating the decoupling process. In principle, both calculations of $-\Delta G_{restr}^{bound}$ and $-\Delta G_{decoupl}^{bound}$ are done in the presence of the flat-bottomed restraint $\mathcal{I}(\xi)$ on the ligand center ξ during the FEP or TI simulations. In Eq. (4.3), ΔG_{restr}^{gas} is the free energy of switching on the ligand-receptor harmonic distance restraint $U_{restr}(\xi)$ and the flat-bottomed restraint $\mathcal{I}(\xi)$ that force the gas-phase ligand to remain inside the receptor binding site; this term is evaluated analytically as

$$\Delta G_{restr}^{gas} = -k_B T \ln \frac{V}{V_0} \frac{Z_{R,N \dots L}^{V_{site}}}{Z_{R,N+L}^{V_{site}}} \frac{Z_{R,N+L}^{gas}}{Z_{R,N} Z_{L}^{gas}} - k_B T \ln \frac{V}{V_0} \quad (5)$$

$$-k_B T \ln \frac{\int_{\xi \in V_{site}} e^{-U_{restr}(\xi)/k_B T} d\xi}{V_{site}} - k_B T \ln \frac{V_{site}}{V} =$$

$$-k_B T \ln \frac{\int_{\xi \in V_{site}} e^{-U_{restr}(\xi)/k_B T} d\xi}{V_0}$$

since it is straightforward to show that $\frac{Z_{R,N \dots L}^{V_{site}}}{Z_{R,N+L}^{V_{site}}} = \frac{\int_{\xi \in V_{site}} e^{-U_{restr}(\xi)/k_B T} d\xi}{V_{site}}$ and

$$\frac{Z_{R,N+L}^{V_{site}}}{Z_{R,N} Z_{L}^{gas}} = \frac{V_{site}}{V}.$$

Note that the system volume V drops out from the final expression of ΔG_{bind}^o , as it should be,

since in general ΔG_{bind}^o can only depend on the standard state concentration $C^o = 1/V_0$ and

V_{site} , in addition to the non-ideal terms calculated by Eq. (4.1), (4.2) and (4.4). It can be shown that when the force constant k_r in $U_{restr}(\xi)$ is sufficiently strong, or when V_{site} is

sufficiently large, i.e. $V_{site}^{1/3} \gg \left(\frac{2k_B T}{k_r}\right)^{1/2}$, we can write

$$\int_{\xi \in V_{site}} e^{-U_{restr}(\xi)/k_B T} d\xi \approx 4\pi \int_0^\infty \Delta r^2 e^{-\frac{1}{2}k_r \Delta r^2/k_B T} d\Delta r = \left(\frac{2\pi k_B T}{k_r}\right)^{3/2} \quad (6)$$

which can be considered the effective binding site volume explored by the harmonically restrained gas-phase ligand. Note that except for very weak binders, when V_{site} or k_r is

properly chosen, ΔG_{bind}° becomes independent of the exact values for these parameters.
10,11,27

Lastly, $-\Delta G_{decoupl}^{bulk}$ in Eq. (4.4) is the free energy of inserting the gas-phase ligand into the solvent. This term is the ligand solvation free energy, which can be calculated using TI.

For the ligands studied in this work, we did not apply the flat-bottom restraint $\mathcal{K}(\xi)$ in either DDM or PMF simulations. This is because all the ligands remain bound on the MD timescales of tens of nanoseconds during the $\lambda_{restr} = 0$ (coupled and unrestrained) simulation. Therefore, applying the flat-bottom restraint $\mathcal{K}(\xi)$ will not change the terms computed in Eq. (4.1) and (4.2). The only effect of $\mathcal{K}(\xi)$ on the binding free energy is from ΔG_{restr}^{gas} in Eq. (4.3), which is calculated analytically in Eq. (5) and Eq. (6). The magnitude of the change due to $\mathcal{K}(\xi)$ is negligibly small unless the V_{site} is chosen to be unreasonably small. For example, for KF134, suppose the upper limit of the integral in Eq. (6), which corresponds to the boundary of V_{site} , is chosen to be just 2 Å beyond the minimum in the PMF curve (Fig. 2). With such a restrictive V_{site} , the corresponding change in the free energy due to the flat-bottom restraint $\mathcal{K}(\xi)$ is still only 0.06 kcal/mol.

An alternative derivation for the expression for the absolute binding free energy using the PMF approach

To derive an expression of ΔG_{bind}° in the PMF approach,¹³ we also start from Eq. (1).^{10,11,28} The denominator of the logarithm in the right-hand side of Eq. (1) is the product of two configurational integrals, $Z_{R,N}$ and $Z_{L,N}$, which represent a box V with receptor R and N solvent molecules, and the a box V with ligand L and N solvent molecules, respectively. When the solvent box V is sufficiently large, we can transfer L from its own solvent box to the solvent box containing R without affecting the product of the two configurational integrals. After the transfer, R and L are in one solvent box V, and remain dissociated. The original solvent box V where L is transferred from now contains only N solvent molecules. This transformation is equivalent to rewriting the denominator in the logarithm of the right-hand side of Eq. (1) as

$$\Delta G_{bind}^{\circ} = -k_B T \ln \frac{V}{V_0} \frac{Z_{RL,N} Z_{0,N}}{Z_{R,N} Z_{L,N}} = -k_B T \ln \frac{V}{V_0} \frac{Z_{RL,N} Z_{0,N}}{Z_{R+L,N} Z_{0,N}} \quad (7)$$

Here $Z_{R+L,N}$ represents the configuration integral of the box V now containing receptor R, the unbound L and N solvent molecules. Now the two $Z_{0,N}$ in the numerator and the denominator cancel and we obtain

$$\Delta G_{bind}^{\circ} = -k_B T \ln \frac{V}{V_0} \frac{z_{RL,N} z_{0,N}}{z_{R+L,N} z_{0,N}} = -k_B T \ln \frac{V}{V_0} \frac{z_{RL,N}}{z_{R+L,N}} \quad (8)$$

To evaluate the right-hand side of Eq. (8) we insert intermediate states

$$\Delta G_{\text{bind}}^{\circ} = -k_B T \ln \frac{V}{V_0} \frac{Z_{RL,N}}{Z_{R+L,N}} = -k_B T \ln \frac{V}{V_0} \frac{Z_{RL,N}}{Z_{RL(\theta,\phi,\Theta,\Phi,\Psi),N}} \frac{Z_{RL(\theta,\phi,\Theta,\Phi,\Psi),N}}{Z_{R+L(r^*,\theta,\phi,\Theta,\Phi,\Psi),N}} \frac{Z_{R+L(r^*,\theta,\phi,\Theta,\Phi,\Psi),N}}{Z_{R+L,N}} \quad (9)$$

Here $Z_{RL(\theta,\phi,\Theta,\Phi,\Psi),N}$ represents a complex RL in which ligand L's external degrees of freedom, which are defined by the polar angles (θ, ϕ) and three Euler angles (Θ, Φ, Ψ) , are restrained to their equilibrium values in the bound state by a set of harmonic restraints:

$$U_{\theta} = \frac{1}{2}k_{\theta}(\theta - \theta_0)^2, U_{\phi} = \frac{1}{2}k_{\phi}(\phi - \phi_0)^2, U_{\Theta} = \frac{1}{2}k_{\Theta}(\Theta - \Theta_0)^2, U_{\Phi} = \frac{1}{2}k_{\Phi}(\Phi - \Phi_0)^2, \text{ and } U_{\Psi} = \frac{1}{2}k_{\Psi}(\Psi - \Psi_0)^2$$

gives the definition of the polar angles (θ, ϕ) and three Euler angles (Θ, Φ, Ψ) which specify the ligand orientation relative to the receptor. $Z_{R+L(r^*,\theta,\phi,\Theta,\Phi,\Psi),N}$ represents a system in which the ligand L is not only subject to the polar and orientational restraints $(U_{\theta}, U_{\phi}, U_{\Theta}, U_{\Phi}, U_{\Psi})$, but is also subject to the harmonic restraint $U_{r^*} = \frac{1}{2}k_r(r - r^*)^2$, which forces the ligand atom A to be close to a bulk location r^* .

The different terms inside the logarithm of the right-hand side of Eq. (9) correspond to the following thermodynamic transformations:

The first term inside the logarithm in the right-hand side of Eq. (9) $\frac{Z_{RL,N}}{Z_{RL(\theta,\phi,\Theta,\Phi,\Psi),N}}$ corresponds to the free energy of switching on the angular restraints $(U_{\theta}, U_{\phi}, U_{\Theta}, U_{\Phi}, U_{\Psi})$ on the ligand in the bound state, i.e.

$$\frac{z_{RL,N}}{z_{RL(\theta,\phi,\Theta,\Phi,\Psi),N}} = e^{\Delta G_{(\theta,\phi,\Theta,\Phi,\Psi)}^{\text{bound}}/k_B T} \quad (10.1)$$

Here $\Delta G_{(\theta,\phi,\Theta,\Phi,\Psi)}^{\text{bound}}$ is the free energy change of slowly switching on the polar angular and orientational restraints $(U_{\theta}, U_{\phi}, U_{\Theta}, U_{\Phi}, U_{\Psi})$. It is computed in this work using 24 λ windows: 12 λ windows ($\lambda = 0.0, 0.01, 0.025, 0.05, 0.075, 0.1, 0.2, 0.35, 0.5, 0.65, 0.8, 1.0$) are used to switch on the polar angle restraints (U_{θ}, U_{ϕ}) , followed by another 12 λ which switch on the ligand orientational restraints $(U_{\Theta}, U_{\Phi}, U_{\Psi})$. In this work the free energy change in Eq. (10.1) is computed by analyzing the data obtained from FEP simulations using the Bennett acceptance ratio (BAR) method²⁹ (FEP or TI can also be used). The definition of V_{site} is implicit in Eq. (10.1) in the same way as it is for the DDM calculation; namely it is defined by the volume accessible to the ligands in the binding site of the receptor as the set of five angular restraints $(U_{\theta}, U_{\phi}, U_{\Theta}, U_{\Phi}, U_{\Psi})$ are turned on.

The second term inside the logarithm of the right-hand side of Eq. (9) is

$\frac{Z_{RL}(\theta, \phi, \Theta, \Phi, \Psi), N}{Z_{R+L}(r^*, \theta, \phi, \Theta, \Phi, \Psi), N}$. In the system represented by the denominator $Z_{R+L}(r^*, \theta, \phi, \Theta, \Phi, \Psi), N$, L is held at r^* in the bulk solution by the distance restraint $U_{r^*} = \frac{1}{2}k_r(r - r^*)^2$. In addition, the orientation of L relative to R is held at near the equilibrium value by the angular restraints $(U_\theta, U_\phi, U_\Theta, U_\Phi, U_\Psi)$. Using the 1D potential of mean force (PMF) $w(r)$ along the axis r_{aA} , we can compute the ratio $\frac{Z_{RL}(\theta, \phi, \Theta, \Phi, \Psi), N}{Z_{R+L}(r^*, \theta, \phi, \Theta, \Phi, \Psi), N}$.

Let $w(r)$ be the 1D PMF as a function of $r = |r_{aA}|$. The PMF function $w(r)$, which encompasses both the complexed and the unbound state, can be computed using umbrella sampling simulation on the ligand in the presence of the angular restraints $(U_\theta, U_\phi, U_\Theta, U_\Phi, U_\Psi)$. The PMF $w(r)$ is defined as

$$e^{-w(r)/kT} = \int dr_R dr_L dr_w \delta(r_{aA} - r) e^{-H(r_R, r_L, r_w)/k_B T} \quad (10.2.1)$$

Here, the $H(r_R, r_L, r_w)$ includes both the normal interactions involving the receptor, ligand and solvent molecules, and the angular restraints $(U_\theta, U_\phi, U_\Theta, U_\Phi, U_\Psi)$. Note that the numerator in $\frac{Z_{RL}(\theta, \phi, \Theta, \Phi, \Psi), N}{Z_{R+L}(r^*, \theta, \phi, \Theta, \Phi, \Psi), N}$ is

$$Z_{RL}(\theta, \phi, \Theta, \Phi, \Psi), N = \int_{r_{aA} \in V_{site}} dr_R dr_L dr_w e^{-H(r_R, r_L, r_w)/k_B T} \quad (10.2.2)$$

which can be written as

$$\begin{aligned} Z_{RL}(\theta, \phi, \Theta, \Phi, \Psi), N &= \int_{r_{aA} \in V_{site}} dr_R dr_L dr_w e^{-\frac{H(r_R, r_L, r_w)}{k_B T}} \\ &= \int_{r \in V_{site}} dr \int dr_R dr_L dr_w \delta(r_{aA} - r) e^{-\frac{H(r_R, r_L, r_w)}{k_B T}} \end{aligned} \quad (10.2.3)$$

Substituting Eq. (10.2.3) into Eq. (10.2.1) yields

$$Z_{RL}(\theta, \phi, \Theta, \Phi, \Psi), N = \int_{r \in V_{site}} dr e^{-w(r)/k_B T} \quad (10.2.4)$$

We now simplify the denominator of $\frac{z_{RL}(\theta, \phi, \Theta, \Phi, \Psi), N}{z_{R+L}(r^*, \theta, \phi, \Theta, \Phi, \Psi), N}$, which is

$$Z_{R+L}(r^*, \theta, \phi, \Theta, \Phi, \Psi), N = \int dr_R dr_L dr_W e^{-H(r_R, r_L, r_W)/k_B T} e^{-U(r_{aA} - r^*)/k_B T} \quad (10.2.5)$$

For every $r_{aA} = r$, the corresponding contribution to the integral in the right-hand side of Eq. (10.2.5) is $e^{-U(r - r^*)/k_B T} \int dr_R dr_L dr_W \delta(r_{aA} - r) e^{-H(r_R, r_L, r_W)/k_B T}$; therefore, Eq. (10.2.5) can be written as

$$Z_{R+L}(r^*, \theta, \phi, \Theta, \Phi, \Psi), N = \int dr_R dr_L dr_W e^{\frac{H(r_R, r_L, r_W)}{k_B T}} e^{-\frac{U(r_{aA} - r^*)}{k_B T}} = \int dr e^{-\frac{U(r - r^*)}{k_B T}} \int dr_R dr_L dr_W \delta(r_{aA} - r) e^{-\frac{H(r_R, r_L, r_W)}{k_B T}} \quad (10.2.6)$$

Substituting Eq. (10.2.6) into Eq. (10.2.1), we obtain

$$Z_{R+L}(r^*, \theta, \phi, \Theta, \Phi, \Psi), N = \int dr e^{-\frac{U(r - r^*)}{k_B T}} \int dr_R dr_L dr_W \delta(r_{aA} - r) e^{-\frac{H(r_R, r_L, r_W)}{k_B T}} \quad (10.2.7)$$

$$= \int dr e^{-U(r - r^*)/k_B T} e^{-w(r)/k_B T}$$

Combine Eq. (10.2.4) and Eq. (10.2.7) yields

$$\begin{aligned}
 \frac{z_{RL}(\theta, \phi, \Theta, \Phi, \Psi), N}{z_{R+L}(r^*, \theta, \phi, \Theta, \Phi, \Psi), N} &= \frac{\int_{bound} e^{-w(r)/k_B T} dr}{\int_0^\infty e^{-w(r)/k_B T} e^{-U_{r^*}/k_B T} dr} \quad (10.2a) \\
 &= \frac{\int_{r_{min}}^{r_{site}} e^{-w(r)/k_B T} dr}{\int_0^\infty e^{-w(r)/k_B T} e^{-\frac{1}{2}k_r(r-r^*)^2/k_B T} dr}
 \end{aligned}$$

Here the line integral $\int_{bound} e^{-w(r)/k_B T} dr$ spans the range of r that corresponds to the bound state. Therefore, the upper limit of the integration equals r_{site} , where the ligand-receptor interaction becomes negligibly small (see Fig. 2, right panel). The lower limit of the integration corresponds to the smallest distance recorded in the PMF (Fig. 2), denoted as r_{min} .

To simplify the right-hand side of Eq. 10.2a, note that in the denominator, $w(r)$ can be taken outside the integral $\int_0^\infty e^{-w(r)/k_B T} e^{-\frac{1}{2}k_r(r-r^*)^2/k_B T} dr$ as a constant. This is because the

integral contains the Gaussian function $e^{-\frac{1}{2}k_r(r-r^*)^2/k_B T}$ centered around r^* , when the force constant k_r is sufficiently large, the Gaussian function acts like a δ -function, i.e. sharply peaked within a small region around r^* and vanishingly small everywhere else. Therefore, in this very small neighborhood of r^* , $w(r)$ is essentially constant with $w(r) \approx w(r^*)$, and can be taken outside the integral as a constant in Eq. 10.2a.

Thus, Eq. (10.2a) becomes

$$\begin{aligned}
\frac{Z_{RL}(\theta, \phi, \Theta, \Phi, \Psi), N}{Z_{R+L}(r^*, \theta, \phi, \Theta, \Phi, \Psi), N} &= \frac{\int_{bound} e^{-w(r)/k_B T} dr}{\int_0^\infty e^{-w(r)/k_B T} e^{-\frac{1}{2}k_r(r-r^*)^2/k_B T} dr} \\
&= \frac{\int_{bound} e^{-w(r)/k_B T} dr}{e^{-w(r^*)/k_B T} \int_{-\infty}^\infty e^{-\frac{1}{2}k_r(r-r^*)^2/k_B T} d(r-r^*)} = \frac{\int_{bound} e^{-[w(r)-w(r^*)]/k_B T} dr}{\int_{-\infty}^\infty e^{-\frac{1}{2}k_r(\Delta r)^2/k_B T} d\Delta r} \\
&= \frac{\int_{bound} e^{-[w(r)-w(r^*)]/k_B T} dr}{\left(\frac{2\pi k_B T}{k_r}\right)^{\frac{1}{2}}}
\end{aligned} \tag{10.2}$$

Finally, the last term inside the logarithm of the right hand side of Eq. (9),

$\frac{Z_{R+L}(r^*, \theta, \phi, \Theta, \Phi, \Psi), N}{Z_{R+L, N}}$, represents the free energy of removing all the distance and angular restraints $(U_{r^*}, U_\theta, U_\phi, U_\Theta, U_\Phi, U_\Psi)$ on the bulk ligand such that the ligand is allowed to explore the system volume V and rotate freely. When the force constants used in the harmonic restraints are sufficiently strong, which is the case in this work, the rigid-rotor approximation applies, allowing this term to be evaluated analytically as¹¹

$$\frac{Z_{R+L}(r^*, \theta, \phi, \Theta, \Phi, \Psi), N}{Z_{R+L, N}} \approx \frac{r^{*2} \sin \theta_0 \sin \Theta_0 (2\pi k_B T)^3}{8\pi^2 V (k_r k_\theta k_\phi k_\Theta k_\Phi k_\Psi)^{\frac{1}{2}}} \tag{10.3}$$

Where θ_0 and Θ_0 are the equilibrium values of θ and Θ , respectively. $k_r k_\theta k_\phi k_\Theta k_\Phi k_\Psi$ in the denominator is the product of the force constants for the corresponding harmonic restraints on $(r, \theta, \phi, \Theta, \Phi, \Psi)$.

Combining the above Eq. (10.1), (10.2), (10.3) and Eq. (9), we obtain

$$\Delta G_{\text{bind}}^{\circ} = -\Delta G_{(\theta, \phi, \theta, \phi, \psi)}^{\text{bound}} - k_B T \ln \frac{\int_{\text{bound}} e^{-\frac{[w(r) - w(r^*)]}{k_B T}} dr}{\left(\frac{2\pi k_B T}{k_r}\right)^{\frac{1}{2}}} \quad (11)$$

$$-k_B T \ln \frac{C^{\circ} r^{*2} \sin \theta_0 \sin \Theta_0 (2\pi k_B T)^3}{8\pi^2 (k_r k_{\theta} k_{\phi} k_{\Theta} k_{\psi})^{\frac{1}{2}}} = -\Delta G_{(\theta, \phi, \theta, \phi, \psi)}^{\text{bound}} - k_B T \ln \frac{\int_{\text{bound}} e^{-\frac{w(r)}{k_B T}} dr}{\left(\frac{2\pi k_B T}{k_r}\right)^{\frac{1}{2}}}$$

$$-k_B T \ln \frac{C^{\circ} r^{*2} \sin \theta_0 \sin \Theta_0 (2\pi k_B T)^3}{8\pi^2 (k_r k_{\theta} k_{\phi} k_{\Theta} k_{\psi})^{\frac{1}{2}}} = -\Delta G_{(\theta, \phi, \theta, \phi, \psi)} - w(r^*)$$

$$-k_B T \ln \frac{\int_{\text{bound}} e^{-w(r)/k_B T} dr}{\left(\frac{2\pi k_B T}{k_r}\right)^{\frac{1}{2}}} + \Delta G_{\text{restr}}^{\text{bulk}}$$

$$\text{where } \Delta G_{\text{restr}}^{\text{bulk}} = -k_B T \ln \frac{C^{\circ} r^{*2} \sin \theta_0 \sin \Theta_0 (2\pi k_B T)^3}{8\pi^2 (k_r k_{\theta} k_{\phi} k_{\Theta} k_{\psi})^{\frac{1}{2}}} \text{ (the quantity in the logarithm is}$$

dimensionless).

Note that although the right-hand side of Eq. (11) contains r^* , the $\Delta G_{\text{bind}}^{\circ}$ calculated using Eq. 11 is actually independent of the choice of the bulk location r^* . To see this we note that the $w(r)$ implicitly contains the Jacobian factor of the spherical coordinates that is proportional to $\ln^2$, because the polar angle restraints (U_{θ} , U_{ϕ}) force the ligand to move along the axis r_{aA} , inside a cone with a cross-section area

$= r^2 \sin \theta_0 \langle \Delta \theta \rangle \langle \Delta \phi \rangle$. $\langle \Delta \theta \rangle$ and $\langle \Delta \phi \rangle$ are the root-mean-square fluctuation in θ and ϕ , which are related to the angular force constants k_{θ} , k_{ϕ} and the temperature factor by $\langle \Delta \theta \rangle^2 = \frac{k_B T}{k_{\theta}}$

and $\langle \Delta \phi \rangle^2 = \frac{k_B T}{k_{\phi}}$, respectively. Thus, the 1D probability density $P(r)$ is a product of the surface area $r^2 \sin \theta_0 \langle \Delta \theta \rangle \langle \Delta \phi \rangle$ and a term $P'(r)$ that depends only on intermolecular forces i.e. $P(r) = r^2 \sin \theta_0 \langle \Delta \theta \rangle \langle \Delta \phi \rangle P'(r)$, and $w(r) \propto -k_B T \ln[P(r)] = -k_B T \ln[r^2 \sin \theta_0 \langle \Delta \theta \rangle \langle \Delta \phi \rangle P'(r)]$. In the bulk solution, $P'(r)$ is isotropic and therefore a constant, so $w(r)$ varies with $\ln^2$ in bulk

solution. Now the $\ln r^{*2}$ -dependence in $w(r^*)$ and the $\ln r^{*2}$ -dependence in ΔG_{restr}^{bulk} cancel out each other, which makes the ΔG_{bind}^o calculated using Eq. (11) independent of the choice of the bulk location r^* .

The explicit geometrical definition of the complex can also be used in a similar way in the PMF calculation of ΔG_{bind}^o . In this case, Eq. (11) does not need to be modified, since the information of the binding site volume is already used in choosing the upper limit of the PMF integral $\int_{bound} e^{-w(r)/k_B T} dr$ in the right hand side of the Eq. (11): see Eq. (10.2). The only change is that the flat-bottomed restraint on the ligand center should be present when running the FEP (or TI) simulations to compute $\Delta G_{(\theta, \phi, \Theta, \Phi, \Psi)}^{bound}$, the free energy of turning on the angular restraints $(U_\theta, U_\phi, U_\Theta, U_\Phi, U_\Psi)$ on the ligand external degrees of freedom in the bound state: see Eq. (10.1).

Comparison of our PMF formula with that derived previously by Woo and Roux

Using a δ -function formulation, Woo and Roux have derived the expression for the absolute binding free energy ΔG_{bind}^o as¹³

$$\Delta G_{bind}^o = -k_B T \ln C^o S^* I^* + \Delta G_o^{bulk} - \Delta G_o^{site} - \Delta G_a^{site} \quad (12)$$

where

$$S^* = (r^*)^2 \int_0^\pi \sin \theta d\theta \int_0^{2\pi} d\phi e^{-u_a(\theta, \phi)/k_B T} \quad (12.1)$$

$$I^* = \int_{bound} dr e^{-[w(r) - w(r^*)]/k_B T} \quad (12.2)$$

$$\Delta G_o^{bulk} = -k_B T \ln \frac{1}{8\pi^2} \int_0^\pi d\Theta \int_{-\pi}^\pi d\Phi \int_{-\pi}^\pi d\Psi \sin \Theta e^{-\frac{u_o(\Theta, \Phi, \Psi)}{k_B T}} \quad (12.3)$$

In Eq. (12), ΔG_o^{site} and ΔG_a^{site} are the free energy of applying the harmonic restraints on (θ, ϕ) and (Θ, Φ, Ψ) for the ligand in the binding site, respectively. The $u_a(\theta, \phi)$ and $u_o(\Theta, \Phi, \Psi)$ in the above equations are the harmonic restraints on (θ, ϕ) and (Θ, Φ, Ψ) , respectively.

Comparing Eq. (11) with Eq. (12), it can be shown that the two expressions yields equivalent results when all the six harmonic restraint force constants are sufficiently large.

First, we recognize that the first term on the right hand side of Eq. (11),

$-\Delta G_{(U_\theta, U_\phi, U_\Theta, U_\Phi, U_\Psi)}^{bound}$ is the same as the last two terms of the Eq. (12), $-\Delta G_o^{site} - \Delta G_a^{site}$,

therefore these terms cancel.

Next, we compare the remaining terms in the right-hand side of Eq. (11)

$$-k_B T \ln \frac{\int_{bound} e^{-[w(r) - w(r^*)]/k_B T} dr}{\left(\frac{2\pi k_B T}{k_r}\right)^{\frac{1}{2}}} - k_B T \ln \frac{C^o r^{*2} \sin \theta_0 \sin \Theta_0 (2\pi k_B T)^3}{8\pi^2 (k_r k_\theta k_\phi k_\Theta k_\Phi k_\Psi)^{\frac{1}{2}}} \quad (13)$$

with the remaining terms in the right-hand side of Eq. (12)

$$\begin{aligned} & -k_B T \ln [C^o (r^*)^2 \int_0^\pi \sin \theta d\theta \int_0^{2\pi} d\phi e^{-u_a/k_B T} \int_{bound} dr e^{-[w(r) - w(r^*)]/k_B T}] \quad (14) \\ & -k_B T \ln \frac{1}{8\pi^2} \int_0^\pi d\Theta \int_{-\pi}^\pi d\Phi \int_{-\pi}^\pi d\Psi \sin \Theta e^{-\frac{u_o(\Theta, \Phi, \Psi)}{k_B T}} \end{aligned}$$

Note that $u_a(\theta, \phi) = U_\theta + U_\phi = \frac{1}{2}k_\theta(\theta - \theta_0)^2 + \frac{1}{2}k_\phi(\phi - \phi_0)^2$ and

$u_o(\Theta, \Phi, \Psi) = U_\Theta + U_\Phi + U_\Psi = \frac{1}{2}k_\Theta(\Theta - \Theta_0)^2 + \frac{1}{2}k_\Phi(\Phi - \Phi_0)^2 + \frac{1}{2}k_\Psi(\Psi - \Psi_0)^2$. When the six harmonic restraint force constants are sufficiently large, the Gaussian functions in (14) are of the form $\int d\xi e^{-k_\xi(\xi - \xi_0)^2/k_B T} = \left(\frac{2\pi k_B T}{k_\xi}\right)^{1/2}$. Therefore, the two terms in (14) can be written as

$$-k_B T \ln \frac{1}{8\pi^2} \int_0^\pi d\Theta \int_{-\pi}^\pi d\Phi \int_{-\pi}^\pi d\Psi \sin \Theta e^{-\frac{u_o(\Theta, \Phi, \Psi)}{k_B T}} \approx -k_B T \ln \frac{1}{8\pi^2} \sin \Theta_0 \frac{(2\pi k_B T)^{\frac{3}{2}}}{(k_\Theta k_\Phi k_\Psi)^{\frac{1}{2}}} \quad (14.1)$$

and

$$\begin{aligned}
 & -k_B T \ln \left\{ C^o(r^*)^2 \int_0^\pi \sin\theta d\theta \int_0^{2\pi} d\phi e^{-u_a/k_B T} \int_{bound} dre^{-[w(r)-w(r^*)]/k_B T} \right\} \approx \quad (14.2) \\
 & -k_B T \ln \left\{ C^o(r^*)^2 \sin\theta_0 \frac{2\pi k_B T}{(k_\theta k_\phi)^{\frac{1}{2}}} \int_{bound} dre^{-[w(r)-w(r^*)]/k_B T} \right\}
 \end{aligned}$$

Comparing the right-hand sides of Eq. (14.1) and (14.2) with the (13), we can see that (13) and (14) are identical. Therefore, Eq. (11) is equivalent to Eq. (12) when the force constants are sufficiently large.

PMF calculation setup

We compute the ligand 1D PMF $w(r)$ using umbrella sampling simulations in explicit solvent³⁰, in which a series of MD simulations is performed with the harmonic distance restraint on the receptor atom a and ligand atom A : see Fig. 1. The biasing potential in the i -th simulation window is $U_{r,i} = \frac{1}{2}k_r(r - r_i^0)^2$, where r_i^0 is the reference distance for the i -th sampling window. The full range of the distance space is covered using between 20 and 24 umbrella windows for the different complexes. A single force constant $k_r = 1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ is used for the distance restraint in all the sampling windows. The force constants used in the angular restraints are: $k_\theta = k_\phi = k_\Theta = k_\Phi = k_\Psi = 1000 \text{ kJ rad}^{-1} \text{ mol}^{-1}$. In each umbrella window, a 30-ns MD simulation is performed, starting from the last simulation snapshot of the previous window. The first 20 ns are treated as equilibration and the last 10 ns of sampling data are used for the calculation of PMF. The biased probability distributions along the distance r accumulated in these sampling windows are unbiased and combined using the Weighted Histogram Analysis Method (WHAM) method to yield the unbiased distribution and the potential of mean force $w(r)$ ^{31,32}. The WHAM program implemented by Grossfield is used to calculate the PMF³³. For each receptor-ligand complex, three independent sets of umbrella sampling simulations were performed and the statistical uncertainties in the calculated PMF are estimated by the standard deviation of the results obtained from the three sets of umbrella sampling simulations. The total simulation time used to compute the PMF for a single receptor-ligand complex is $\sim 2.1 \mu\text{s}$.

The paths along which ligands are pulled out in the PMF calculation are identified by trial and error. First, a ligand atom **A** and a protein atom **a** are chosen to define axis $\mathbf{r}_{aA}$; using the Discovery Studio Visualizer (Biovia Inc), the ligand is translated manually along this axis with its orientation relative to the receptor fixed. If the unbinding ligand collides with the nearby protein atoms, (a clash is shown as red dashed line in the DS Visualizer), then a different set of atom pairs will be tried. After a few tries, one or more low energy paths may be identified; these paths are then used to run umbrella-sampling simulation to get PMF. As an example, for BI-224436, we identified two low free energy paths (Fig. S1 and S2,

Supporting Information), with the $\Delta G_{\text{bind}}^{\circ}$ estimated using path 2 lower than that from path 1. The PMF result reported in Table 1 and Table 2 are for the lowest free energy paths.

DDM calculation setup

A DDM calculation involves two legs of decoupling simulations in order to compute the $\Delta G_{\text{decoupl}}^{\text{bound}}$ and $-\Delta G_{\text{decoupl}}^{\text{bulk}}$. In the each leg of the decoupling in this work, the Coulomb intermolecular interaction is turned off first, and then the Lennard-Jones intermolecular interaction is switched off. Thus, the calculation provides estimates for the following free energy terms

$$\begin{aligned}\Delta G_{\text{decoupl}}^{\text{bound}} &= \Delta G_{\text{decoupl} - \text{Coulomb}}^{\text{bound}} + \Delta G_{\text{decoupl} - \text{LJ}}^{\text{bound}} \\ \Delta G_{\text{decoupl}}^{\text{bulk}} &= \Delta G_{\text{decoupl} - \text{Coulomb}}^{\text{bulk}} + \Delta G_{\text{decoupl} - \text{LJ}}^{\text{bulk}}\end{aligned}\quad (15)$$

Combining Eq. (15) with the Eq. (4), the absolute binding free energy $\Delta G_{\text{bind}}^{\circ}$ can be rearranged as

$$\begin{aligned}\Delta G_{\text{bind}}^{\circ} &= -\Delta G_{\text{restr}}^{\text{bound}} - \Delta G_{\text{decoupl} - \text{Coulomb}}^{\text{bound}} + \Delta G_{\text{decoupl} - \text{Coulomb}}^{\text{bulk}} - \Delta G_{\text{decoupl} - \text{LJ}}^{\text{bound}} \\ &+ \Delta G_{\text{decoupl} - \text{LJ}}^{\text{bulk}} + \Delta G_{\text{restr}}^{\text{gas}} = -\Delta G_{\text{restr}}^{\text{bound}} + \Delta \Delta G_{\text{Coulomb}} + \Delta \Delta G_{\text{LJ}} + \Delta G_{\text{restr}}^{\text{gas}}\end{aligned}\quad (16)$$

Here $\Delta \Delta G_{\text{Coulomb}} = -\Delta G_{\text{decoupl} - \text{Coulomb}}^{\text{bound}} + \Delta G_{\text{decoupl} - \text{Coulomb}}^{\text{bulk}}$, and $\Delta \Delta G_{\text{LJ}} = -\Delta G_{\text{decoupl} - \text{LJ}}^{\text{bound}} + \Delta G_{\text{decoupl} - \text{LJ}}^{\text{bulk}}$, which can be interpreted as the contributions of the effective electrostatic interactions and the nonpolar interactions to the total binding free energy, respectively^{35,39}.

It has been shown that since the simulations of the DDM calculation are performed in a finite-size, periodic solvent box, for charged ligands this can have a non-negligible effect on the calculated electrostatic decoupling free energies $\Delta G_{\text{decoupl} - \text{Coulomb}}^{\text{bound}}$ and

$\Delta G_{\text{decoupl} - \text{Coulomb}}^{\text{bulk}}$ due to lattice sum artifact.^{19,37} Here we used a scheme developed by Rocklin and coworkers to compute the correction for these effects¹⁹. The electrostatic finite-size corrections include the following contributions: periodicity-induced net-charge interactions, periodicity-induced net-charge undersolvation, discrete solvent effects, and residual integrated potential effects.¹⁹ The calculation of the residual integrated potential requires three Poisson-Boltzmann calculations (PB) of electrostatic potentials generated by different combinations of receptor and ligand charge distributions. Here the PB calculations are performed using the APBS program.³⁸ With the inclusion of these corrections, the final expression for the $\Delta G_{\text{bind}}^{\circ}$ from DDM is

$$\Delta G_{\text{bind}}^{\circ} = -\Delta G_{\text{restr}}^{\text{bound}} + G_{\text{Coulomb}} + \Delta \Delta G_{\text{LJ}} + \Delta G_{\text{elec_corr}}^{\text{finite_size}} + \Delta G_{\text{restr}}^{\text{gas}}\quad (17)$$

For the systems studied here, 11 λ s are used for the Coulomb decoupling: $\lambda_{\text{Coulomb}} = 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0$; and 17 λ s are used for the Lennard-Jones decoupling: $\lambda_{\text{LJ}} = 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.94, 0.985, 1.0$. TI is used to compute the resulting decoupling free energies i.e.

$$\Delta G_{\text{decoupl_Coulomb}} = \int_{\lambda_{\text{Coulomb}}=0}^{\lambda_{\text{Coulomb}}=1} \left\langle \frac{dU}{d\lambda_{\text{Coulomb}}} \right\rangle_{\lambda_{\text{Coulomb}}} d\lambda_{\text{Coulomb}} \text{ and}$$

$$\Delta G_{\text{decoupl_LJ}} = \int_{\lambda_{\text{LJ}}=0}^{\lambda_{\text{LJ}}=1} \left\langle \frac{dU}{d\lambda_{\text{LJ}}} \right\rangle_{\lambda_{\text{LJ}}} d\lambda_{\text{LJ}}. \text{ (FEP can also be used.)}$$

The soft-core potential

implemented in GROMACS is used for the Lennard-Jones decoupling step in the DDM calculation to avoid the problem of end point instability.³⁴ At each λ , a decoupling simulation is performed for 15 ns. The first 5 ns of the trajectory were treated as equilibration and the last 10 ns trajectory were used to compute $\Delta G_{\text{bind}}^{\circ}$. For the $\Delta G_{\text{bind}}^{\circ}$ calculation of each ligand, three independent DDM runs with different starting velocities were carried out. The statistical error bars are estimated from the standard deviations of the three independent sets of DDM simulations. A total of $\approx 2.5 \mu\text{s}$ simulation data are collected to compute a single $\Delta G_{\text{bind}}^{\circ}$ value.

MD Simulation Setup

In this work, the MD binding free energy simulations using PMF and DDM were performed using the GROMACS 4.6.4⁴⁰. The AMBER parm99ILDN force field is used to model the HIV-1 integrase catalytic core dimer, and the Amber GAFF parameters set⁴¹ and the AM1-BCC charge model⁴² are used to describe the different ligands. TIP3P water⁴³ boxes previously equilibrated at 300 K and 1 atm pressure were used to solvate the protein-ligand complex. The dimension of the solvent box is set up to ensure that the distance between solute atoms from nearest walls of the box is at least 10 Å. To maintain charge neutrality, 1 Cl^- and 1 Na^+ are added to the solvent boxes containing the protein-ligand complex and that containing the ligand, respectively. The electrostatic interactions were computed using the particle-mesh Ewald (PME) method⁴⁴ with a real space cutoff of 10 Å and a grid spacing of 1.0 Å. MD simulations were performed in the NPT ensemble with a time step of 2 fs.

RESULTS

Comparison of PMF and DDM in calculating the $\Delta G_{\text{bind}}^{\circ}$ for the charged ligands

Using the expressions of PMF and DDM derived here, we have computed the absolute binding free energies for three charged ligands that bind at the allosteric site in the dimer interface of the HIV-1 integrase catalytic core domain (CCD): see Table 1. Comparing with the experimental results, the errors in the $\Delta G_{\text{bind}}^{\circ}$ computed from PMF and DDM are in the range of 1.5–3.4 kcal/mol and 1.6–4.3 kcal/mol, respectively. While the two methods produced comparable results for the two stronger binders BI-224436 and KF116, the DDM method overestimated the binding affinity for the weak inhibitor KF134. Thus, only PMF

correctly reproduced the rank order of the binding affinities of these charged ligands. However, as seen from Table 1, the error bars in both sets of results are substantial, especially in the PMF-calculation of BI-224436. Although the PMF method provides a somewhat better estimate for the $\Delta G_{\text{bind}}^{\circ}$ for these charged ligands in the same amount of computing time, the statistical uncertainty in the calculated $\Delta G_{\text{bind}}^{\circ}$ makes it less straightforward to clearly favor one method over the other.

Overestimation in the PMF-calculated $\Delta G_{\text{bind}}^{\circ}$ for BI-224436 is likely related to the bulky tricyclic group carried by the ligand

We analyze the results of the PMF-calculated absolute binding free energies. As shown in Table 1, while the error in the calculated $\Delta G_{\text{bind}}^{\circ}$ for KF134 and KF116 are about 2 kcal/mol, the PMF calculation overestimates the absolute binding free energy for BI-224436 by -3.4 kcal/mol. To investigate the source of this overestimation of the $\Delta G_{\text{bind}}^{\circ}$, we examine the breakdown of the different contributions to the $\Delta G_{\text{bind}}^{\circ}$ computed in the PMF method: see Table 2. The term $-w(r_1^*)$ provides the most dominant contribution to the absolute binding free energy. In fact $-w(r_1^*)$ alone correctly predicts the rank order of the binding affinities of these charged ligands, since the $\Delta G_{\text{restr}}^{\text{bulk}}$ term (see Eq. (6)) is almost a constant, and the differences in $-\Delta G_{(\theta, \phi, \Theta, \Phi, \Psi)}^{\text{bound}}$, the free energy cost of turning on the ligand angular restraints, are smaller than the differences in the $-w(r_1^*)$.

As seen from Table 2 and Fig. 2, the error bar in the reversible work value $-w(r_1^*)$ for BI-224436 is 3.5 kcal/mol, which is much larger than the statistical errors in the other components. In contrast, the error bar in the PMF calculated for the other two ligands KF134 and KF116 are smaller: see Fig. 2. The different behavior in the calculated PMF for the different ligands can be traced to the chemical groups carried by the ligands: Fig. 3 shows the three ligands in the binding cavity of integrase. The binding modes of the three ligands are almost identical. However, BI-224436 carries a bulky tricyclic arene which is buried in a sub-pocket in the binding site, while in both KF134 and KF116 a single benzene ring occupies the same sub-pocket. Because of its large size, the tricyclic group in the BI-224436 may cause steric clash with the neighboring protein residues when the ligand is pulled out of the binding cavity; this could raise the value of the PMF $w(r)$ and leads to the overestimation of the binding free energy.

Fig. 2 also shows that in general, the error bars in the PMF increase with r . This is likely related to the fact that the ligand moves inside a cone with a cross-section area proportional to r^2 . Therefore, the accessible volume in the sampling of ligand external degrees of freedom grows with r^2 . It is possible that using a Cartesian coordinate system as the one described by Doudou et al.⁴⁵ in which the accessible volume is constant with distance from the binding site may help reducing the sampling problem at larger r .

Insights from the DDM calculation for charged ligands

Table 3 shows the results of the different free energy terms in the DDM-calculated ΔG_{bind}° for the three charged ligands. It can be seen that the electrostatic free energy of decoupling charged ligands from the protein binding pocket $-\Delta G_{decoupl}^{bound} - Coulomb$ is very large, which is due to both short and long range electrostatic effects and is consistent with the fact that the carboxylate moiety shared by the three charged ligands form multiple correlated hydrogen bonds (Fig. 4) with Glu170/His171/Thr174 in the binding pocket and also with neighboring waters.⁴⁶ However, this large favorable electrostatic free energy is almost exactly cancelled out by the equally large electrostatic free energy of decoupling the charged ligand from the solvent $\Delta G_{decoupl}^{bulk} - Coulomb$. The latter is equally large because the same carboxylate group on the ligands also forms multiple hydrogen bonds with the waters in the bulk (data not shown). As a result, the net electrostatic contribution to binding $\Delta \Delta G_{Coulomb}$ is very small, e.g. ranging from 0.9 to -1.5 kcal/mol. In fact, the ligand binding is dominated by nonpolar interactions $\Delta \Delta G_{LJ}$, which ranges from -11 to -12 kcal/mol. Although the ligand binding is dominated by nonpolar interactions, the intermolecular electrostatic interactions described by $-\Delta G_{decoupl}^{bound} - Coulomb$ is essential to compensate for the free energy cost $\Delta G_{decoupl}^{bulk} - Coulomb$ for the desolvation of the charged ligands. Therefore both electrostatic intermolecular interaction and nonpolar interactions are important to the binding of these charged ligands at the HIV-1 integrase CCD dimer interface.

It should also be noted that, while the electrostatic decoupling free energies are of large magnitude (~ -100 kcal/mol), the fluctuations in the $-\Delta G_{decoupl}^{bound} - Coulomb$ are on the order of only ~ 1 kcal/mol. This is somewhat surprising, as one would expect the fluctuations in the large electrostatic decoupling free energies to also be large.

As seen from Table 3, the electrostatic correction for the finite-size periodic system $\Delta G_{elec_corr}^{finite_size}$ is on the order of ~ -2 kcal/mol. Among the different contributions to the $\Delta G_{elec_corr}^{finite_size}$, the largest contribution is from the discrete solvent correction term $\Delta \Delta G_{DSC}$, while the correction terms for the periodicity-induced net-charge interactions and the periodicity-induced net-charge undersolvation nearly cancel out each other.

DISCUSSION

In this study, we compare the performance of DDM and PMF in computing the absolute binding free energies of charged ligands using examples taken from drug discovery projects we are working on.²⁶ For the charged ligands studied here, the PMF and DDM methods yield comparable results, although the PMF result is in somewhat better agreement with the experiment, including recapitulating the ranking order of the experimental binding affinities. A similar comparison study has been reported by Gumbart et al. who calculated the absolute binding free energies for an uncharged peptide using PMF and DDM¹⁶ and found that the ΔG_{bind}° computed from the two methods differ by just 0.2 kcal/mol. In our study, we focus on

the binding of charged small molecules. For the three charged ligands studied here, the difference in the ΔG_{bind}° calculated from PMF and DDM is considerably larger, at 1.2 – 2.8 kcal/mol. In addition, we decompose the ΔG_{bind}° from DDM into electrostatic and Lennard-Jones contributions as well as electrostatic corrections for the finite-size, to help dissect the driving force of binding and to study the effect of different terms on the accuracy of the DDM calculation. This free energy decomposition was not reported in the paper by Gumbart et al.¹⁶

We find that the main disadvantage of DDM is that the net binding free energy is obtained as the difference between two very large electrostatic decoupling free energies, which can be 100 kcal/mol. In contrast, all the terms in the PMF method have smaller magnitudes (i.e. 20 kcal/mol), since the ligand never leaves the solvent phase in the PMF simulations. Furthermore, in treating the charged ligands, complex procedures need to be applied with DDM to account for the systematic electrostatic errors due to use of finite-size periodic system. Such procedures are not needed for the PMF approach. In addition, the PMF method may have an advantage in treating large flexible ligands: since the DDM method involves a vacuum state, the rugged energy landscape for those large and flexible ligands in the vacuum state can lead to very slow convergence in conformational sampling.

The main advantage of DDM is that it allows more automatic setup, since the calculation does not rely on the geometrical definition of a physical pathway to pull the ligand out of the binding pocket and can therefore be readily applied to a diverse set of ligands in virtual screening. By comparison, the PMF method requires carefully identifying a pulling pathway to avoid potential atomic collisions with the nearby receptor atoms. This makes it difficult to be applied automatically for a diverse set of ligands. For some ligands/binding pocket combinations, such collision-free pathways can be difficult to identify or do not even exist. In such cases, the PMF computed by pulling will lead to overestimated binding free energy estimates. This is because, although binding free energy is a state function, and in principle only depends on the end states and not on pathway, but in practice the receptor atoms involved in the atomic clashes do not have enough time to relax, leading to errors in the estimate of the free energy difference.

Another advantage of DDM is that, in addition to providing absolute binding free energy estimates, the method also provides a thermodynamic decomposition of the total binding free energy into physically meaningful contributions such as electrostatic contributions versus nonpolar ones, which can provide insights about the nature of the interactions between the ligand and the binding pockets. However, since such decomposition schemes are pathway dependent, caution should be applied in their physical interpretation.

Lastly, we explain the rationale for explicitly including the V_{site} in the DDM expression in this work. V_{site} , the effective volume of the binding site, is a fundamental quantity in measuring binding affinity either experimentally or computationally. This is because although binding affinity experiments do not explicitly use restraint, they implicitly contain the information of V_{site} , since what is measured in experiment involves a binding signal which changes when a complex is formed; the onset of the detection of the binding signal

provide an experimental definition of $V_{site}^{10,47}$. In order to make direct comparison between calculated ΔG_{bind}° with that determined from experiment, the V_{site} used in the calculation needs to be consistent with what is detected as a bound state in the experiment, and this depends on the nature of the experimental probe being employed to detect binding. In practice, such a rigorous direct comparison has seldom been done partly because except for weak binders, the calculated ΔG_{bind}° is insensitive to the exact choice of V_{site} as long as the major configurations contributing to the binding affinity have been included in the calculation.¹⁰

The present work makes the connection for the first time between fundamental equations of absolute binding free energy (Eq. 1–3) which use the step function $\mathcal{I}(\xi)$ to define V_{site}^{10} and the standard DDM implementation^{11,12} which uses harmonic restraint to accelerate sampling. Our modification of using explicitly the V_{site} to define the complexed state in addition to the use of harmonic restraint to accelerate sampling is therefore of conceptual importance, because without explicitly including V_{site} , it is strictly speaking less rigorous to directly compare calculated absolute binding free energy with experiment. In some problems such as weak or non-specific binding, our modification can be of practical importance since in such cases the unrestrained ligand can dissociate from the receptor even in the fully coupled state. Overall, this modification makes DDM broad enough to cover all cases of small molecule binding.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

This research was supported by grants from the National Institutes of Health (NIH U54 401356 and GM30580 (RL)). We thank Dr. Emilio Gallicchio for helpful discussions.

REFERENCES

- (1). Mobley DL; Gilson MK Predicting Binding Free Energies: Frontiers and Benchmarks. Annual Review of Biophysics 2017, 46 (1), 531–558.
- (2). Chodera JD; Mobley DL; Shirts MR; Dixon RW; Branson K; Pande VS Alchemical Free Energy Methods for Drug Discovery: Progress and Challenges. Current Opinion in Structural Biology 2011, 21 (2), 150–160.21349700
- (3). Zwanzig RW High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. The Journal of Chemical Physics 1954, 22 (8), 1420–1426.
- (4). Lybrand TP; McCammon JA; Wipff G Theoretical Calculation of Relative Binding Affinity in Host-Guest Systems. Proc Natl Acad Sci U S A. 1986, 83 (4), 833–835.3456569
- (5). Jorgensen WL; Ravimohan C Monte Carlo Simulation of Differences in Free Energies of Hydration. The Journal of Chemical Physics 1985, 83 (6), 3050.
- (6). Wang L; Wu Y; Deng Y; Kim B; Pierce L; Krilov G; Lupyan D; Robinson S; Dahlgren MK; Greenwood J; Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. J. Am. Chem. Soc. 2015, 137 (7), 2695–2703.25625324
- (7). Limongelli V; Bonomi M; Parrinello M Funnel Metadynamics as Accurate Binding Free-Energy Method. Proceedings of the National Academy of Sciences 2013, 110 (16), 6358–6363.

- (8). Moraca F; Amato J; Ortuso F; Artese A; Pagano B; Novellino E; Alcaro S; Parrinello M; Limongelli V Ligand Binding to Telomeric G-Quadruplex DNA Investigated by Funnel-Metadynamics Simulations. *Proceedings of the National Academy of Sciences* 2017, 114 (11), E2136–E2145.
- (9). Xie B; Nguyen TH; Minh DDL Absolute Binding Free Energies between T4 Lysozyme and 141 Small Molecules: Calculations Based on Multiple Rigid Receptor Configurations. *Journal of Chemical Theory and Computation* 2017, 13 (6), 2930–2944.28430432
- (10). Gilson M; Given J; Bush B; Mccammon J The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review. *Biophysical Journal* 1997, 72 (3), 1047–1069.9138555
- (11). Boresch S; Tettinger F; Leitgeb M; Karplus M Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *Journal of Physical Chemistry B* 2003, 107 (35), 9535–9551.
- (12). Deng Y; Roux B Calculation of Standard Binding Free Energies: Aromatic Molecules in the T4 Lysozyme L99A Mutant. *Journal of Chemical Theory and Computation* 2006, 2 (5), 1255–1273.26626834
- (13). Woo H-J; Roux B Calculation of Absolute Protein-Ligand Binding Free Energy from Computer Simulations. *Proceedings of the National Academy of Sciences* 2005, 102 (19), 6825–6830.
- (14). Velez-Vega C; Gilson MK Overcoming Dissipation in the Calculation of Standard Binding Free Energies by Ligand Extraction. *Journal of Computational Chemistry* 2013, n/a-n/a.
- (15). Heinzlmann G; Henriksen NM; Gilson MK Attach-Pull-Release Calculations of Ligand Binding and Conformational Changes on the First BRD4 Bromodomain. *Journal of Chemical Theory and Computation* 2017, 13 (7), 3260–3275.28564537
- (16). Gumbart JC; Roux B; Chipot C Standard Binding Free Energies from Computer Simulations: What Is the Best Strategy? *Journal of Chemical Theory and Computation* 2013, 9 (1), 794–802.23794960
- (17). Lee MS; Olson MA Calculation of Absolute Protein-Ligand Binding Affinity Using Path and Endpoint Approaches. *Biophysical Journal* 2006, 90 (3), 864–877.16284269
- (18). Figueirido F; Del Buono GS; Levy RM On Finite-size Effects in Computer Simulations Using the Ewald Potential. *The Journal of Chemical Physics* 1995, 103 (14), 6133–6142.
- (19). Rocklin GJ; Mobley DL; Dill KA; Hünenberger PH Calculating the Binding Free Energies of Charged Species Based on Explicit-Solvent Simulations Employing Lattice-Sum Methods: An Accurate Correction Scheme for Electrostatic Finite-Size Effects. *The Journal of Chemical Physics* 2013, 139 (18), 184103.24320250
- (20). Dadarlat VM; Skeel RD Dual Role of Protein Phosphorylation in DNA Activator/Coactivator Binding. *Biophysical Journal* 2011, 100 (2), 469–477.21244843
- (21). Gumbart JC; Roux B; Chipot C Efficient Determination of Protein–Protein Standard Binding Free Energies from First Principles. *Journal of Chemical Theory and Computation* 2013, 9 (8), 3789–3798.
- (22). Cui D; Ou S; Patel S Free Energetics of Rigid Body Association of Ubiquitin Binding Domains: A Biochemical Model for Binding Mediated by Hydrophobic Interaction: Ubiquitin: Model Hydrophobic Interactions. *Proteins: Structure, Function, and Bioinformatics* 2014, 82 (7), 1453–1468.
- (23). Engelman A; Kessl JJ; Kvaratskhelia M Allosteric Inhibition of HIV-1 Integrase Activity. *Current Opinion in Chemical Biology* 2013, 17 (3), 339–345.23647983
- (24). Gallicchio E; Deng N; He P; Wickstrom L; Perryman AL; Santiago DN; Forli S; Olson AJ; Levy RM Virtual Screening of Integrase Inhibitors by Large Scale Binding Free Energy Calculations: The SAMPL4 Challenge. 2014, 28, 475–490.
- (25). Slaughter A; Jurado KA; Deng N; Feng L; Kessl JJ; Shkriabai N; Larue RC; Fadel HJ; Patel PA; Jena N; The Mechanism of H171T Resistance Reveals the Importance of Ndelta-Protonated His171 for the Binding of Allosteric Inhibitor BI-D to HIV-1 Integrase. *Retrovirology* 2014, 11 (1), 100.25421939
- (26). Deng N; Hoyte A; Mansour YE; Mohamed MS; Fuchs JR; Engelman AN; Kvaratskhelia M; Levy R Allosteric HIV-1 Integrase Inhibitors Promote Aberrant Protein Multimerization by Directly Mediating Inter-Subunit Interactions: Structural and Thermodynamic Modeling Studies:

Mechanism of Allosteric HIV-1 Integrase Inhibitors. *Protein Science* 2016, 25 (11), 1911–1917.27503276

- (27). Gallicchio E; Lapelosa M; Levy RM Binding Energy Distribution Analysis Method (BEDAM) for Estimation of Protein–Ligand Binding Affinities. *Journal of Chemical Theory and Computation* 2010, 6 (9), 2961–2977.21116484
- (28). Gallicchio E; Levy RM Recent Theoretical and Computational Advances for Modeling Protein–ligand Binding Affinities In *Advances in Protein Chemistry and Structural Biology*; Elsevier, 2011; Vol. 85, pp 27–80.21920321
- (29). Bennett CH Efficient Estimation of Free Energy Differences from Monte Carlo Data. *Journal of Computational Physics* 1976, 22 (2), 245–268.
- (30). Deng N-J; Cieplak P Free Energy Profile of RNA Hairpins: A Molecular Dynamics Simulation Study. *Biophysical Journal* 2010, 98 (4), 627–636.20159159
- (31). Kumar S; Rosenberg JM; Bouzida D; Swendsen RH; Kollman PA THE Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *Journal of Computational Chemistry* 1992, 13 (8), 1011–1021.
- (32). Gallicchio E; Andrec M; Felts AK; Levy RM Temperature Weighted Histogram Analysis Method, Replica Exchange, and Transition Paths $\dagger$. *Journal of Physical Chemistry B* 2005, 109 (14), 6722–6731.
- (33). Grossfield, A. *WHAM: The Weighted Histogram Analysis Method*; <http://membrane.urmc.rochester.edu/content/wham>.
- (34). Beutler TC; Mark AE; van Schaik RC; Gerber PR; van Gunsteren WF Avoiding Singularities and Numerical Instabilities in Free Energy Calculations Based on Molecular Simulations. *Chemical Physics Letters* 1994, 222 (6), 529–539.
- (35). Deng N; Zhang P; Cieplak P; Lai L Elucidating the Energetics of Entropically Driven Protein–Ligand Association: Calculations of Absolute Binding Free Energy and Entropy. *The Journal of Physical Chemistry B* 2011, 115 (41), 11902–11910.21899337
- (36). Deng Y; Roux B Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. *Journal of Physical Chemistry B* 2009, 113 (8), 2234–2246.
- (37). Reif MM; Oostenbrink C Net Charge Changes in the Calculation of Relative Ligand-Binding Free Energies via Classical Atomistic Molecular Dynamics Simulation. *Journal of Computational Chemistry* 2014, 35 (3), 227–243.24249099
- (38). Baker NA; Sept D; Joseph S; Holst MJ; McCammon JA Electrostatics of Nanosystems: Application to Microtubules and the Ribosome. *Proceedings of the National Academy of Sciences* 2001, 98 (18), 10037–10041.
- (39). Lin Y-L; Meng Y; Huang L; Roux B Computational Study of Gleevec and G6G Reveals Molecular Determinants of Kinase Inhibitor Selectivity. *Journal of the American Chemical Society* 2014, 136 (42), 14753–14762.25243930
- (40). Pronk S; Pall S; Schulz R; Larsson P; Bjelkmar P; Apostolov R; Shirts MR; Smith JC; Kasson PM; van der Spoel D; GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* 2013, 29 (7), 845–854.23407358
- (41). Wang J; Wolf RM; Caldwell JW; Kollman PA; Case DA Development and Testing of a General Amber Force Field. *Journal of Computational Chemistry* 2004, 25 (9), 1157–1174.15116359
- (42). Jakalian A; Bush BL; Jack DB; Bayly CI Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *Journal of Computational Chemistry* 2000, 21 (2), 132–146.
- (43). Jorgensen WL; Chandrasekhar J; Madura JD; Impey RW; Klein ML Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* 1983, 79 (2), 926.
- (44). Essmann U; Perera L; Berkowitz ML; Darden T; Lee H; Pedersen LG A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* 1995, 103 (19), 8577.
- (45). Doudou S; Burton NA; Henchman RH Standard Free Energy of Binding from a One-Dimensional Potential of Mean Force. *Journal of Chemical Theory and Computation* 2009, 5 (4), 909–918.26609600

- (46). Cui D; Zhang BW; Matubayasi N; Levy RM The Role of Interfacial Water in Protein–Ligand Binding: Insights from the Indirect Solvent Mediated Potential of Mean Force. *Journal of Chemical Theory and Computation* 2018, 14 (2), 512–526.29262255
- (47). Hill TL Theory of Protein Solutions. II. *The Journal of Chemical Physics* 1955, 23 (12), 2270–2274.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.

The coordinate frame in which the orientation and the position of the ligand relative to the receptor are defined. The ligand position is defined in the polar coordinates by the distance r_{aA} and two polar angles θ : b-a-A; ϕ : c-b-a-A. The ligand orientation is defined by the three Euler angles: Θ : a-A-B; Φ : a-A-B-C; Ψ : b-a-A-B.

Figure 2.

The PMF computed from independent umbrella sampling simulations for the protein-ligand complexes containing KF134, KF116 and BI-224436.

Figure 3.

The three charged ligands in the binding site of the HIV-1 integrase CCD dimer interface.

The functional groups circled by the red dashed line are buried in the protein binding cavity.

Figure 4.
Correlated H-bonds (green dashed lines) between the carboxylate moiety on the ligand and the protein binding pocket.

Table 1.

Comparisons of absolute binding free energies from DDM and PMF with experiment. Unit: kcal/mol.^a

Ligand	Structure	PMF	DDM	Experiment ^b
KF134		-7.0 ± 1.7	-9.8 ± 1.4	-5.5
KF116		-7.6 ± 1.4	-8.4 ± 1.4	-9.9
BI-224436		-13.7 ± 3.5	-12.5 ± 0.3	-10.3

^a. There are two symmetry related ALLINI binding sites in one HIV-1 Integrase CCD dimer. All the calculated ΔG_{bind}° are for the intrinsic binding constant at a single site.

^b. The experimental ΔG_{bind}° are obtained from surface plasmon resonance (SPR) measurements (private communication with the Kvaratskhelia lab).

Table 2.

Contributions to the absolute binding free energy $\Delta G_{\text{bind}}^{\circ}$ calculated using the PMF approach (See Eq. (11) for the meaning of different terms).

Component	KF134	KF116	BI-224436
$-w(r_1^*)$	-11.5 ± 1.5	-12.8 ± 1.1	-19.2 ± 3.5
$\Delta G_{\text{restr}}^{\text{bulk}}$	8.3	8.0	9.7
$-k_B T \ln \frac{\int_{\text{bound}} e^{-w(r)/k_B T} dr}{(2\pi k_B T / k_r)^{\frac{1}{2}}}$	-0.2 ± 0.1	0.04 ± 0.04	0.05 ± 0.03
$-\Delta G_{(\theta, \phi, \Theta, \Phi, \Psi)}^{\text{bound}}$	-3.6 ± 0.8	-2.8 ± 0.5	-4.2 ± 0.15
$\Delta G^{\circ}(\text{bind})$	-7.0 ± 1.7	-7.6 ± 1.4	-13.7 ± 3.5

Table 3.

The binding free energy decomposition according to DDM

	KF134	KF116	BI-224436
$\Delta G_{decoupl}^{bulk} - Coulomb$	-105.5 ± 0.9	-179.2 ± 0.7	-92.3 ± 0.2
$-\Delta G_{decoupl}^{bound} - LJ$	-0.95 ± 0.8	2.9 ± 0.9	-3.9 ± 0.2
$\Delta G_{decoupl}^{bulk} - Coulomb$	105.96 ± 0.06	180.1 ± 0.06	90.8 ± 0.01
$\Delta G_{decoupl}^{bulk} - LJ$	-10.5 ± 0.08	-13.4 ± 0.3	-8.4 ± 0.2
$-\Delta G_{restr}^{bound}$	-0.04 ± 0.005	-0.1 ± 0.02	-0.03 ± 0.002
ΔG_{restr}^{gas}	3.23	3.23	3.23
$\Delta G_{PBC} - Corr$	-2.0	-1.9	-1.9
ΔG_{bind}°	-9.8 ± 1.4	-8.4 ± 1.4	-12.5 ± 0.3