Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Binding Affinity Estimation from Restrained Umbrella Sampling Simulations

Vivek Govind Kumar
University of Arkansas
Shilpi Agrawal
University of Arkansas
Thallapuranam Kumar
University of Arkansas
Mahmoud Moradi (moradi@uark.edu)
University of Arkansas https://orcid.org/0000-0002-0601-402X

Article

Keywords:

Posted Date: August 1st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1443106/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Binding Affinity Estimation From Restrained Umbrella Sampling Simulations

*Vivek Govind Kumar, Shilpi Agrawal, Thallapuranam Krishnaswamy Suresh Kumar, and Mahmoud Moradi**

Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, AR 72701

* Correspondence: moradi@uark.edu

ABSTRACT

The protein-ligand binding affinity quantifies the binding strength between a protein and its ligand. Computer modeling and simulations can be used to estimate the binding affinity or binding free energy using data- or physics-driven methods or a combination thereof. Here, we discuss a purely physics-based sampling approach based on biased molecular dynamics (MD) simulations. Our proposed method generalizes and simplifies previously suggested stratification strategies that use umbrella sampling (US) or other enhanced sampling simulations with additional collective-variable based restraints. The novel approach presented here uses a flexible scheme that can be easily tailored for any system of interest. We estimate the binding affinity of human fibroblast growth factor 1 (hFGF1) to heparin hexasaccharide based on the available crystal structure of the complex as the initial model and four different variations of the proposed method to compare against the experimentally determined binding affinity obtained from isothermal titration calorimetry (ITC) experiments. Our results indicate that enhanced sampling methods that sample along the ligand-protein distance without restraining other degrees of freedom do not perform as well as those with additional restraints. In particular, restraining the orientation of the ligand plays the most crucial role in reaching a reasonable estimate for binding affinity.

We also compare our results with those from free energy perturbation (FEP) simulations and show the higher accuracy of our method in the case of hFGF1 heparin binding. The framework presented here provides a general scheme for designing practical binding free energy estimation methods for flexible ligands and proteins.

INTRODUCTION

Accurate quantification of absolute binding affinities remains a problem of major importance in computational biophysics[1,2,3,4]. In principle, accurate binding free energy calculations should be the cornerstone of any study investigating protein-ligand interactions. However, the high computational costs that typically accompany such calculations necessitate the improvement of the computational methods traditionally used to investigate complex biomolecular interactions[3,4,5]. Experimentally determined binding affinities are commonly used as benchmarks to judge the accuracy of various computational binding affinity estimation methods[5,6,7]. Several experimental techniques can be used to study protein-ligand binding equilibria[5,8]. For instance, ITC can detect the interaction of binding partners based on changes in solution heat capacity and binding partner concentration[8,9,10]. Other methods such as fluorescence spectroscopy rely on changes in fluorescence intensity upon ligand binding[8,11,12]. Surface plasmon resonance (SPR) can be used to calculate binding affinities based on changes in refractive index that occur when an immobilized binding partner interacts with a free binding partner[8,13,14]. Studies have found that experimental binding affinities can vary depending on the experimental method used[5,6,15]. Therefore, a thorough understanding of the experimental

conditions used to generate reference data is essential when comparing computationally determined binding affinities with experimental values.

Several computational methods at varying levels of rigor and complexity have been used to determine binding affinities for biomolecular interactions[3,16-28]. Knowledge-based statistical potentials and force field scoring potentials are typically used to rank docked protein-ligand or protein-protein complexes but can also be used for binding affinity prediction[29,30,31]. A major disadvantage of these methods is that they do not treat the entropic effects rigorously, which effectively decreases the accuracy of such binding affinity predictions[5,32]. This is also the case for methods like Molecular Mechanics/Poisson-Boltzmann-Surface Area (MM-PBSA) and Molecular Mechanics/Generalized Born-Surface Area (MM-GBSA), which combine sampling of conformations from explicit solvent MD simulations with free energy estimation based on implicit continuum solvent models[33,34,35]. Adequate sampling of protein and ligand conformational dynamics as well as ligand roto-translational movements with respect to the protein is essential for accurately quantifying the entropic reduction arising from the binding event[35,36,37]. MM-PBSA/GBSA methods typically neglect such entropic contributions to the binding free energy or do not treat them rigorously[34,35].

One of the best-known binding free energy estimation methods is alchemical FEP, where scaling of non-bonded interactions enables reversal decoupling of the ligand from its environment in the bound state as well as the unbound state[38,39,40,41]. Most entropic and enthalpic contributors to changes in binding affinity are typically considered during FEP simulations, thus avoiding the approximations used by methods like MM-PBSA/GBSA[5,42]. A disadvantage of FEP is the fact that ligands tend to move away from the binding site

during the decoupling process, which results in poorly defined target states of the FEP calculation being used as starting states for the re-coupling process[43]. Using receptor-ligand restraints to resolve this issue[17,40,44,45] introduces some ambiguity to the way a standard state is defined, with a level of correlation between the size of the simulation cell and the standard state[46]. This can be corrected via the use of appropriate geometrical restraints[16,47,23].

Unrestrained long timescale MD simulations should theoretically allow for the investigation and quantification of protein-ligand or protein-protein binding events[48,49]. While microsecond-level MD simulations provide a more accurate description of protein conformational dynamics as compared to shorter simulations[50], efficient sampling of the conformational landscape remains a major issue and requires access to timescales beyond the capabilities of current MD simulations[51,52]. Several methods have been developed to tackle the sampling problem. Markov state models allow the sampling and characterization of native as well as alternative binding states[53,54,55]. Similarly, weighted ensemble (WE) simulations sample the conformational landscape along one or more discretized reaction coordinates based on the assignment of a statistical weight to each simulation[56,57]. More traditionally, US along such reaction coordinates can be used to guide the binding or unbinding of a ligand, after which algorithms like the weighted histogram analysis method (WHAM) can be used to calculate a unidimensional potential of mean force (PMF) which quantifies ligand binding and unbinding along a reaction coordinate[58,59]. Better convergence of the calculated free energy profiles can be achieved by the exchange of conformations between successive US windows as in the bias-exchange umbrella sampling (BEUS)[60,61,62]. Other methods based on similar principles include umbrella integration[63],

well-tempered metadynamics[64], adaptive biasing force (ABF) simulations[65] and variations of these techniques.

Incomplete sampling of important degrees of freedom, such as orientation of the ligand with respect to the protein, remains a major disadvantage of unidimensional PMF-based methods[3,4]. To resolve this problem, Woo and Roux[3] have devised a method wherein explicitly defined geometrical restraints on the orientation and conformation of the binding partners are used to reduce the conformational entropy of the biomolecular system being studied[3,4]. This results in improved convergence of the PMF calculation[3,4]. The introduction of a restraining potential based on the root-mean-square deviation (RMSD) of the ligand relative to its average bound conformation, reduces the flexibility of the ligand and the number of conformations that need to be sampled[3,4]. This method avoids the need to decouple the ligand from its surrounding environment as required by alchemical FEP[3,4,38-41]. Recent studies have described applications and extensions of the methodology proposed by Woo and Roux[4,66].

Here, we describe a purely physics-based enhanced sampling method based on biased MD simulations, which is similar in principle to the stratification strategy proposed by Woo and Roux[3,4]. Although we use the US method as our enhanced sampling technique, the methodology is generalizable to other techniques as long as they can be combined with additional restraints. There are several important differences between our method and that of Woo and Roux[3,4] including: (i) providing a general scheme that can be easily adapted to any number of restraints; (ii) the non-parametric reconstruction of the grid PMF, as defined below; and (iii) the use of the unidimensional orientation angle of the ligand with respect to the protein as a collective variable for restraining, as opposed to the use of three Euler

angles. We have used this methodology to calculate the binding affinity for the interaction of human fibroblast growth factor 1 (hFGF1) with heparin hexasaccharide, its glycosaminoglycan (GAG) binding partner. hFGF1 is an important signaling protein that is implicated in physiological processes such as cell proliferation and differentiation, neurogenesis, wound healing, tumor growth and angiogenesis[67,68,69,70,71]. GAGs are linear anionic polysaccharides that interact with positively charged regions of FGF binding partners to regulate their biological activity[70,72-80]. The hFGF1-heparin complex is the most well-known and broadly characterized protein-GAG complex[81,82]. Heparin binding is thought to stabilize hFGF1 and impart protection against proteolytic degradation. In this study, we show that the absolute binding affinity for the hFGF1-heparin interaction calculated using our novel approach, is in good agreement with binding affinity data from ITC experiments. Four alternative methods are used for estimating the absolute binding affinity within the formalism presented here to determine the workings of the methodology and the effect of the application of different (or no) restraints. We also compare our results with those obtained from FEP simulations and show that although performing longer FEP simulations could improve the accuracy of binding affinity estimates when compared to short FEP simulations, our novel approach is still more accurate than FEP when similar simulation times are used.

THEORETICAL FOUNDATION

Binding affinity is often quantified using the equilibrium dissociation constant (K_d), defined as:

$$K_d = [P][L]/[P:L] \qquad (1)$$

where $[P]$, $[L]$, and $[P:L]$ are the concentrations of protein, ligand, and the protein-ligand complex, respectively. Computationally, the absolute binding free energy ($\Delta G°$), which is the standard molar free energy of binding, is more convenient to calculate. The dissociation constant and the absolute binding free energy are related via

$$\Delta G° = RT \ln \frac{K_d}{1M} \qquad (2)$$

where R is the gas constant and T is the temperature. Various strategies have been used to estimate $\Delta G°$, some of which were briefly discussed above. The methodology proposed here has a significant resemblance to the stratification strategy of Woo and Roux[3,4]. However, the two methods have major differences as will be discussed later.

Absolute binding free energy or $\Delta G°$ is the free energy change associated with moving the ligand from the bulk to the binding pocket. Within the formalism presented in this work, $\Delta G°$ is determined from the grid PMF $G(\boldsymbol{x})$, where $\boldsymbol{x}$ is the position of the ligand mass center from the center of the binding pocket, $G(\boldsymbol{x})$ is the potential of mean force (PMF) associated with the ligand position $\boldsymbol{x}$. In practice, we need to bin the 3D space and define the PMF at every bin or grid point as:

$$G(\boldsymbol{x}) = -RT \ln p(\boldsymbol{x}) \qquad (3)$$

where $p(\boldsymbol{x})$ is the probability of finding the ligand at bin $\boldsymbol{x}$.

We define $\Delta G(\boldsymbol{x}) = G(\boldsymbol{x}) - G(\mathbf{0})$, where $\boldsymbol{x} = \mathbf{0}$ (i.e., the center of the binding pocket) is defined as the grid point associated with the lowest grid PMF. One can show:

$$\Delta G° = -RT \ln \frac{\int_{pocket} e^{-\frac{G(\boldsymbol{x})}{RT}} dV}{\int_{bulk} e^{-\frac{G(\boldsymbol{x})}{RT}} dV} = -RT \ln \frac{\int_{pocket} e^{-\frac{\Delta G(\boldsymbol{x})}{RT}} dV}{\int_{bulk} e^{-\frac{\Delta G(\boldsymbol{x})}{RT}} dV} \qquad (4)$$

in which the binding "pocket" refers to all $\boldsymbol{x}$ where the ligand is considered bound and "bulk" refers to all $\boldsymbol{x}$ where the ligand is not interacting with the protein. Since $\Delta G(\boldsymbol{x})$ is the same everywhere in the bulk, we can simplify Relation (4) as follows:

$$\Delta G^{\circ} = -RT \ln \frac{V_P}{e^{-\frac{\Delta G(\boldsymbol{x}_B)}{RT}} V_B} = -\Delta G(\boldsymbol{x}_B) - RT \ln \frac{V_P}{V_B} \qquad (5)$$

where V_B is the bulk volume per protein associated with the standard concentration (i.e., $1\ M$), $\boldsymbol{x}_B$ is any grid point in the bulk, and V_P is the binding pocket volume defined as:

$$V_P = \int_{pocket} e^{-\frac{\Delta G(\boldsymbol{x})}{RT}} dV \qquad (6)$$

Defining ΔG_V as the contribution of the difference between the volume of the binding pocket and the bulk to the binding free energy:

$$\Delta G_V = -RT \ln \frac{V_P}{V_B} \qquad (7)$$

Combining (5) and (7), we have:

$$\Delta G^{\circ} = -\Delta G(\boldsymbol{x}_B) + \Delta G_V \qquad (8)$$

We can find the bulk volume (V_B) associated with the standard concentration for a single protein approximately as:

$$V_B = \frac{\frac{1}{N_A}\ \text{mol}}{1\ M} = \frac{1}{N_A}\ L \approx 1661\ \text{Å}^3 \qquad (9)$$

where N_A is the Avogadro's constant. We can now rewrite ΔG_V as:

$$\Delta G_V = -RT \ln \frac{V_P}{V_B} = -RT \ln \frac{V_P}{\text{Å}^3} + RT \ln \frac{V_B}{\text{Å}^3} = \Delta G_P - \Delta G_B \qquad (10)$$

in which ΔG_B is the bulk volume contribution and ΔG_P is the binding pocket contribution:

$$\begin{cases} \Delta G_B = -RT \ln \frac{V_B}{\text{Å}^3} \approx -7.42RT \\ \Delta G_P = -RT \ln \frac{V_P}{\text{Å}^3} = -RT \ln \int_{pocket} e^{-\frac{\Delta G(\boldsymbol{x})}{RT}} \frac{dV}{\text{Å}^3} \end{cases} \quad (11)$$

Determining both $\Delta G(\boldsymbol{x}_B)$ and ΔG_P requires finding the grid PMF $\Delta G(\boldsymbol{x})$. $\Delta G(\boldsymbol{x}_B)$ is the PMF difference between the binding pocket center and the bulk and ΔG_P also requires an estimate for $\Delta G(\boldsymbol{x})$ within the binding pocket. We therefore do not need to find $\Delta G(\boldsymbol{x})$ for all $\boldsymbol{x}$ if we have a good estimate for $\Delta G(\boldsymbol{x})$ within the binding pocket and in the bulk. Ideally, $\Delta G(\boldsymbol{x})$ for these points can be determined by pulling the ligand out of the binding pocket towards the bulk and using an enhanced sampling technique such as US to sample the space of a collective variable such as d, i.e., the distance between the mass centers of the ligand and protein. $\Delta G(\boldsymbol{x})$ can be estimated for all sampled grid points $\boldsymbol{x}$ using this distance-based US simulation. Note that the collective variable used for biasing would be d, while the collective variable used for the PMF calculations would be the 3D position vector of the mass center of ligand with respect to protein's binding pocket center. One may estimate the grid PMF from the distance-based US simulations using a non-parametric reweighting algorithm as discussed in the Methods section. $\Delta G(\boldsymbol{x})$ can also be used to estimate ΔG_P as defined in Relation (11). There is often no need to strictly define the binding pocket since only low $\Delta G(\boldsymbol{x})$ values have nonnegligible contribution to V_P and thus even if we include all sampled grid points, only those close to the binding pocket center have nonnegligible contributions.

A practical issue with determining $\Delta G(\boldsymbol{x}_B)$ is the convergence. The key obstacles for the sampling that slow down the convergence are the orientation of the ligand, and the conformational changes of the ligand and protein. Using an approach similar in spirit to

the previously proposed stratification strategy[3,4,35], we can circumvent extensive sampling of these degrees of freedom. Let us first focus on the orientation of the ligand (Ω). We can restrain Ω during the distance-based US simulations using a biasing potential ($\frac{1}{2}k\Omega^2$) and later correct the free energy difference based on the PMF associated with the Ω, which is different in the bulk ($F(\boldsymbol{x}_{\mathrm{B}},\Omega)$) and in the binding pocket ($F(\mathbf{0},\Omega)$). More generally, for any grid point $\boldsymbol{x}$, we may determine $\Delta G(\boldsymbol{x})$ based on the PMF associated with the Ω at $\boldsymbol{x}$ ($F(\boldsymbol{x},\Omega)$) and $\mathbf{0}$ ($F(\mathbf{0},\Omega)$):

$$e^{-\frac{\Delta G(\boldsymbol{x})}{RT}} = \frac{\int_0^\pi e^{-\frac{F(\boldsymbol{x},\Omega)}{RT}}\,d\Omega}{\int_0^\pi e^{-\frac{F(\mathbf{0},\Omega)}{RT}}\,d\Omega} \qquad (12)$$

Note that $F(\boldsymbol{x},\Omega)$ is the PMF associated with $\boldsymbol{x}$ and Ω, defined such that:

$$G(\boldsymbol{x}) = c - RT\ln\int_0^\pi e^{-\frac{F(\boldsymbol{x},\Omega)}{RT}}\,d\Omega \qquad (13)$$

where c is an arbitrary constant. We therefore have:

$$e^{-\frac{\Delta G(\boldsymbol{x})}{RT}} = \frac{\int_0^\pi e^{-\frac{F(\boldsymbol{x},\Omega)}{RT}}\,d\Omega}{\int_0^\pi e^{-\frac{F(\boldsymbol{x},\Omega)+\frac{1}{2}k\Omega^2}{RT}}\,d\Omega} \times \frac{\int_0^\pi e^{-\frac{F(\mathbf{0},\Omega)+\frac{1}{2}k\Omega^2}{RT}}\,d\Omega}{\int_0^\pi e^{-\frac{F(\mathbf{0},\Omega)}{RT}}\,d\Omega} \times \frac{\int_0^\pi e^{-\frac{F(\boldsymbol{x},\Omega)+\frac{1}{2}k\Omega^2}{RT}}\,d\Omega}{\int_0^\pi e^{-\frac{F(\mathbf{0},\Omega)+\frac{1}{2}k\Omega^2}{RT}}\,d\Omega} \qquad (14)$$

We now define $G_\Omega(\boldsymbol{x})$ as the grid PMF of the restrained system (by Ω):

$$G_\Omega(\boldsymbol{x}) = -RT\ln\int_0^\pi e^{-\frac{F(\boldsymbol{x},\Omega)+\frac{1}{2}k\Omega^2}{RT}}\,d\Omega \qquad (15)$$

We also define $U_\Omega(\boldsymbol{x})$ as the average biasing potential at grid point $\boldsymbol{x}$:

$$U_\Omega(\boldsymbol{x}) = -RT\ln\langle e^{-\frac{\frac{1}{2}k\Omega^2}{RT}}\rangle_{\boldsymbol{x}} = -RT\ln\frac{\int_0^\pi e^{-\frac{F(\mathbf{x},\Omega)+\frac{1}{2}k\Omega^2}{RT}}\,d\Omega}{\int_0^\pi e^{-\frac{F(\mathbf{x},\Omega)}{RT}}\,d\Omega} \qquad (16)$$

Now we have from Relations (14), (15), and (16):

$$\Delta G(\boldsymbol{x}) = \Delta G_{\Omega}(\boldsymbol{x}) - \Delta U_{\Omega}(\boldsymbol{x}) \qquad (17)$$

where, the free energy of grid point $\boldsymbol{x}$ from the center $\mathbf{0}$ ($\Delta G(\boldsymbol{x})$) is calculated based on its equivalent free energy ($\Delta G_{\Omega}(\boldsymbol{x})$) in a system biased by a harmonic restraint on Ω and a correction term $\Delta U_{\Omega}(\boldsymbol{x})$. For $\boldsymbol{x} = \boldsymbol{x}_B$:

$$\Delta U_{\Omega}(\boldsymbol{x}_B) = -RT \ln \frac{\langle e^{-\frac{\frac{1}{2}k\Omega^2}{RT}} \rangle_{bulk}}{\langle e^{-\frac{\frac{1}{2}k\Omega^2}{RT}} \rangle_{pocket}} \qquad (18)$$

To determine the above ensemble averages, we need to determine the PMF along Ω for the bound and unbound ligand and calculate the ensemble averages analytically using Relation (16). $\Delta G_{\Omega}(\boldsymbol{x}_B)$ can be determined from PMF calculations, where the distance between the protein and ligand is varied and the orientation of the ligand is restrained (distance-based BEUS with restrained orientation). We note that:

$$V_P = \int_{pocket} e^{-\frac{\Delta G(x)}{RT}} dV = \int_{pocket} e^{-\frac{\Delta G_{\Omega}(x) - \Delta U_{\Omega}(x)}{RT}} dV \approx \int_{pocket} e^{-\frac{\Delta G_{\Omega}(x)}{RT}} dV \quad (19)$$

where we assume $\Delta U_{\Omega}(\boldsymbol{x})$ is negligible for $\boldsymbol{x}$ within the binding pocket. In other words, $\langle e^{-\frac{\frac{1}{2}k\Omega^2}{RT}} \rangle_{\boldsymbol{x}} \approx \langle e^{-\frac{\frac{1}{2}k\Omega^2}{RT}} \rangle_{\mathbf{0}}$ for $\boldsymbol{x}$ close to $\mathbf{0}$.

In brief, if we choose to restrain the orientation, our absolute binding free energy estimate includes the following terms (using Relations (8) and (17)):

$$\Delta G^{\circ} = -\Delta G_{\Omega}(\boldsymbol{x}_B) + \Delta \mathrm{U}_{\Omega}(\boldsymbol{x}_B) + \Delta G_V \qquad (20)$$

$F(\boldsymbol{x}_B, \Omega)$ can be calculated numerically from orientation angle distribution of a free ligand: $\mathrm{F}(\boldsymbol{x}_B, \Omega) = -RT \ln p(\Omega)$, where $p(\Omega)$ is determined from the distribution of Euler angles ($p(\phi, \theta, \psi) = \frac{1}{8\pi^2} \sin\theta$, where $0 \leq \phi, \psi \leq 2\pi$ and $0 \leq \theta \leq \pi$) given that:

$$\cos\frac{\Omega}{2} = \cos\frac{\phi}{2}\cos\frac{\theta}{2}\cos\frac{\psi}{2} + \sin\frac{\phi}{2}\sin\frac{\theta}{2}\sin\frac{\psi}{2} \qquad (21)$$

$\langle e^{-\frac{\frac{1}{2}k\Omega^2}{RT}}\rangle_{bulk}$ can then be calculated using Relation (16) with numerically estimated $F(\boldsymbol{x}_B, \Omega)$ and the k value used in the simulations. $F(\mathbf{0}, \Omega)$ can be determined approximately using orientation-based US simulations of bound ligand. F$(\mathbf{0}, \Omega)$ can then be used to estimate $\langle e^{-\frac{\frac{1}{2}k\Omega^2}{RT}}\rangle_{pocket}$ using Relation (16).

The above strategy can be extended to other degrees of freedom for which unbiased sampling may hinder the convergence. Most notably, the internal conformational changes of the ligand and that of the protein may also play a crucial role in slowing down the convergence. In the following, we show how one can restrain not only the orientation of the ligand but also the root-mean-square-deviation (RMSD) of the ligand (denoted here by r) in distance-based US simulations (along d) to speed up convergence. In this case, the grid PMF difference $\Delta G(\boldsymbol{x})$ is calculated based on $\Delta G_{\Omega,r}(\boldsymbol{x})$, the grid PMF of a system whose Ω and r are both restrained:

$$e^{-\frac{\Delta G(\boldsymbol{x})}{RT}} = \frac{\int_0^\infty \int_0^\pi e^{-\frac{G(\boldsymbol{x},\Omega,\mathrm{r})}{RT}}\, d\Omega\, dr}{\int_0^\infty \int_0^\pi e^{-\frac{G(\mathbf{0},\Omega,\mathrm{r})}{RT}}\, d\Omega\, dr} \qquad (22)$$

Using a similar strategy as in Relation (14), we have:

$$e^{-\frac{\Delta G(\boldsymbol{x})}{RT}} = \frac{\int_0^{\infty}\int_0^{\pi} e^{-\frac{F(\boldsymbol{x},\Omega,\mathrm{r})}{RT}}\, d\Omega\, dr}{\int_0^{\infty}\int_0^{\pi} e^{-\frac{F(\mathbf{x},\Omega,\mathrm{r})+\frac{1}{2}k'\mathrm{r}^2}{RT}}\, d\Omega\, dr} \times \frac{\int_0^{\infty}\int_0^{\pi} e^{-\frac{F(\mathbf{x},\Omega,\mathrm{r})+\frac{1}{2}k'\mathrm{r}^2}{RT}}\, d\Omega\, dr}{\int_0^{\infty}\int_0^{\pi} e^{-\frac{F(\mathbf{x},\Omega,\mathrm{r})+\frac{1}{2}k'\mathrm{r}^2+\frac{1}{2}k\Omega^2}{RT}}\, d\Omega\, dr}$$

$$\times \frac{\int_0^{\infty}\int_0^{\pi} e^{-\frac{F(\mathbf{0},\Omega,\mathrm{r})+\frac{1}{2}k'\mathrm{r}^2+\frac{1}{2}k\Omega^2}{RT}}\, d\Omega\, dr}{\int_0^{\infty}\int_0^{\pi} e^{-\beta\left(F(\mathbf{0},\Omega,\mathrm{r})+\frac{1}{2}k'\mathrm{r}^2\right)}\, d\Omega\, dr} \times \frac{\int_0^{\infty}\int_0^{\pi} e^{-\frac{F(\mathbf{0},\Omega,\mathrm{r})+\frac{1}{2}k'\mathrm{r}^2}{RT}}\, d\Omega\, dr}{\int_0^{\infty}\int_0^{\pi} e^{-\beta F(\mathbf{0},\Omega,\mathrm{r})}\, d\Omega\, dr}$$

$$\times \frac{\int_0^{\infty}\int_0^{\pi} e^{-\frac{F(\mathbf{x},\Omega,\mathrm{r})+\frac{1}{2}k'\mathrm{r}^2+\frac{1}{2}k\Omega^2}{RT}}\, d\Omega\, dr}{\int_0^{\infty}\int_0^{\pi} e^{-\frac{F(\mathbf{0},\Omega,\mathrm{r})+\frac{1}{2}k'\mathrm{r}^2+\frac{1}{2}k\Omega^2}{RT}}\, d\Omega\, dr} \qquad (23)$$

which results in:

$$e^{-\frac{\Delta G(\boldsymbol{x})}{RT}} = \frac{\langle e^{-\frac{\frac{1}{2}k'\mathrm{r}^2}{RT}}\rangle_{\mathbf{0}}}{\langle e^{-\frac{\frac{1}{2}k'\mathrm{r}^2}{RT}}\rangle_{\boldsymbol{x}}} \times \frac{\langle e^{-\frac{\frac{1}{2}k\Omega^2}{RT}}\rangle_{\mathbf{0}}^{r}}{\langle e^{-\frac{\frac{1}{2}k\Omega^2}{RT}}\rangle_{\boldsymbol{x}}^{r}} \times \frac{e^{-\beta G_{\Omega,\mathrm{r}}(\boldsymbol{x})}}{e^{-\beta G_{\Omega,\mathrm{r}}(\mathbf{0})}} \qquad (24)$$

Here we have defined $G_{\Omega,\mathrm{r}}(\boldsymbol{x})$ as:

$$G_{\Omega,\mathrm{r}}(\boldsymbol{x}) = -RT\ln\int_0^{\infty}\int_0^{\pi} e^{-\frac{F(\mathbf{x},\Omega,\mathrm{r})+\frac{1}{2}k'\mathrm{r}^2+\frac{1}{2}k\Omega^2}{RT}}\, d\Omega\, dr \qquad (25)$$

We also define $U_r(\boldsymbol{x})$ similar to $U_\Omega(\boldsymbol{x})$ in Relation (15) except for using r instead of Ω. $U_\Omega^r(\boldsymbol{x})$ is also defined similar to $U_\Omega(\boldsymbol{x})$ except for the additional restraint on r:

$$U_\Omega^r(\boldsymbol{x}) = -RT\ln\langle e^{-\frac{\frac{1}{2}k\Omega^2}{RT}}\rangle_{\boldsymbol{x}}^{r} = -RT\ln\frac{\int_0^{\infty}\int_0^{\pi} e^{-\frac{F(\mathbf{x},\Omega,\mathrm{r})+\frac{1}{2}k'\mathrm{r}^2+\frac{1}{2}k\Omega^2}{RT}}\, d\Omega\, dr}{\int_0^{\infty}\int_0^{\pi} e^{-\beta\left(F(\mathbf{x},\Omega,\mathrm{r})+\frac{1}{2}k'\mathrm{r}^2\right)}\, d\Omega\, dr} \qquad (26)$$

Finally, we have:

$$\Delta G(\boldsymbol{x}) = \Delta G_{\Omega,\mathrm{r}}(\boldsymbol{x}) - \Delta U_{\mathrm{r}}(\boldsymbol{x}) - \Delta U_\Omega^r(\boldsymbol{x}) \qquad (27)$$

In brief, if we choose to restrain both the orientation and RMSD, our absolute binding free energy estimate includes the following terms:

$$\Delta G^\circ = -\Delta G_{\Omega,\mathrm{r}}(\boldsymbol{x}_B) + \Delta \mathrm{U}_\mathrm{r}(\boldsymbol{x}_B) + \Delta \mathrm{U}^\mathrm{r}_\Omega(\boldsymbol{x}_B) + \Delta G_V \qquad (28)$$

Here we are using an approximation similar to that in Relation (19):

$$V_P \approx \int_{pocket} e^{-\frac{\Delta G_{\Omega,\mathrm{r}}(\boldsymbol{x})}{RT}} dV \qquad (29)$$

Using Relations (20) and (28), we can generalize the stratification strategy to include three restraints on arbitrary collective variables α, β, and γ:

$$\Delta G^\circ = -\Delta G_{\alpha,\beta,\gamma}(\boldsymbol{x}_B) + \Delta \mathrm{U}_\gamma(\boldsymbol{x}_B) + \Delta \mathrm{U}^\gamma_\beta(\boldsymbol{x}_B) + \Delta \mathrm{U}^{\beta,\gamma}_\alpha(\boldsymbol{x}_B) + \Delta G_V \qquad (30)$$

where:

$$\Delta G_V \approx -RT \ln \int_{pocket} e^{-\frac{\Delta G_{\alpha,\beta,\gamma}(\boldsymbol{x})}{RT}} \frac{dV}{\mathrm{\AA}^3} - \Delta G_B \qquad (31)$$

METHODS

Isothermal titration calorimetry of hFGF-1 with heparin hexasaccharide

ITC data was obtained using MicroCal iTC 200 (Malvern Inc.). The change in heat during the biomolecular interaction was measured by titrating heparin (loaded in the syringe) into the hFGF1 solution in the calorimetric cell. Both the protein and the heparin samples were prepared in the buffer containing 10 mM phosphate buffer with 100 mM NaCl at pH 7.2 and were degassed prior to loading. The protein to heparin ratio was maintained at 1:10 with the protein concentration being 100 μM and the heparin concentration being 1mM. A total of 30 injections were conducted with a constant temperature of 25 °C and stirring speed of 300 rpm. One set of sites binding model was used for the ITC binding curve[85]. The standard binding free energy ΔG° was determined from dissociation constant via relation (2) at $T = 25℃$. The experiment was repeated three times with the same sample

and the results obtained were very similar to each other. The mean and standard deviation were reported for both K_d and $\Delta G°$.

All-atom MD simulations

Our simulations were based on the x-ray crystal structure of the dimeric complex with a heparin hexasaccharide (PDB:2AXM, resolution: 3.0 Å)[86]. One of the hFGF1 protomers was removed leaving one protein and one ligand in the model of the holo protein. The model for the apo protein was based on the x-ray crystal structure of unbound monomeric hFGF1 (PDB: 1RG8, resolution: 1.1 Å)[87]. The simulations (residues 12-137 in the PDB file correspond to residues 26-151 in the experimental sequence) and experiments were performed using a truncated version of hFGF1 (residues 13-154) which did not contain the unstructured 12 amino acid N-terminal segment. Heparin hexasaccharide consists of N, O6 disulfo-glucosamine and 2-O-sulfo-alpha-L-idopyranuronic acid repeats[86]. The models were solvated in a box of TIP3P waters and 0.15 M NaCl. MD simulations were performed using the NAMD 2.13[88] simulation package with the CHARMM36m all-atom additive force field[89]. Initially, we energy-minimized the systems for 10,000 steps using the conjugate gradient algorithm[90]. Subsequently, we relaxed the systems using restrained MD simulations in a stepwise manner (for a total of ∼1 ns) using the standard CHARMM-GUI protocol[91,92]. The initial relaxations were performed in an NVT ensemble while the production runs were performed in an NPT ensemble. Simulations were carried out using a 2-fs time step at 300 K using a Langevin integrator with a damping coefficient of $\gamma = 0.5$ ps^{-1}. The pressure was maintained at 1 atm using the Nosé−Hoover Langevin piston method[90,93]. The smoothed cutoff distance for non-bonded interactions was set to 10−12 Å

and long-range electrostatic interactions were computed with the particle mesh Ewald (PME) method[94]. The initial runs lasted 15 nanoseconds, followed by the productions run on the supercomputer Anton 2 (Pittsburgh Supercomputing Center) for 4.8 μs, with a time-step of 2.5 fs. These equilibrium simulations have previously been described in a related study[95].

MD simulations of free heparin hexasaccharide

Heparin hexasaccharide[86] was simulated in a rectangular water box without the protein. The system was set up as described previously. The final conformation after relaxation was then used as the starting conformation for 10 production runs for 40 ns each. The total simulation time was around 400 ns.

Steered Molecular Dynamics (SMD) simulations

The final conformations of the hFGF1-heparin[95], apo hFGF1[95] and free heparin hexasaccharide equilibrium simulations were used to generate starting conformations for the non-equilibrium pulling simulations. Four collective variables[96] were used for the SMD simulations[97]: (1) distance between the heavy-atom center of mass of heparin and that of the protein (d), (2) the orientation angle of heparin with respect to the protein (Ω), (3) RMSD of the protein (r_P) and (4) RMSD of heparin (r_L). Six independent sets of simulations were performed. The distance-based SMD simulation was run for 9.5 ns, while the orientation based SMD simulation was run for 8 ns. The distance-based SMD

simulation was used to pull the heparin away from the protein by approximately 30 Å (10→40 Å) with a force constant of 100 kcal/(mol.Å^2). The orientation angle was also restrained in this simulation with a force constant of 0.5 kcal/(mol.$degree^2$) to stay close to its initial orientation in the bound state. The orientation-based SMD simulation was used to rotate the bound heparin locally with respect to the protein (0°→73°) with a force constant of 100 kcal/(mol.$degree^2$). Four RMSD-based SMD simulations were run for 10 ns each using a force constant of 50 kcal/(mol.Å^2): (1) to change the RMSD of the bound protein (0.5→2 Å). The RMSD of heparin was restrained in this simulation with a force constant of 1 kcal/(mol.Å^2); (2) to change the RMSD of the bound heparin (1.5→4 Å); (3) to change the RMSD of the unbound protein (0.8→3.2 Å); (4) to change the RMSD of the free heparin (1.5→5.5 Å).

Bias Exchange Umbrella Sampling (BEUS) simulations

Bias exchange umbrella sampling[62,98,99] (BEUS), which is a variation of the US simulation method, was performed to estimate grid PMF. Four independent sets of distance (d) based BEUS simulations were performed, with no restraints, restraint on Ω, restraint on r_L and r_P, and restraints on Ω, r_L, and r_P. Two sets of BEUS simulations were also performed using the Ω collective variable, one with and one without a restraint on r_L and r_P. In addition, two sets of BEUS simulations were performed using the r_P collective variable (bound protein with restraint on r_L; unbound protein) and two sets were performed using the r_L collective variable (bound ligand; free ligand). Selected SMD conformations were assigned to individual BEUS windows with equal spacing in each one of these BEUS simulations. The distance-based BEUS simulations ran for 10 ns with 31 replicas/windows

and the orientation-based simulations ran for 10 ns with 30 replicas/windows. The RMSD-based BEUS simulations ran for 10 ns with 12 replicas/windows. The force constant used for ligand-protein distance (d) in distance-based BEUS was 2 kcal/(mol.Å^2) while the orientation was restrained as in SMD simulations using a force constant of 0.5 kcal/(mol.$degree^2$). For orientation-based BEUS simulations, the force constant for the ligand orientation angle (as in SMD simulations) was set to 0.5 kcal/(mol.$degree^2$). The force constant used for r_L and r_P in all cases was 1 kcal/(mol.Å^2).

Free energy calculations using non-parametric reweighting

Once the BEUS simulations described above were converged, a non-parametric reweighting method[98,100], which is somewhat similar to the multi-state Bennett acceptance ratio method[101], was used to construct PMF. In this method[98], each sampled configuration will be assigned a weight, which can be used to construct the PMF in terms of a desired collective variable. Suppose that a system is biased (for instance, within a BEUS scheme) using N different biasing potentials $U_i(\boldsymbol{r})$, where $i = 1, \dots, N$, and $\boldsymbol{r}$ represents all atomic coordinates. Typically, $U_i(\boldsymbol{r})$ is a harmonic potential defined in terms of a collective variable with varying centers for different i. Assuming an equal number of sampled configurations from each of the N generated trajectories, we can combine them in a single set of samples $\{\boldsymbol{r}_k\}$ (irrespective of which bias was used to generate each sample $\boldsymbol{r}_k$) and determine the weight of each sample as:

$$w_k = c/\sum_i e^{-\beta(U_i(\boldsymbol{r}_k)-F_i)}$$

where, c is the normalization constant such that $\sum_k w_k = 1$ and both $\{w_k\}$ and $\{F_i\}$ are determined iteratively using the above equation and the following:

$$e^{-\beta F_i} = \sum_k w_k e^{-\beta U_i(\boldsymbol{r}_k)}$$

Converged w_k values can be used to construct any ensemble averages including any PMF (e.g., $G(\boldsymbol{\zeta})$) not only in terms of the collective variable used for biasing but also any other collective variables that are sufficiently sampled. One may use a weighted histogram method to construct the PMF as follows:

$$G(\boldsymbol{\zeta}_i) = -RT \ln \sum_k w_k \, \delta(\boldsymbol{\zeta}(\boldsymbol{r}_k) - \boldsymbol{\zeta}_i),$$

$$\delta(\boldsymbol{\zeta}(\boldsymbol{r}_k) - \boldsymbol{\zeta}_i) = \begin{cases} 1, & |\boldsymbol{\zeta}(\boldsymbol{r}_k) - \boldsymbol{\zeta}_i| < |\boldsymbol{\zeta}(\boldsymbol{r}_k) - \boldsymbol{\zeta}_j| \; for \; j \neq i \\ 0, & otherwise \end{cases}$$

To estimate the uncertainty of any of PMF calculations described above, one may use bootstrapping. Here, we have used a block Bayesian bootstrapping technique[99], where 100 alternative data sets are resampled from the existing data set and the same non-parametric reweighting algorithm and the same PMF calculation is repeated for each set to generate 100 alternative PMFs. The standard deviation of the PMF at any point along the reaction coordinate provides an estimate for the error.

Alchemical free energy perturbation (FEP) simulations

Alchemical FEP simulations were performed to calculate the absolute binding free energy for the interaction of hFGF1 with heparin hexasaccharide. We used a double annihilation protocol[102], wherein the heparin hexasaccharide is annihilated in both the free and bound states. The final conformations of the hFGF1-heparin complex[95] and free heparin hexasaccharide equilibrium simulations (discussed previously) were used to generate starting conformations for the bound hFGF1-heparin and free heparin FEP simulations respectively. Two sets of FEP simulations (short and long) were performed for both the hFGF1-heparin complex and the free heparin. The short FEP simulations were performed bi-directionally using 20 λ-windows. Each λ-window included a 0.5-ns equilibration and 2.5 ns of averaging for both the unbound and bound states, for a total of 240 ns. Similarly, the long FEP simulations were also performed bi-directionally using 50 λ-windows. Each λ-window included a 0.5-ns equilibration and 5 ns of averaging for both the unbound and bound states, for a total of 1.1 μs. A soft-core potential was used for vdW interactions with a radius-shifting coefficient of 5.0. The vdW and electrostatic interactions use a different schedule for changing, where the electrostatic interactions were completely decoupled at $\lambda = 0.5$. All FEP simulations were performed using the NAMD 2.13[88] simulation package with the CHARMM36m all-atom additive force field[89], using the protocol discussed previously for the equilibrium simulations. The ParseFEP plugin in VMD[103] was used to analyze the FEP data. Estimates for the absolute binding free energy calculations were obtained using the Bennett acceptance-ratio (BAR) method[104.]

RESULTS AND DISCUSSION

We have calculated the absolute binding free energy for the interaction of hFGF1 with heparin hexasaccharide using four variations of the stratification scheme described above, based on a combination of SMD and BEUS simulations. The details of the methodology are discussed in the Methods section. Four different methods are used with varying effectiveness in estimating the absolute binding free energy. These methods include (1) the traditional distance-based BEUS simulations that do not employ any additional restraining, (2) distance-based BEUS simulations employing a restraint on the orientation of the ligand (Ω) defined based on the orientation quaternion, (3) distance-based BEUS simulations employing a restraint on the RMSD of both ligand and protein ($\mathrm{r_L}, r_P$), (4) distance-based BEUS simulations employing a restraint on the RMSD of both ligand and protein as well as the orientation of the ligand ($\Omega, \mathrm{r_L}, r_P$). In each case, appropriate correction terms are calculated as discussed in the Theoretical Foundation section above and shown in Table 1.

Table 1: Summary of the results of free energy calculations.

	No restraints	Ω restraint	r_L, r_P restraint	Ω, r_L, r_P restraint
Grid PMF difference (kcal/mol)	$\Delta G(\boldsymbol{x}_B) = -19.7 \pm 1.1^{*}$	$\Delta G_{\Omega}(\boldsymbol{x}_B) = -13.2 \pm 0.3$	$\Delta G_{r_L, r_P}(\boldsymbol{x}_B) = -17.7 \pm 1.0$	$\Delta G_{\Omega, r_L, r_P}(\boldsymbol{x}_B) = -17.0 \pm 0.5$
Orientation correction (kcal/mol)	N/A	$\Delta U_{\Omega}(\boldsymbol{x}_B) = 4.4 \pm 0.3$	N/A	$\Delta U_{\Omega}^{r_L, r_P}(\boldsymbol{x}_B) = 4.4 \pm 0.3$
Ligand RMSD correction (kcal/mol)	N/A	N/A	$\Delta U_{\mathrm{r_L}}(\boldsymbol{x}_B) = 0.7 \pm 0.1$	$\Delta U_{\mathrm{r_L}}(\boldsymbol{x}_B) = 0.7 \pm 0.1$
Protein RMSD correction (kcal/mol)	N/A	N/A	$\Delta U_{\mathrm{rP}}^{r_L}(\boldsymbol{x}_B) = 0.4 \pm 0.1$	$\Delta U_{\mathrm{rP}}^{r_L}(\boldsymbol{x}_B) = 0.4 \pm 0.1$
ΔG_V (kcal/mol)	3.7 ± 0.2	2.5 ± 0.2	2.3 ± 0.2	2.7 ± 0.2
$\Delta G°$ (kcal/mol)	-16.0 ± 1.2	-6.3 ± 0.5	-14.3 ± 1.0	-8.7 ± 0.7
$K_d\ (\mu M)^{**}$	$O(10^{-6})$	25	$O(10^{-5})$	0.5
K_d range $(\mu M)^{***}$	$10^{-7} - 10^{-5}$	$11 - 58$	$10^{-6} - 10^{-4}$	$0.2 - 2.0$

* All error estimates are based on one standard deviation (s.d.). ** K_d values are determined directly from mean $\Delta G°$ values using Relation (2). *** K_d range is determined from the lower and upper limits of $\Delta G°$ values (mean $\pm$ s.d.) using Relation (2).

We denote the PMF of the ligand at a given position $\boldsymbol{x}$ (with respect to the center of the heparin binding pocket) as the grid PMF, since the PMF is estimated at different grid points in this approach (Figure 1). The average grid PMF profiles along the ligand-protein distance for the four different methods used here (as shown in Figure 1), confirm the differential behavior of these methods. The most successful method is expected to be the one employing restraints on $\Omega, \mathrm{r_L}, r_P$ (Table 1, Figure 1). The largest contributor to the free energy is the difference between the grid PMF associated with the heparin hexasaccharide at a grid point at the center of the binding pocket and at any grid point in the bulk, which is -17.0 ± 0.5 kcal/mol (Figure 2A, Table 1).

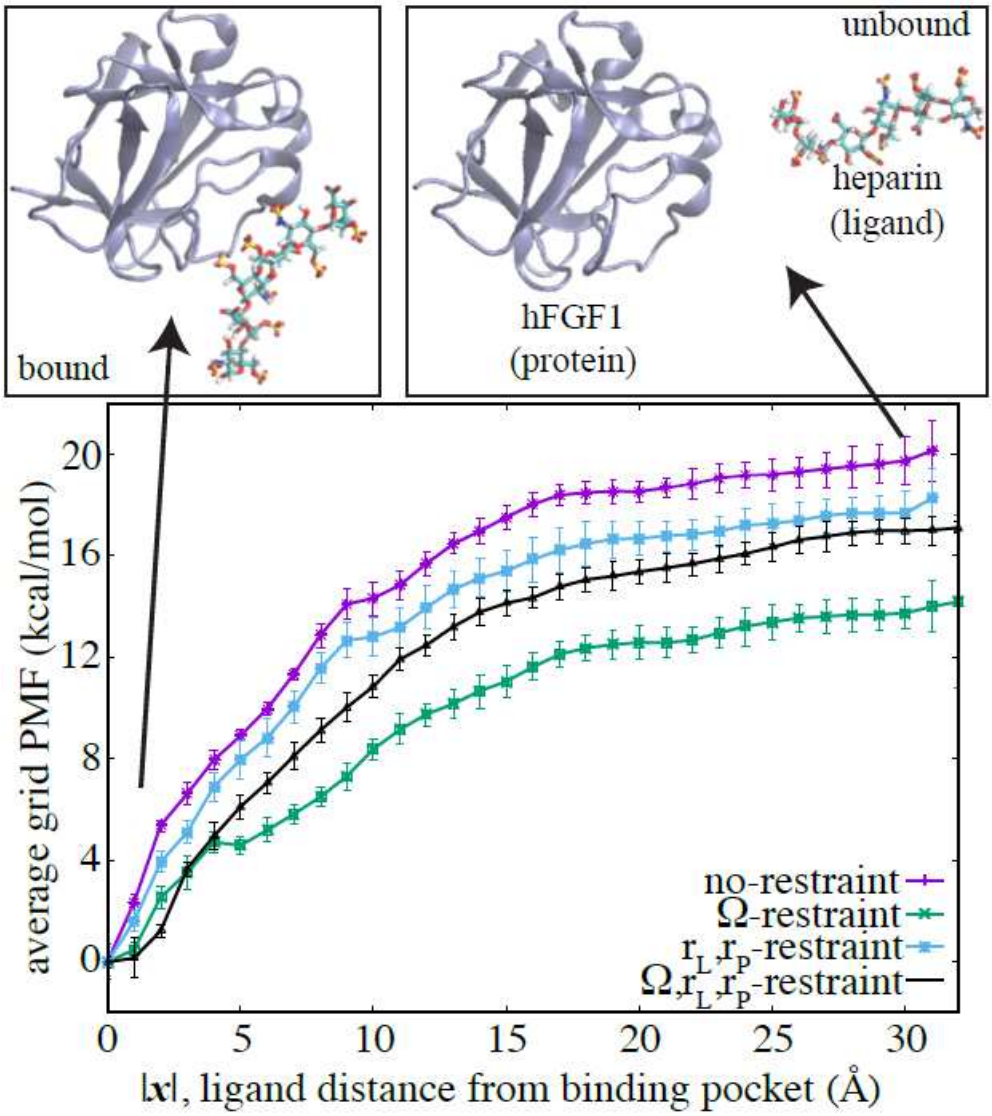


Figure 1. Average grid PMF based on four alternative distance-based BEUS simulations. Average grid PMFs in terms of $|\boldsymbol{x}|$,where $\boldsymbol{x}$ is the 3D position vector of the ligand with respect to the center of binding pocket. The x axis represents $|\boldsymbol{x}|$ and the y axis represents $\Delta F(|\boldsymbol{x}|)$, which is an average over all $\Delta F(\boldsymbol{x})$ with the same $|\boldsymbol{x}|$, i.e., the ligand distance from the center of binding pocket. The error bars represent the standard deviation obtained from all values of $\Delta F(\boldsymbol{x})$ at various grid points $\boldsymbol{x}$ with the same $|\boldsymbol{x}|$. Representative images of heparin hexasaccharide bound to hFGF1 ($|\boldsymbol{x}| = 0$ Å) and unbound ($|\boldsymbol{x}| = 30$ Å) are shown above the grid PMF plot.

The PMF calculations above are based on the BEUS simulations along the protein-ligand distance; however, the orientation and RMSD of the ligand and the RMSD of the protein are restrained to speed up convergence.

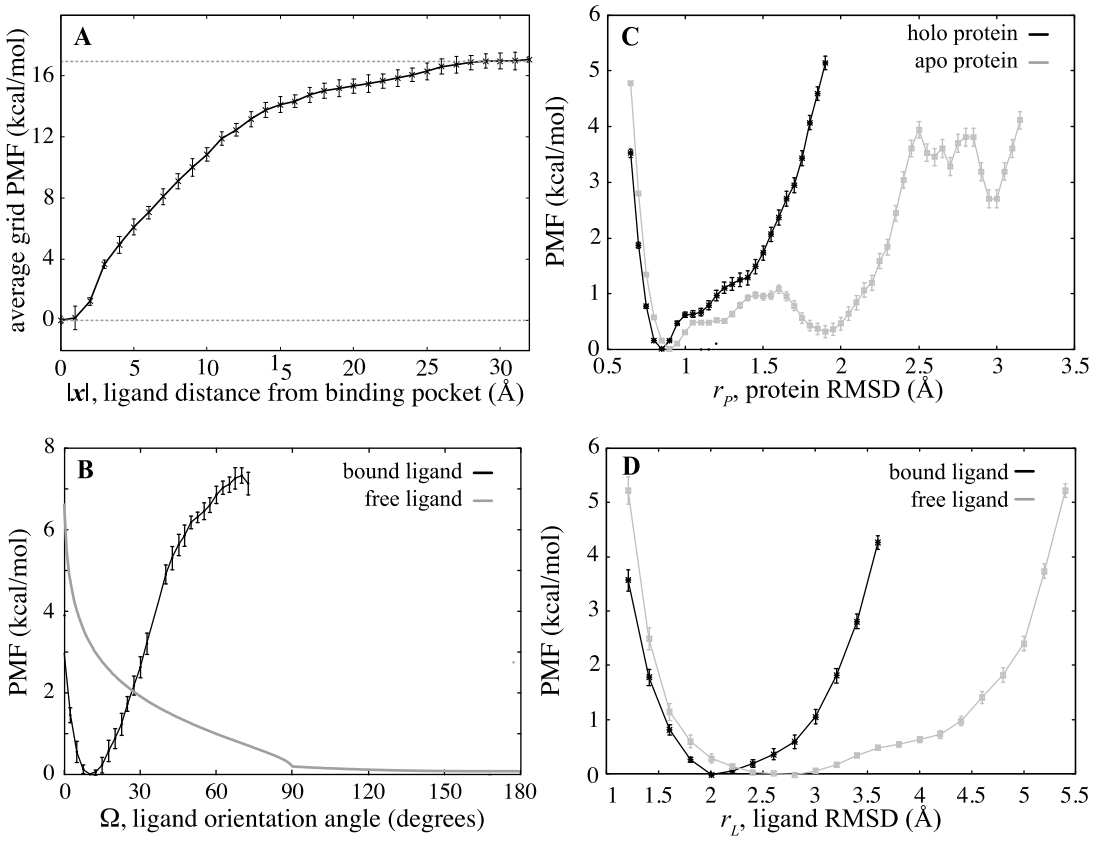


Figure 2. Computational heparin-hFGF1 binding free energy measurements. (A) Average grid PMF in terms of $|\boldsymbol{x}|$, where $\boldsymbol{x}$ is the 3D position vector of the ligand with respect to the center of binding pocket determined from distance-based BEUS simulations with Ω, r_L, r_P restraints. The x axis represents $|\boldsymbol{x}|$ and the y axis represents $\Delta G_{\Omega,r_L,r_P}(|\boldsymbol{x}|)$, which is an average over all $\Delta G_{\Omega,r_L,r_P}(\boldsymbol{x})$ with the same $|\boldsymbol{x}|$, i.e., the ligand distance from the center of binding pocket. The error bars represents the standard deviation obtained from all values of $\Delta G_{\Omega,r_L,r_P}(\boldsymbol{x})$ at various grid points $\boldsymbol{x}$ with the same $|\boldsymbol{x}|$. The dashed line represents the value associated with $\Delta G_{\Omega,r_L,r_P}(|\boldsymbol{x}|)$ at $|\boldsymbol{x}| = 30$ Å. (B) The PMF associated with the ligand orientation angle (Ω)

for the bound heparin (i.e., $\boldsymbol{x} \approx \mathbf{0}$, ligand in the binding pocket) and free heparin (i.e., $\boldsymbol{x} \approx \boldsymbol{x}_B$, ligand in the bulk). The latter is calculated analytically with the help of Relation (21). (C,D) PMF in terms of internal conformational fluctuations of the protein and ligand – (C) PMF associated with the internal RMSD of heparin-bound (solid line) and apo (dashed line) hFGF1 (r_P); (D) PMF associated with the internal RMSD of FGF1-bound (solid line) and free (dashed line) heparin hexasaccharide (r_L). The error bars in (B-D) represent the standard deviation determined from the bootstrapping algorithm described in Methods.

To account for the orientation bias, a correction term needs to be applied, which is calculated from the PMF associated with the ligand orientation angle at the bulk and binding pocket (Figure 2B). The orientation bias is estimated to be 4.4 ± 0.3 kcal/mol (Table 1). Similarly, a correction term is calculated based on the PMF of the ligand RMSD and that of the protein (Figure 2C). These correction terms are estimated to be 0.5 ± 0.1 and 0.8 ± 0.1 kcal/mol, for the ligand and protein, respectively (Table 1).

Finally, another term is needed to account for the difference in the volume accessible to the ligand in the binding pocket and in the bulk (volume contribution). Figure 3 shows that ΔG_P (or V_P) for the distance-based BEUS simulations with no restraint as determined from 20 lowest free energy grid points is almost equal to that obtained from all visited grid points inside or outside the binding pocket. For the distance-based BEUS simulations with $\Omega, \mathrm{r_L}, r_P$ restraints, this term is estimated to be 2.7 ± 0.2 kcal/mol (Table 1), which results in an absolute binding free energy of -8.5 ± 0.7 kcal/mol (Table 1).

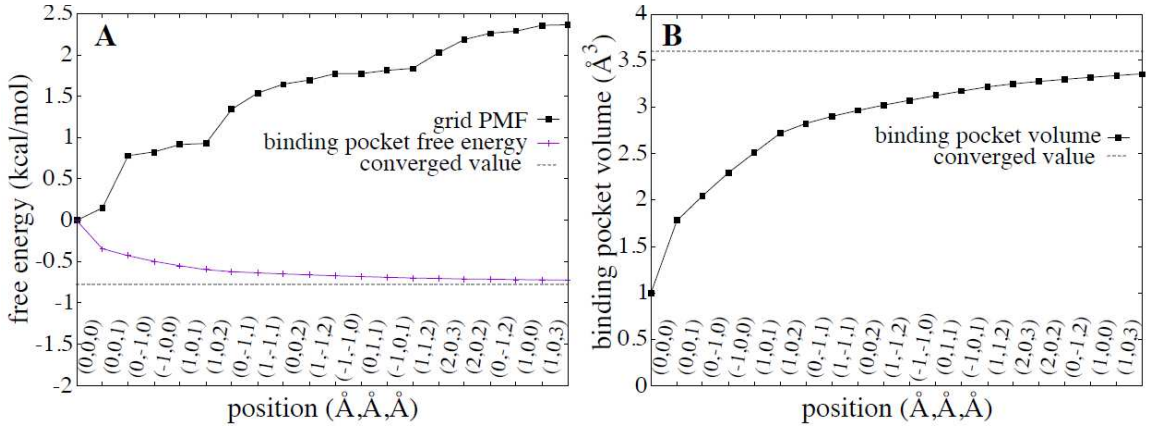


Figure 3. Estimating binding pocket volume (V_P) and its contribution to absolute binding free energy (ΔG_P). (A) Grid PMF ($\Delta G(\boldsymbol{x})$) associated with grid points with the 20 lowest PMF values (black) along with estimated ΔG_P based on the first 20 grid points (shown in an accumulative manner in magenta). The distance-based BEUS simulations with no restraints are used for these calculations. Dashed line shows the estimated ΔG_P based on all visited grid points inside or outside the binding pocket. The x axis shows the position vector of these 20 grid points. (B) Binding pocket volume (V_P) calculated from 20 lowest grid PMF values (similar to A). Dashed line shows the V_P estimated from all visited grid points inside or outside the binding pocket. See the Methods section for more details.

Based on our error analysis, K_d values calculated from the absolute binding free energy were found to be in the micromolar range with an average value of 0.6 μM (using the mean $\Delta G°$ estimate) and ranging from 0.2 to 2.0 μM (based on the lower and upper bounds of free energy estimates) (Table 1). These are in very good agreement with the K_d value obtained from ITC experiments. We performed the ITC experiments in triplicate resulting in a K_d of 1.68 ± 0.03 μM (as shown in Figure 4), 1.65 ± 0.07 μM, and 1.69 ± 0.05 μM in three independent experiments. The binding free energy calculated from the experimental K_d (-7.87, -7.88, or -7.89 kcal/mol depending on the experiment) is also in good agreement with the computationally calculated binding free energy (Figure 4 and Table 1).

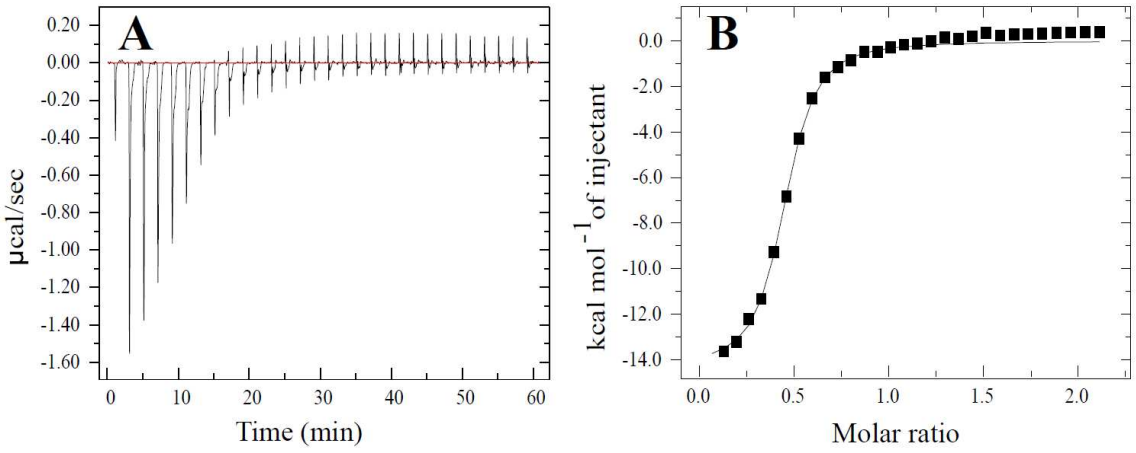


Figure 4. Experimental heparin-hFGF1 binding free energy measurements. Isothermogram representing the titration of hFGF1 with heparin hexasaccharide. Panel (A) represents the raw data and panel (B) represents the best-fit of the raw data (one-set of sites binding model). In the specific experiment shown, K_d is 1.68 ± 0.03 µM and the free energy from experimental K_d is -7.88 ± 0.01 kcal/mol.

The quantitative agreement between the computational and experimental binding affinity estimates is a great indicator of the accuracy of our absolute binding free energy calculation method. However, if proper restraining is not used as in the distance-based BEUS simulations with no restraints or only RMSD restraints, the binding affinity estimates would be off by several orders of magnitude. The simulations that only restrain the orientation of the ligand are interestingly quite successful as well, being off only by one order of magnitude in terms of binding affinity, which is generally considered a good estimate. This provides some evidence that the orientation of the ligand is perhaps the degree of freedom with the most significant contribution to the absolute binding free energy besides the ligand-protein distance. While the average grid PMF profiles along the ligand protein-distance (as shown in Figure 1) confirm that the four methods used here produce different results, it is important to note that the correction terms should ideally eliminate

these differences. This is seen to some extent when comparing the two methods involving orientation restraints that happen to estimate binding affinities that are reasonably close (Table 1) to the experimentally determined value.

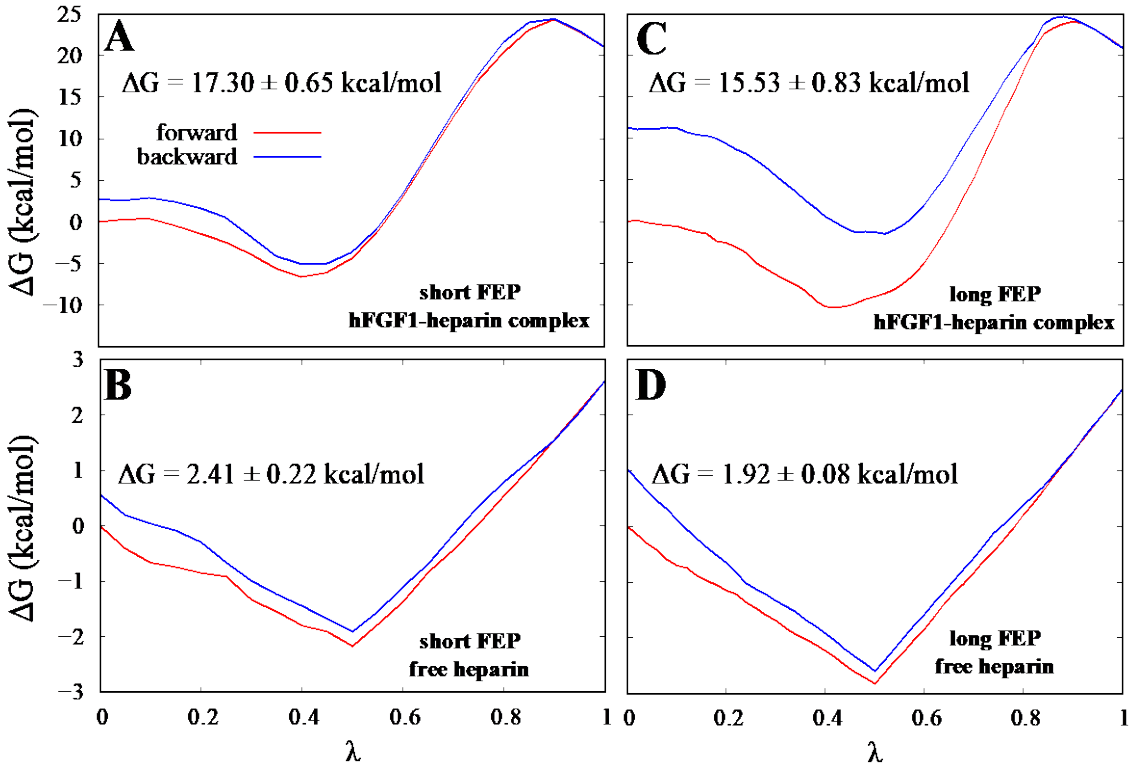


Figure 5. Binding free energy measurements from FEP simulations. (A-B) Free energy change for the double annihilation of heparin hexasaccharide in its bound state (A) and in its free state (B) - short FEP. The forward and backward transformations are shown in red and blue respectively. (C-D) Free energy change for the double annihilation of heparin hexasaccharide in its bound state (C) and in its free state (D) - long FEP. Inset: reported free energies and the error bars are based on the BAR method[104]. The error bars are not provided since the plots simply demonstrate the hysteresis and are not directly used to estimate the free energy and uncertainty (provided in the inset).

Recent computational studies have used the MM-GBSA method to calculate the binding free energy of the hFGF1-heparin interaction, with values ranging from -84.2 kcal/mol to -106.1 kcal/mol[105]. The results obtained from the MM-GBSA approach are very different from our own results, which is to be expected given that methods like MM-

PBSA/GBSA typically do not treat entropic or enthalpic contributions to the binding free energy rigorously[33,34,35]. On the other hand, these contributors to the binding affinity are typically taken into account during FEP simulations, thus obviating the need for the approximations used in MM-PBSA/GBSA[5,42]. It is widely accepted that binding free energy estimates from MM-PBSA/GBSA are less accurate than those from FEP, which is considered to be the gold standard for the calculation of absolute binding affinities[106]. We performed double annihilation FEP to calculate the absolute binding free energy of the hFGF1-heparin complex. The BAR method was used to estimate binding affinities from both the short (60×4 ns $= 240$ ns) and long (275×4 ns $= 1.1$ μs) FEP simulations. The long FEP simulations here were designed to have an aggregate simulation time similar to that used in our BEUS simulations. An absolute binding free energy of -14.9 ± 0.7 kcal/mol was obtained for the short FEP (Figure 5A-B), while an absolute binding free energy of -13.61 ± 0.8 kcal/mol was obtained for the long FEP simulations (Figure 5C-D). Unlike the absolute binding free energy estimated from the BEUS simulations (-8.7 ± 0.7 kcal/mol) (Table 1), neither of the estimates from the FEP simulations are in agreement with the binding free energy determined from ITC experiments (-7.88 ± 0.01 kcal/mol) (Figure 4). Hysteresis between the forward and backward directions increases during the long FEP simulations of the hFGF1-heparin complex (Figure 5C). Our results thus show that running longer FEP simulations does not necessarily result in a more accurate binding free energy estimate.

Studies have shown that the binding affinity and free energy results derived from computational methods can be compared to experimental binding affinities obtained from

ITC experiments[9,10]. However, for a reliable computational free energy estimate, employing purely physics-based free energy calculation methods such as those employed here has proven to be difficult. Herein, we showed that using a careful strategy that considers all relevant free energy terms and ensures the use of powerful enhanced sampling techniques, could result in good quantitative agreements between the computational and experimental binding affinity estimates.

The formalism presented in this work has significant similarities to the method previously proposed by Woo and Roux[3], and later implemented[4,66]. However, there are major differences that make the current method more practical. The grid PMF and its various estimates provide a simple conceptual framework to understand how restraining can be accounted for with appropriate correction terms. The average grid PMF in terms of the ligand-protein distance provides an alternative to the PMF in terms of d as is often constructed. Relation (30) is a general scheme that can be easily adapted to any number of restraints. The orientation angle of the ligand with respect to the protein as determined using the orientation quaternion formalism, provides a simple way of determining the absolute binding free energy with a feasible computational cost. Among the four different sets of restraints, the two involving orientation restraints predict binding free energies similar to that determined experimentally.

ACKNOWLEDGEMENTS

This research is supported by National Science Foundation grant CHE 1945465 and OAC 1940188 and the Arkansas Biosciences Institute. This work is also supported by the Department of Energy (DE-FG02-01ER15161) and the National Institutes of Health

(R15GM139140, R01CA172631, P30GM103450). Anton 2 computer time was provided by the Pittsburgh Supercomputing Center (PSC) through Grant R01GM116961 from the National Institutes of Health. The Anton 2 machine at PSC was generously made available by D.E. Shaw Research. This research is also part of the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) and the state of Illinois. This work also used the Extreme Science and Engineering Discovery Environment (allocation MCB150129), which is supported by National Science Foundation grant number ACI-1548562. This research is also supported by the Arkansas High Performance Computing Center, which is funded through multiple National Science Foundation grants and the Arkansas Economic Development Commission. TKSK was also supported by the National Institute of General Medical Sciences of the National Institutes of Health under award P20GM139768, and the Arkansas Integrative Metabolic Research Center at the University of Arkansas.

AUTHOR CONTRIBUTIONS

M.M. designed the research. V.G.K performed the simulations and analyzed the simulation data. T.K.S.K. designed the experiments. S.A. performed the experiments and analyzed the experimental data. V.G.K, M.M., T.K.S.K., and S.A. wrote the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

AVAILABILITY OF DATA

All data will be shared upon request to corresponding author.

AVAILABILITY OF CODE

All code will be shared upon request to corresponding author. Sample scripts as well as the full source code for non-parametric reweighting is available on github: https://github.com/bslgroup/binding_affinity_US.

REFERENCES

1. Mobley, D.L.; Gilson, M.K. Predicting binding free energies: frontiers and benchmarks. *Annu. Rev. Biophys*. **2017,** *46*, 531–58.
2. Wan, S.; Bhati, A.P.; Zasada, S.J.; Coveney, P.V. Rapid, accurate, precise and reproducible ligand–protein binding free energy prediction. *Interface Focus*. **2020**, *10*, 20200007.
3. Woo, H.-J.; Roux, B. Calculation of absolute protein–ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6825-6830.
4. Gumbart, J.C.; Roux, B.; Chipot, C. Standard binding free energies from computer simulations: What is the best strategy? *J. Chem. Theory Comput.* **2013**, *9***,** 794–802.
5. Siebenmorgen, T.; Zacharias, M. Computational prediction of protein–protein binding affinities. *WIREs Comput. Mol. Sci*. **2020**, *10*, e1448.
6. Kastritis, P.L., Bonvin, A.M. Are scoring functions in protein-protein docking ready to predict interactomes? Clues from a novel binding affinity benchmark. *J. Proteome Res*. **2010**, *9*, 2216–2225.
7. Gromiha, M.M.; Yugandhar, K.; Jemimah, S. Protein–protein interactions: Scoring schemes and binding affinity. *Curr. Opin. Struct. Biol*. **2017**, *44*, 31–38.

8. Du ,X.; Li, Y.; Xia, Y.L.; Ai, S.M.; Liang, J.; Sang, P.; Ji, X.L.; Liu, S.Q. Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods. *Int. J. Mol. Sci.* **2016**, *17*, 144.
9. Fenley, A. T.; Henriksen, N. M.; Muddana, H. S.; Gilson, M. K. Bridging Calorimetry and Simulation through Precise Calculations of Cucurbituril-Guest Binding Enthalpies. *J. Chem. Theory Comput.* **2014**, *10*, 4069–4078.
10. Talhout, R.; Villa, A.; Mark, A. E.; Engberts, J. B. F. N. Understanding binding affinity: A combined isothermal titration calorimetry/molecular dynamics study of the binding of a series of hydrophobically modified benzamidinium chloride inhibitors to trypsin. *J. Am. Chem. Soc*. **2003**, *125*, 10570-10579.
11. Weiss, S. Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy. *Nat. Struct. Biol.* **2000**, *7*, 724–729.
12. Rossi, A.; Taylor, C. Analysis of protein-ligand interactions by fluorescence polarization. *Nat. Protoc.* **2011**, *6*, 365–387.
13. Jönsson, U.; Fägerstam, L.; Ivarsson, B.; Johnsson, B.; Karlsson, R.; Lundh, K.; Löfås, S.; Persson, B.; Roos, H.; Rönnberg, I. et al. Real-time biospecific interaction analysis using surface plasmon resonance and a sensor chip technology. *Biotechniques.* **1991**, *11*, 620–627.
14. Torreri, P.; Ceccarini, M.; Macioce, P.; Petrucci, T.C. Biomolecular interactions by Surface Plasmon Resonance technology. *Ann. Ist. Super Sanita.* **2005**, *41*, 437–441.
15. Kastritis, P.L.; Moal, I.H.; Hwang, H. et al. A structure-based benchmark for protein–protein binding affinity. *Protein Sci*. **2011**, *20*, 482–491.

16. Dixit, S.B.; Chipot, C. Can absolute free energies of association be estimated from molecular mechanical simulations ? The biotin–streptavidin system revisited. *J. Phys. Chem. A.* **2001**, *105*, 9795–9799.
17. Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute binding free energies: A quantitative approach to their calculation. *J. Phys. Chem. B.* **2003**, *107*, 9535–9551.
18. Huang, D.; Caflisch, A. Efficient evaluation of binding free energy using continuum electrostatics solvation. *J. Med. Chem.* **2004**, *47*, 5791–5797.
19. Rodinger, T.; Howell, P.L.; Pomès, R. Absolute free energy calculations by thermodynamic integration in four spatial dimensions. *J. Chem. Phys.* **2005**, *123*, 34104.
20. Ytreberg, F.M.; Zuckerman, D.M. Simple estimation of absolute free energies for biomolecules. *J. Chem. Phys.* **2006**, *124*, 104105.
21. Wang, J.; Deng, Y.; Roux, B. Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophys. J.* **2006**, *91*, 2798–2814.
22. Rodinger, T.; Howell, P.L.; Pomès, R. Calculation of absolute protein-ligand binding free energy using distributed replica sampling. *J. Chem. Phys.* **2008**, *129*, 155102.
23. Deng, Y.; Roux, B. Computations of standard binding free energies with molecular dynamics simulations. *J. Phys. Chem. B.* **2009**, *113*, 2234–2246.
24. Boyce, S.E.; Mobley, D.L.; Rocklin, G.J.; Graves, A.P.; Dill, K.A.; Shoichet, B.K. Predicting ligand binding affinity with alchemical free energy methods in a polar model binding site. *J. Mol. Biol.* **2009**, *394*, 747–763.

25. Doudou, S.; Burton, N.A.; Henchman, R.H. Standard free energy of binding from a one-dimensional potential of mean force. *J. Chem. Theory Comput.* **2009**, *5*, 909–918.

26. Jiang, W.; Roux, B. Free energy perturbation Hamiltonian replica-exchange molecular dynamics (FEP/H-REMD) for absolute ligand binding free energy calculations. *J. Chem. Theory Comput.* **2010**, *6*, 2559–2565.

27. Singh, N.; Warshel, A. Absolute binding free energy calculations: On the accuracy of computational scoring of protein-ligand interactions. *Proteins.* **2010**, *78*, 1705–1723.

28. General, I.J.; Dragomirova, R.; Meirovitch, H. Absolute free energy of binding of avidin/biotin, revisited. *J. Phys. Chem. B.* **2012**, *116*, 6628–6636.

29. Zhang, C.; Liu, S.; Zhu, Q.; Zhou, Y. A knowledge-based energy function for protein−ligand, protein−protein, and protein−DNA complexes. *J. Med. Chem*. **2005**, *48*, 2325–2335.

30. Chéron, J.-B.; Zacharias, M.; Antonczak, S.; Fiorucci, S. Update of the ATTRACT force field for the prediction of protein–protein binding affinity. *J. Comput. Chem*. **2017**, *38*, 1887–1890.

31. Lensink, M.F.; Wodak, S.J. Docking and scoring protein interactions: CAPRI 2009. *Proteins Struct. Funct. Bioinf*. **2010**, *78*, 3073–3084.

32. Böhm, H.-J. Prediction of Non-bonded Interactions in Drug Design. *In Protein-Ligand Interactions*. Mannhold, R.; Kubinyi, H.; Folkers, G.; Böhm, H.-J.; Schneider, G., Eds.; 2003.

33. Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. *J. Am. Chem. Soc*. **1998**, *120*, 9401– 9409.

34. Wang, C. H.; Nguyen, P. H.; Pham, K.; Huynh, D.; Le, T. B. N.; Wang, H. L. et al. Calculating protein-ligand binding affinities with MMPBSA: method and error analysis. *J. Comput. Chem.* **2016**, *37*, 2436–2446.

35. Fu, H.; Gumbart, J. C.; Chen, H.; Shao, X.; Cai, W.; Chipot, C. BFEE: A User-Friendly Graphical Interface Facilitating Absolute Binding Free-Energy Calculations. *J. Chem. Inf. Model*. **2018**, *58*, 556–560.

36. Chipot, C. Frontiers in free-energy calculations of biological systems. In *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **2014**, *4*, 71– 89.

37. Chodera, J. D.; Mobley, D. L. Entropy-enthalpy compensation: Role and ramifications in biomolecular ligand recognition and design. *Annu. Rev. Biophys*. **2013**, *42*, 121– 142

38. *Free energy calculations. Theory and applications in chemistry and biology.* Chipot, C.; Pohorille, A., Eds; Springer Verlag: 2007.

39. Kollman, P.A. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.

40. 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*, 1047–1069.

41. Hermans, J.; Wang, L. Inclusion of loss of translational and rotational freedom in theoretical estimates of free energies of binding. Application to a complex of benzene and mutant T4 lysozyme. *J. Am. Chem. Soc.* **1997**, *119*, 2707–2714.

42. Fratev, F.; Sirimulla, S. An Improved Free Energy Perturbation FEP+ Sampling Protocol for Flexible Ligand-Binding Domains. *Sci. Rep.* **2019***, 9***,** 16829.

43. Jorgensen, W.L. Free-energy calculations: A breakthrough for modeling organic chemistry in solutions. *Acc. Chem. Res.* **1989**, *22*, 184–189.
44. Hermans, J.; Shankar, S. The free energy of xenon binding to myoglobin from molecular dynamics simulation. *Isr. J. Chem.* **1986**, *27*, 225–227.
45. Roux, B.; Nina, M.; Pomès, 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*, 670–681.
46. 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*, 084108.
47. Deng, Y.; Roux, B. Calculation of standard binding free energies: Aromatic molecules in the T4 lysozyme L99A mutant. *J. Chem. Theory Comput.* **2006**, *2*, 1255–1273.
48. Karplus, M.; McCammon, J.A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol*. **2002**, *9*, 646–652.
49. Dror, R.O.; Dirks, R.M.; Grossman, J.P.; Xu, H.; Shaw, D.E. Biomolecular simulation: A computational microscope for molecular biology. *Annu. Rev. Biophys*. **2012**, *41*, 429–452.
50. Immadisetty, K.; Hettige, J.; Moradi, M. What Can and Cannot Be Learned from Molecular Dynamics Simulations of Bacterial Proton-Coupled Oligopeptide Transporter GkPOT? *J. Phys. Chem. B* **2017**, *121*, 3644-3656.
51. van Gunsteren, W.F.; Mark, A.E. Validation of molecular dynamics simulation. *J. Chem. Phys*. **1998**, *108*, 6109–6116.

52. van Gunsteren, W.F.; Dolenc, J.; Mark, A.E. Molecular simulation as an aid to experimentalists. *Curr. Opin. Struct. Biol.* **2008**, *18*, 149–153.
53. Plattner, N.; Doerr, S.; De Fabritiis, G.; Noé, F. Complete protein–protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling. *Nat. Chem*. **2017**, *9*, 1005–1011.
54. Thayer, K. M.; Lakhani, B.; Beveridge, D. L. Molecular Dynamics–Markov State Model of Protein Ligand Binding and Allostery in CRIB-PDZ: Conformational Selection and Induced Fit. *J. Phys. Chem. B* **2017**, *121*, 5509–5514.
55. Chodera, J.D.; Noé, F. Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol*. **2014**, *25*, 135–144.
56. Zuckerman, D.M.; Chong, L.T. Weighted ensemble simulation: Review of methodology, applications, and software. *Annu. Rev. Biophys*. **2017**, *46*, 43–57.
57. Zwier, M.C.; Adelman, J.L.; Kaus, J.W. et al. WESTPA: An interoperable, highly scalable software package for weighted ensemble simulation and analysis. *J. Chem. Theory Comput.* **2015**, *11*, 800–809.
58. Kumar, S.; Rosenberg, J.M.; Bouzida, D.; Swendsen, R.H.; Kollman, P.A. The weighted histogram analysis method for free-energy calculations on biomolecules. *J. Comput. Chem*. **1992**, *13*, 1011–1021.
59. Souaille, M.; Roux, B. Extension to the weighted histogram analysis method: Combining umbrella sampling with free energy calculations. *Comput. Phys. Commun*. **2001**, *135*, 40–57.

60. Luitz, M.; Bomblies, R.; Ostermeir, K.; Zacharias, M. Exploring biomolecular dynamics and interactions using advanced sampling methods. *J. Phys. Condens. Matter*. **2015**, *27*, 323101.
61. Kokubo, H.; Tanaka, T.; Okamoto, Y. Ab initio prediction of protein–ligand binding structures by replica-exchange umbrella sampling simulations. *J. Comput. Chem*. **2011**, *32*, 2810–2821.
62. Moradi, M.; Tajkhorshid, E. Mechanistic picture for conformational transition of a membrane transporter at atomic resolution. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110,* 18916–18921.
63. Kästner, J.; Thiel, W. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: "Umbrella integration.". *J. Chem. Phys*. **2005**, *123*, 144104.
64. Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, 020603.
65. Comer, J.; Gumbart, J.C.; Hénin, J.; Lelièvre, T.; Pohorille, A.; Chipot, C. The adaptive biasing force method: Everything you always wanted to know but were afraid to ask. *J. Phys. Chem. B*. **2015**, *119*, 1129–1151.
66. Gumbart, J.C.; Roux, B.; Chipot, C. Efficient determination of protein–protein standard binding free energies from first principles. *J. Chem. Theory Comput.* **2013**, *9*, 3789–3798.
67. Eswarakumar, V. P.; Lax, I.; Schlessinger, J. Cellular Signaling by Fibroblast Growth Factor Receptors. *Cytokine Growth Factor Rev.* **2005**, *16*, 139-149.

68. Beenken, A.; Mohammadi, M. The FGF Family: Biology, Pathophysiology and Therapy. *Nat. Rev. Drug Discovery* **2009**, *8*, 235-253.

69. Kuro-o, M. Endocrine FGFs and Klothos: Emerging Concepts. *Trends Endocrinol. Metab.* **2008**, *19*, 239-245.

70. Ornitz, D. M.; Herr, A. B.; Nilsson, M.; Westman, J.; Svahn, C. M.; Waksman, G. FGF Binding and FGF Receptor Activation by Synthetic Heparan-Derived Di- and Trisaccharides. *Science* **1995**, *268*, 432-436.

71. Culajay, J. F.; Blaber, S. I.; Khurana, A.; Blaber, M. Thermodynamic Characterization of Mutants of Human Fibroblast Growth Factor 1 with an Increased Physiological Half-Life. *Biochemistry* **2000**, *39*, 7153-7158.

72. Mohammadi, M.; Olsen, S. K.; Ibrahimi, O. A. Structural Basis for Fibroblast Growth Factor Receptor Activation. *Cytokine Growth Factor Rev.* **2005**, *16*, 107-137.

73. Johnson, D. E.; Williams, L. T. Structural and Functional Diversity in the FGF Receptor Multigene Family. *Adv. Cancer Res.* **1993***, 60*, 1-41.

74. Lin, X.; Buff, E. M.; Perrimon, N.; Michelson, A. M. Heparan Sulfate Proteoglycans Are Essential for FGF Receptor Signaling during Drosophila Embryonic Development. *Development* **1999**, *126*, 3715-3723.

75. Yayon, A.; Klagsbrun, M.; Esko, J. D.; Leder, P.; Ornitz, D. M. Cell Surface, Heparin-like Molecules Are Required for Binding of Basic Fibroblast Growth Factor to Its High Affinity Receptor. *Cell* **1991**, *64*, 841-848.

76. Rapraeger, A. C.; Krufka, A.; Olwin, B. B. Requirement of Heparan Sulfate for BFGF-Mediated Fibroblast Growth and Myoblast Differentiation. *Science* **1991**, *252*, 1705-1708.

77. Taipale, J.; Keski-Oja, J. Growth Factors in the Extracellular Matrix. *FASEB J.* **1997**, *11*, 51-59.

78. McKeehan, W. L.; Wang, F.; Kan, M. The Heparan Sulfate-Fibroblast Growth Factor Family: Diversity of Structure and Function. *Prog. Nucleic Acid Res. Mol. Biol.* **1998**, *59*, 135-176.

79. Agostino, M.; Gandhi, N. S.; Mancera, R. L. Development and Application of Site Mapping Methods for the Design of Glycosaminoglycans. *Glycobiology* **2014**, *24*, 840-851.

80. Imberty, A.; Lortat-Jacob, H.; Pérez, S. Structural View of Glycosaminoglycan-Protein Interactions. *Carbohydr. Re*s. **2007**, *342*, 430-439.

81. Babik, S.; Samsonov, S. A.; Pisabarro, M. T. Computational Drill down on FGF1-Heparin Interactions through Methodological Evaluation. *Glycoconj. J.* **2017***, 34*, 427-440.

82. Muñoz-García, J. C.; García-Jiménez, M. J.; Carrero, P.; Canales, A.; Jiménez-Barbero, J.; Martín-Lomas, M.; Imberty, A.; de Paz, J. L.; Angulo, J.; Lortat-Jacob, H. et al. Importance of the Polarity of the Glycosaminoglycan Chain on the Interaction with FGF-1. *Glycobiology* **2014**, *24* 1004-1009.

83. Carter, E. P.; Fearon, A. E.; Grose, R. P. Careless Talk Costs Lives: Fibroblast Growth Factor Receptor Signalling and the Consequences of Pathway Malfunction. *Trends Cell Biol*. **2015,** *25*, 221-233.

84. Goetz, R.; Mohammadi, M. Exploring Mechanisms of FGF Signalling through the Lens of Structural Biology. *Nat. Rev. Mol. Cell Biol*. **2013**, *14*, 166-180.

85. Le, V. H.; Buscaglia, R.; Chaires, J. B.; Lewis, E. A. Modeling complex equilibria in isothermal titration calorimetry experiments: thermodynamic parameters estimation for a three-binding-site model. *Anal. Biochem.* **2013**, *434*, 233–241.

86. DiGabriele, A. D.; Lax, I.; Chen, D. I.; Svahn, C. M.; Jaye, M.; Schlessinger, J.; Hendrickson, W. A. Structure of a Heparin-Linked Biologically Active Dimer of Fibroblast Growth Factor. *Nature* **1998**, *393*, 812-817.

87. Bernett, M. J.; Somasundaram, T.; Blaber, M. An Atomic Resolution Structure for Human Fibroblast Growth Factor 1. *Proteins* **2004**, *57*, 626-634.

88. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem*. **2005**, *26*, 1781-1802.

89. Huang, J.; Rauscher, S.; Nawrocki, G. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods.* **2017**, *14*, 71–73.

90. *Large Sparse Sets of Linear Equations*. Reid, J.K., Ed.; Academic Press: London and New York, 1971.

91. Jo, S.; Kim, T.; Im, W. Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLoS One* **2007**, 2, e880.

92. Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y. et al. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **2016**, *12*, 405-413.

93. Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys.* **1994**, *101*, 4177-4189.

94. Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.

95. Govind Kumar, V.; Agrawal, S.; Kumar, T. K. S.; Moradi, M. Mechanistic Picture for Monomeric Human Fibroblast Growth Factor 1 Stabilization by Heparin Binding. *J. Phys. Chem. B*. **2021,** *125***,** 12690-12697.

96. Fiorin, G.; Klein, M. L.; Henin, J. Using collective variables to drive molecular dynamics simulations. *Mol. Phys.* **2013***, 111*, 3345-3362.

97. Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.* **1997***, 72*, 1568-1581.

98. Moradi, M.; Tajkhorshid, E. Computational recipe for efficient description of large-scale conformational changes in biomolecular systems. *J. Chem. Theory Comput.* **2014**, *10,* 2866–2880.

99. Moradi, M.; Enkavi, G.; Tajkhorshid, E. Atomic-level characterization of transport cycle thermo-dynamics in the glycerol-3-phosphate:phosphate transporter. *Nat. Commun.* **2015**, *6*, 8393.

100. Bartels, C. Analyzing biased Monte Carlo and molecular dynamics simulations. *Chem. Phys. Lett.* **2000***, 331*, 446-454.

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

102. 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, 1047–1069.

103. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics*. **1996**, *14*, 33– 38.

104. Bennett, C.H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245-268.

105. Bojarski, K. K.; Sieradzan, A. K.; Samsonov, S. A. Molecular dynamics insights into protein-glycosaminoglycan systems from microsecond-scale simulations. *Biopolymers.* **2019***, 110*, e23252.

106. Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov*. **2015**, *10*, 449-461.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- MoradiCSflat.pdf