

# Binding affinity estimation from restrained umbrella sampling simulations

Received: 11 March 2022

Accepted: 5 December 2022

Published online: 29 December 2022

 Check for updates

Vivek Govind Kumar, Adithya Polasa , Shilpi Agrawal, Thallapuram Krishnaswamy Suresh Kumar & Mahmoud Moradi  

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 simulations. Our proposed method generalizes and simplifies previously suggested stratification strategies that use umbrella sampling or other enhanced sampling simulations with additional collective-variable-based restraints. The 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 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 experiments.

Accurate quantification of absolute binding affinities remains a problem of major importance in computational biophysics<sup>1–4</sup>. 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<sup>3–5</sup>. Experimentally determined binding affinities are commonly used as benchmarks to judge the accuracy of various computational binding affinity estimation methods<sup>5</sup>. Several experimental techniques can be used to study protein–ligand binding equilibria<sup>5,6</sup>. For instance, isothermal titration calorimetry (ITC) can detect the interaction of binding partners based on changes in solution heat capacity and binding partner concentration<sup>6–8</sup>. Other methods such as fluorescence spectroscopy rely on changes in fluorescence intensity upon ligand binding<sup>6,9,10</sup>. Surface plasmon resonance 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<sup>6</sup>. Studies have found that experimental binding affinities can vary depending on the experimental method used<sup>5</sup>. 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<sup>3,11–18</sup>. 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<sup>19–21</sup>. 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<sup>5</sup>. This is also the case for methods such as 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 molecular dynamics (MD) simulations with free-energy estimation based on implicit continuum solvent models<sup>22–24</sup>. 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<sup>24–26</sup>. MM-PBSA and MM-GBSA methods typically neglect such entropic contributions to the binding free energy or do not treat them rigorously<sup>23,24</sup>. Binding Free-Energy Estimator 2 (BFEE2)

is a state-of-the-art protein–ligand binding affinity calculation software that addresses the substantial shift in configurational enthalpy and entropy that follows ligand–protein binding, which is hard to represent in brute-force simulations<sup>18,27</sup>. An energy–entropy approach, energy–entropy multiscale cell correlation, has been introduced to compute the free energy of binding and has been applied for binding-free energy calculations, which take into consideration the entropy of all flexible degrees of freedom in the system in a consistent and generic way<sup>28</sup>.

One of the best-known binding-free-energy estimation methods is alchemical free-energy perturbation (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<sup>29–32</sup>. Most entropic and enthalpic contributors to changes in binding affinity are typically considered during FEP simulations, thus avoiding the approximations used by methods such as MM-PBSA and MM-GBSA<sup>5,33</sup>. 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<sup>34</sup>. Using receptor–ligand restraints to resolve this issue<sup>30,35–37</sup> 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<sup>38</sup>. This can be corrected by the use of appropriate geometrical restraints<sup>39–41</sup>.

Unrestrained long-timescale MD simulations should theoretically allow for the investigation and quantification of protein–ligand or protein–protein binding events<sup>42,43</sup>. While microsecond-level MD simulations provide a more accurate description of protein conformational dynamics compared with shorter simulations<sup>44</sup>, efficient sampling of the conformational landscape remains a major issue and requires access to timescales beyond the capabilities of current MD simulations<sup>45,46</sup>. 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<sup>57</sup>. Similarly, weighted ensemble simulations sample the conformational landscape along one or more discretized reaction coordinates based on the assignment of a statistical weight to each simulation<sup>47,48</sup>. More traditionally, umbrella sampling (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 can be used to calculate a unidimensional potential of mean force (PMF) that quantifies ligand binding and unbinding along a reaction coordinate<sup>49,50</sup>. 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)<sup>51–53</sup>. Other methods based on similar principles include umbrella integration<sup>54</sup>, well-tempered metadynamics<sup>55</sup>, adaptive biasing force (ABF) simulations<sup>56</sup> 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<sup>3,4</sup>. To resolve this problem, ref. 3 reported 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<sup>3,4</sup>. This results in improved convergence of the PMF calculation<sup>3,4</sup>. 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<sup>3,4</sup>. This method avoids the need to decouple the ligand from its surrounding environment as required by alchemical FEP<sup>3,4,29–32</sup>. Recent studies<sup>4,57</sup> have described applications and extensions of the methodology proposed by ref. 3.

In this Article, 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 refs. 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 refs. 3,4. Our method includes: (1) providing a general scheme that can be easily adapted to any number of restraints; (2) the non-parametric reconstruction of the grid PMF, as defined below; and (3) 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 note that the method of refs. 3,4 can in principle be generalized as well; the generalization is not as straightforward as it is in our proposed method, particularly in removing some of the restraints. In other words, while adding more restraints is somewhat similar to our approach in the method of refs. 3,4, removing some of the restraint requires less trivial changes to the formalism that makes it distinct from our method. 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<sup>58–62</sup>. GAGs are linear anionic polysaccharides that interact with positively charged regions of FGF binding partners to regulate their biological activity<sup>63</sup>. The hFGF1–heparin complex is the most well-known and broadly characterized protein–GAG complex<sup>64,65</sup>. 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 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 with short FEP simulations, our approach is still more accurate than FEP when similar simulation times are used.

## Results

### Calculation of binding free energy using four different strategies

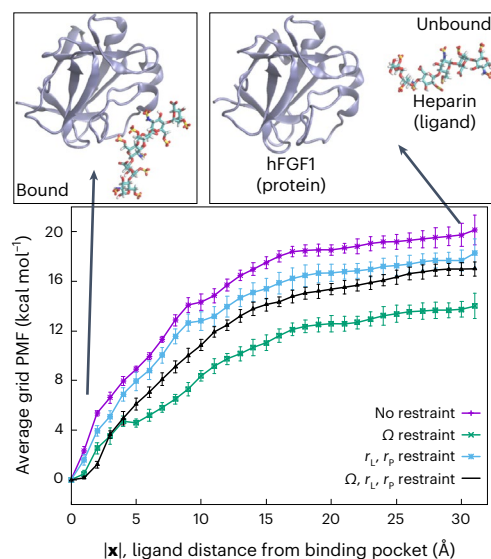
We have calculated the absolute binding free energy for the interaction of hFGF1 with heparin hexasaccharide using four variations of the stratification scheme described above, based on a combination of steered MD (SMD) and BEUS simulations. The details of the methodology are discussed in Methods. Four different methods are used with varying effectiveness in estimating the absolute binding free energy. These methods are: (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 ( $\Omega$ ) defined based on the orientation quaternion formalism; (3) distance-based BEUS simulations employing a restraint on the RMSD of both ligand and protein ( $r_l$  and  $r_p$ ); (4) distance-based BEUS simulations employing a restraint on the RMSD of both ligand and protein as well as the orientation of the ligand ( $\Omega$ ,  $r_l$  and  $r_p$ ). In each case, appropriate correction terms are calculated as discussed in the ‘Theoretical foundation’ section and shown in Table 1.

We denote the PMF of the ligand at a given position  $\mathbf{x}$  (with respect to the center of the heparin binding pocket) as the grid PMF, as the PMF is estimated at different grid points in this approach (Fig. 1). The average grid PMF profiles along the ligand–protein distance for the four different methods used here (as shown in Fig. 1) confirm the differential behavior of these methods (see Supplementary Fig. 1 for a schematic representation of these simulations). Note that since  $\mathbf{x} = \mathbf{0}$  is the grid point associated with the lowest PMF by definition, the average PMF along  $|\mathbf{x}|$  has its global minimum at  $|\mathbf{x}| = 0$ . The most successful method

**Table 1 | Summary of the results of free-energy calculations**

Quantity	No restraints	$\Omega$ restraint	$r_L$ and $r_p$ restraint	$\Omega$ , $r_L$ and $r_p$ restraint
Grid PMF difference (kcal mol <sup>-1</sup> )	$\Delta G(\mathbf{x}_B) = -19.7 \pm 1.1^a$	$\Delta G_{\Omega}(\mathbf{x}_B) = -13.2 \pm 0.3$	$\Delta G_{r_L, r_p}(\mathbf{x}_B) = -17.7 \pm 1.0$	$\Delta G_{\Omega, r_L, r_p}(\mathbf{x}_B) = -17.0 \pm 0.5$
Orientation correction (kcal mol <sup>-1</sup> )	NA	$\Delta U_{\Omega}(\mathbf{x}_B) = 4.4 \pm 0.3$	NA	$\Delta U_{\Omega, r_L, r_p}(\mathbf{x}_B) = 4.6 \pm 0.3$
Ligand RMSD correction (kcal mol <sup>-1</sup> )	NA	N/A	$\Delta U_{r_L}(\mathbf{x}_B) = 0.6 \pm 0.1$	$\Delta U_{r_L}(\mathbf{x}_B) = 0.6 \pm 0.1$
Protein RMSD correction (kcal mol <sup>-1</sup> )	NA	NA	$\Delta U_{r_p}^1(\mathbf{x}_B) = 0.3 \pm 0.1$	$\Delta U_{r_p}^1(\mathbf{x}_B) = 0.3 \pm 0.1$
$\Delta G_V$ (kcal mol <sup>-1</sup> )	$3.7 \pm 0.2$	$2.5 \pm 0.2$	$2.3 \pm 0.2$	$2.7 \pm 0.2$
$\Delta G^\circ$ (kcal mol <sup>-1</sup> )	$-16.0 \pm 1.2$	$-6.3 \pm 0.5$	$-14.5 \pm 1.0$	$-8.7 \pm 0.7$
$K_d$ ( $\mu\text{M}$ ) <sup>b</sup>	$O(10^{-6})$	25	$O(10^{-5})$	0.5
$K_d$ range ( $\mu\text{M}$ ) <sup>c</sup>	$10^{-7}$ – $10^{-5}$	11–58	$10^{-6}$ – $10^{-4}$	0.2–2.0

We are comparing the  $\Delta G^\circ$  of four different restraining methods (see 'BEUS simulations' in Methods for details). <sup>a</sup>All error estimates are based on 1 s.d. <sup>b</sup>Equilibrium dissociation constant ( $K_d$ ) values are determined directly from mean absolute binding-free-energy ( $\Delta G^\circ$ ) values using relation (2). The experimentally determined  $K_d$  and  $\Delta G^\circ$  were  $1.68 \pm 0.03 \mu\text{M}$  and  $-7.88 \pm 0.01 \text{ kcal mol}^{-1}$ , respectively (see Fig. 4). The orientation angle of heparin with respect to the protein ( $\Omega$ ), RMSD of the protein ( $r_p$ ), RMSD of heparin ( $r_L$ ), and the contribution of the difference between the volume of the binding pocket and the bulk to the binding free energy ( $\Delta G_V$ ),  $\Delta G(\mathbf{x}_B)$ ,  $\Delta G_{\Omega}(\mathbf{x}_B)$ ,  $\Delta G_{r_L, r_p}(\mathbf{x}_B)$  and  $\Delta G_{\Omega, r_L, r_p}(\mathbf{x}_B)$  are the PMF difference between the binding pocket center and the bulk associated with respective restraints. NA means the data are not applicable in the corresponding section.  $\Delta U_{\Omega}(\mathbf{x}_B)$ ,  $\Delta U_{\Omega, r_L, r_p}(\mathbf{x}_B)$ ,  $\Delta U_{r_L}(\mathbf{x}_B)$  and  $\Delta U_{r_p}^1(\mathbf{x}_B)$  are correction terms associated with restraints (see 'Theoretical foundation' in Methods for details).



**Fig. 1 | Average grid PMF based on our alternative distance-based BEUS simulations.** Average grid PMFs in terms of  $|x|$ , where  $\mathbf{x}$  is the three-dimensional position vector of the ligand with respect to the center of the binding pocket. The  $x$  axis represents  $|x|$  and the  $y$  axis represents  $\Delta F(|x|)$ , which is an average over all  $\Delta F(\mathbf{x})$  with the same  $|x|$ , that is, the ligand distance from the center of the binding pocket. The error bars represent the standard deviation obtained from all values of  $\Delta F(\mathbf{x})$  at various grid points  $\mathbf{x}$  with the same  $|x|$ . Representative images of heparin hexasaccharide bound to hFGF1 ( $|x| = 0 \text{ \AA}$ ) and unbound ( $|x| = 30 \text{ \AA}$ ) are shown above the grid PMF plot. We used  $n = 31$  MD replicas to generate the data in all cases.

is expected to be the one employing restraints on  $\Omega$ ,  $r_L$  and  $r_p$  (Table 1 and Fig. 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  $-17.0 \pm 0.5 \text{ kcal mol}^{-1}$  (Fig. 2a and Table 1).

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. 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 (Fig. 2b). The orientation bias is estimated to be  $4.4 \pm 0.3 \text{ kcal mol}^{-1}$  (Table 1).

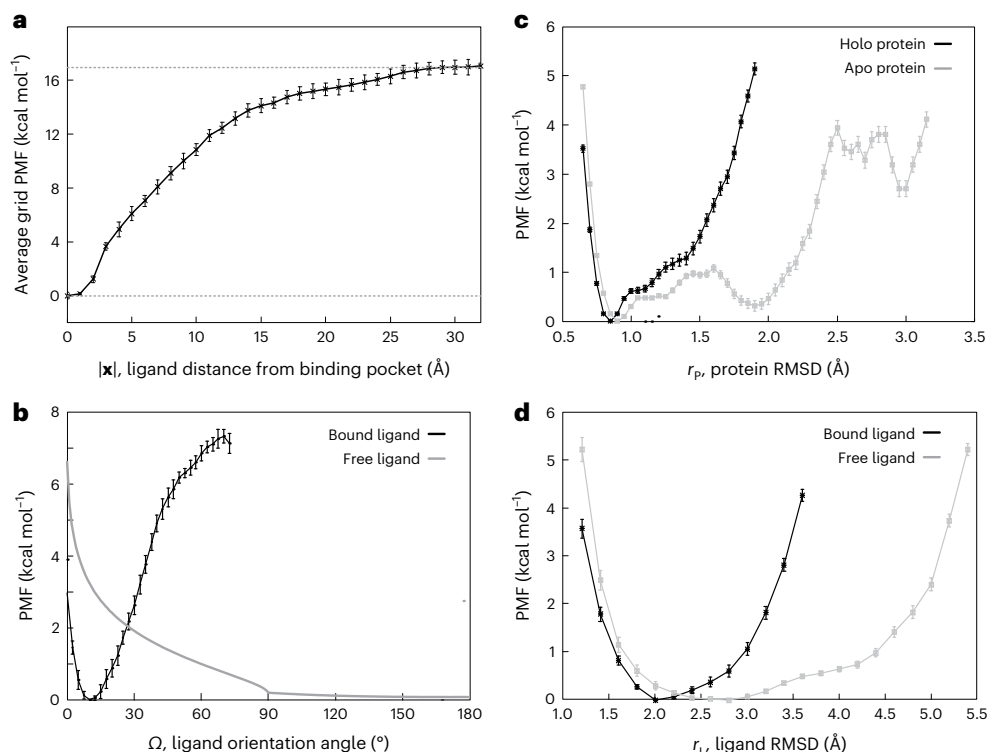
Similarly, a correction term is calculated based on the PMF of the ligand RMSD and that of the protein (Fig. 2c). These correction terms are estimated to be  $0.7 \pm 0.1 \text{ kcal mol}^{-1}$  and  $0.4 \pm 0.1 \text{ kcal mol}^{-1}$  for the ligand and protein, respectively (Table 1).

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 the binding pocket contribution ( $\Delta G_V$ ) (or binding pocket volume ( $V_p$ )) for the distance-based BEUS simulations with no restraint as determined from the 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  $\Omega$ ,  $r_L$  and  $r_p$  restraints, this term is estimated to be  $2.7 \pm 0.2 \text{ kcal mol}^{-1}$  (Table 1), which results in an absolute binding free energy of  $-8.7 \pm 0.7 \text{ kcal mol}^{-1}$  (Table 1).

On the basis of our error analysis, equilibrium dissociation constant ( $K_d$ ) values calculated from the absolute binding free energy were found to be in the micromolar range with an average value of  $0.6 \mu\text{M}$  (using the mean absolute binding free energy ( $\Delta G^\circ$ ) estimate) and ranging from  $0.2 \mu\text{M}$  to  $2.0 \mu\text{M}$  (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  $1.68 \pm 0.03 \mu\text{M}$  (as shown in Fig. 4),  $1.65 \pm 0.07 \mu\text{M}$  and  $1.69 \pm 0.05 \mu\text{M}$  in three independent experiments. The binding free energy calculated from the experimental  $K_d$  ( $-7.87 \text{ kcal mol}^{-1}$ ,  $-7.88 \text{ kcal mol}^{-1}$  or  $-7.89 \text{ kcal mol}^{-1}$ , depending on the experiment) is also in good agreement with the computationally calculated binding free energy (Fig. 4 and Table 1).

### Comparison between computationally and experimentally calculated binding free energy of heparin–hFGF1

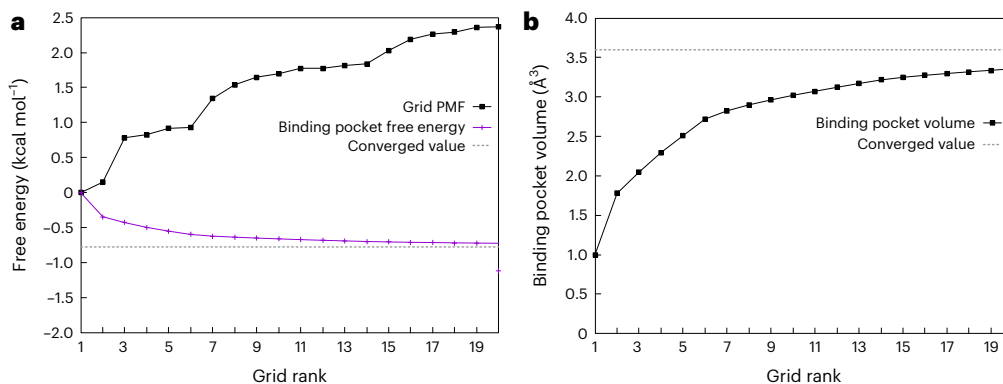
The quantitative agreement between the computational and experimental binding affinity estimates is a great indicator of the accuracy of our absolute binding-free-energy calculation method. However, if proper restraining is not used as in the distance-based BEUS simulations with no restraints or only RMSD restraints, the binding affinity estimates would be off by several orders of magnitude. The simulations that restrain only the orientation of the ligand are interestingly quite successful as well, being off by only 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 substantial 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 Fig. 1) confirm that the four methods used here produce different



**Fig. 2 | Computational heparin-hFGF1 binding-free-energy measurements.**

**a**, Average grid PMF in terms of  $|\mathbf{x}|$ , where  $\mathbf{x}$  is the three-dimensional position vector of the ligand with respect to the center of the binding pocket determined from distance-based BEUS simulations with  $\Omega$ ,  $r_L$  and  $r_p$  restraints. The  $x$  axis represents  $|\mathbf{x}|$  and the  $y$  axis represents  $\Delta G_{\Omega, r_L, r_p}(|\mathbf{x}|)$ , which is an average over all  $\Delta G_{\Omega, r_L, r_p}(\mathbf{x})$  with the same  $|\mathbf{x}|$ , that is, the ligand distance from the center of the binding pocket. The error bars represent the standard deviation obtained from all values of  $\Delta G_{\Omega, r_L, r_p}(\mathbf{x})$  at various grid points  $\mathbf{x}$  with the same  $|\mathbf{x}|$ . The dashed line represents the value associated with  $\Delta G_{\Omega, r_L, r_p}(|\mathbf{x}|)$  at  $|\mathbf{x}| = 30 \text{ \AA}$ . **b**, The PMF associated with the ligand orientation angle ( $\Omega$ ) for the bound heparin (that is,

$\mathbf{x} \approx \mathbf{0}$ , ligand in the binding pocket) and free heparin (that is,  $\mathbf{x} \approx \mathbf{x}_B$ , 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 (black line) and apo (gray line) hFGF1 ( $r_p$ ). **d**, PMF associated with the internal RMSD of FGF1-bound (black line) and free (gray line) heparin hexasaccharide ( $r_L$ ). The error bars in **b–d** represent the standard deviation determined from the bootstrapping algorithm described in Methods. We used  $n = 31$ ,  $n = 30$  and  $n = 12$  MD replicas to generate the data shown in **a**, **b**, and **c** and **d**, respectively.

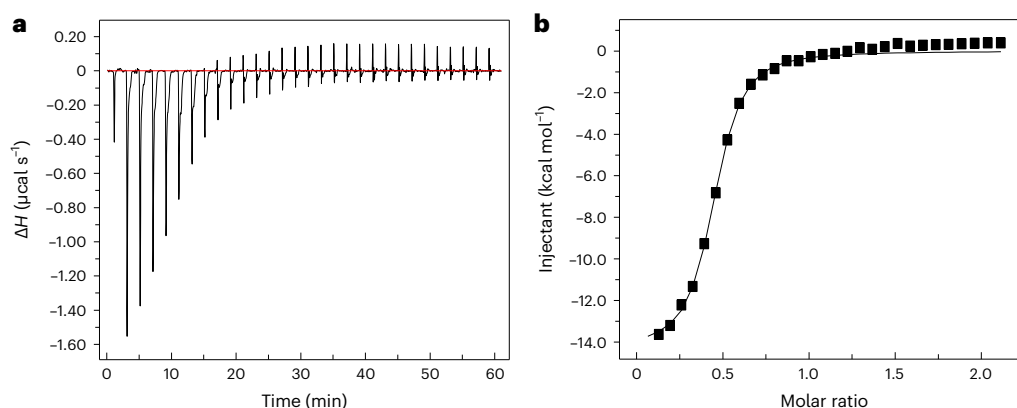


**Fig. 3 | Estimating binding pocket volume and its contribution to absolute binding free energy.** **a**, Grid PMF ( $\Delta G(\mathbf{x})$ ) associated with grid points with the 20-lowest PMF values (black) along with estimated  $\Delta G_p$  based on these grid points (shown in an accumulative manner in magenta). The distance-based BEUS simulations with no restraints are used for these calculations. The dashed line shows the estimated  $\Delta G_p$  based on all visited grid points inside or outside the

binding pocket. The  $x$  axis shows the ranking of these 20 grid points from lowest PMF onwards. **b**, Binding pocket volume ( $V_p$ ) calculated from the 20-lowest grid PMF values (similar to **a**). The dashed line shows the  $V_p$  estimated from all visited grid points inside or outside the binding pocket. See 'Theoretical foundation' in Methods for more details.

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. Another source of error in our calculations

could be in estimating the  $V_p$  and eventually the contribution of the difference between the volume of the binding pocket and the bulk to the binding free energy ( $\Delta G_v$ ). In doing so, we have made an assumption that the  $V_p$  can be calculated from relation (31) approximating the grid PMF with that obtained from biased simulations. Comparing



**Fig. 4 | Experimental heparin–hFGF1 binding-free-energy measurements.** Isothermogram representing the titration of hFGF1 with heparin hexasaccharide. **a**, The raw data. The horizontal red line represents the zero

axis in the plot. **b**, The best fit of the raw data (one set of sites binding model). In the specific experiment shown,  $K_d$  is  $1.68 \pm 0.03 \mu\text{M}$  and the free energy from experimental  $K_d$  is  $-7.88 \pm 0.01 \text{ kcal mol}^{-1}$ .  $\Delta H$ , heat change.

$\Delta G_v$  values from Table 1 shows that different biases result in different approximating values ranging from  $2.3 \text{ kcal mol}^{-1}$  to  $3.7 \text{ kcal mol}^{-1}$ . For more information on these results and the convergence of data, see Supplementary Table 1 and Figs. 2–5 ref. 27.

### Examining how this approach compares with other prevalent binding-free-energy calculation methods

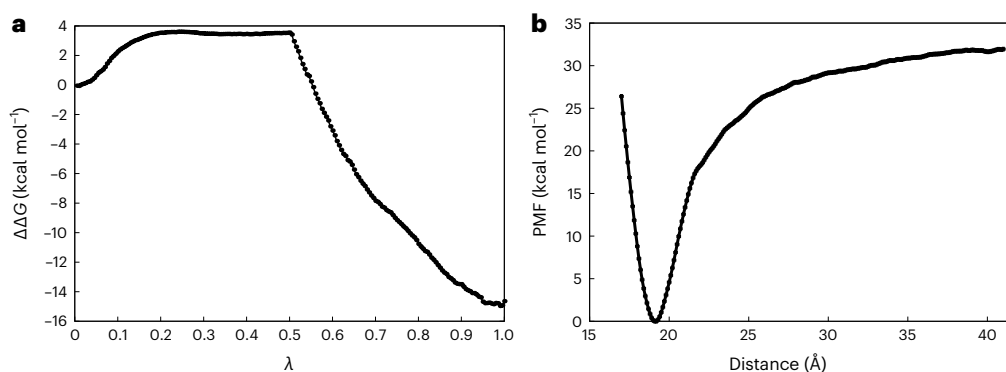
Recent computational studies have used the MM-GBSA method to calculate the binding free energy of the hFGF1–heparin interaction, with values ranging from  $-84.2 \text{ kcal mol}^{-1}$  to  $-106.1 \text{ kcal mol}^{-1}$  (ref. 66). The outcomes of the MM-GBSA technique are considerably different from those of ours. The MM-PBSA and MM-GBSA, which is not an all-atom simulation approach like ours, has drawbacks including the continuum solvent approximation. The intrinsic dielectric constant's appropriate setting presents another challenge. It has long been known that the selection of the intrinsic dielectric constant has a significant impact on the computed electrostatic energy<sup>22–24</sup>. However, 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 and MM-GBSA<sup>5,33</sup>. It is widely accepted that binding-free-energy estimates from MM-PBSA and MM-GBSA are less accurate than those from FEP, which is considered to be the gold standard for the calculation of absolute binding affinities<sup>67</sup>. We performed double annihilation FEP to calculate the absolute binding free energy of the hFGF1–heparin complex. The BFEE2 method was used to estimate binding affinities from FEP simulations with the consideration of several restraints to improve sampling within the framework of the method of refs. 3,4. The FEP simulations here were designed to have an aggregate simulation time comparable to that used in our BEUS simulations ( $\sim 2.3 \mu\text{s}$  for FEP compared with  $\sim 1.1 \mu\text{s}$  for SMD + BEUS). An absolute binding free energy of  $0.55 \pm 30.25 \text{ kcal mol}^{-1}$  was obtained for the FEP (Fig. 5a and Supplementary Table 2). Unlike the absolute binding free energy estimated from the BEUS simulations ( $-8.7 \pm 0.7 \text{ kcal mol}^{-1}$ ) (Table 1), the estimates from the FEP simulations are not in a good agreement with the binding free energy determined from ITC experiments ( $-7.88 \pm 0.01 \text{ kcal mol}^{-1}$ ) (Fig. 4). More importantly, a large uncertainty is associated with the FEP results that is due to the relative large size of the ligand. To show the effectiveness of our plan, we also calculated the binding free energies of hFGF1 and heparin directly using the method of refs. 3,4 as implemented within the BFEE2 package (the geometrical route). To make a fair comparison, we ran  $1.6 \mu\text{s}$  of aggregate simulation using the ABF free-energy calculations. Heparin and hFGF1 have a binding free energy of  $-19.04 \pm 2.95 \text{ kcal mol}^{-1}$  based on the geometric approach (Fig. 5b and Supplementary Table 3). In contrast to our technique, the BFEE2 geometrical route anticipated a value that was twice as high as the experimental

value of free energy. This comparison shows the effectiveness of our method over well-established binding-free-energy calculation methods. To firmly establish the efficiency of our strategy, additional research with a bigger data sample will be required in the future. In particular, it is important to determine what parameters make our method more efficient than the BFEE2 geometric protocol. For instance, it could be due to use of BEUS simulation scheme or a more fundamental difference regarding the use of simpler restraints and analysis schemes.

Studies have shown that the binding affinity and free-energy results derived from computational methods can be compared with experimental binding affinities obtained from ITC experiments<sup>7,8</sup>. 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. Our methodology could serve as a robust free-energy calculation method for determining the binding affinities of any protein and ligand of interest. However, the accuracy of the resulting binding free energies is still limited by the reliability of the force field parameters, which is at least equally as important as sampling for accurate physics-based binding affinity estimation.

## Discussion

The formalism presented in this work has notable similarities to the method previously proposed by ref. 3, and later implemented<sup>4,57</sup>. 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. The non-parametric reweighting allows for calculating the grid PMF in terms of the distance from the center of the binding pocket, as defined in this work, eliminating the need for calculating the PMF in terms of the polar and azimuthal angles as in the method of refs. 3,4. Relation (30) is a general scheme that can be easily adapted to any number of restraints. For instance, one may or may not add the polar and azimuthal angles to the restraints using the trivial generalization of relation (30). 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



**Fig. 5 | Binding-free-energy measurements using BFEE2 software. a**, Free-energy difference for the double annihilation of heparin hexasaccharide in its bound state and in water from FEP calculations.  $\lambda$  is a coupling parameter that gradually switches on interactions between a fully coupled state ( $\lambda = 0$ ) and a final, fully decoupled state ( $\lambda = 1$ ). **b**, PMF obtained during the separation of hFGF1 and heparin hexasaccharide using the geometric route. We used the BFEE2

(ref. 27) package to estimate the absolute free energy of binding in silico for an alchemical or geometrical route with multiple subprocesses and geometric constraints. The alchemical and geometric methods' binding-free-energy estimates were  $0.55 \pm 30.25$  kcal mol<sup>-1</sup> and  $-19.04 \pm 2.95$  kcal mol<sup>-1</sup>, respectively. See Supplementary Tables 2 and 3 for details of each geometric restriction and alchemical FEP calculations, respectively.

experimentally. Again, if restraining the orientation angle does not allow for a rapid convergence, one can add more restraints including the tilt and/or spin restraints. While the traces of the method of refs. 3,4 is clear in our derivation of the binding free energy, there are also clear differences in the use of the concept of the grid PMF that allows treating any restraints within the general formalism expressed in relation (30). A more extensive work is needed to determine when restraints in addition to those used in this work are necessary.

The outcomes of a simulation are significantly influenced by a variety of other factors, including model quality and the precision of docking. The efficiency of the method's findings may also be impacted by the degree to which the force field of ligands is accurately modeled. The restrictions that are often associated with these kinds of approach are connected to sampling, which might vary from project to project; for instance, bigger ligands may demand greater sampling than smaller ligands do. Also extremely crucial is the beginning structure of the bound state; the more precise the bound state, the more accurate the binding affinities will be.

## Methods

### Theoretical foundation

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

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

where [P], [L] and [P:L] are the concentrations of protein, ligand and the protein–ligand complex, respectively. Computationally, the absolute binding free energy ( $\Delta G^\circ$ ), 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

$$\Delta G^\circ = RT \ln \frac{K_d}{1M} \quad (2)$$

where  $R$  is the gas constant,  $T$  is the temperature and 1 M is 1 molar concentration. Various strategies have been used to estimate  $\Delta G^\circ$ , some of which were briefly discussed above. The methodology proposed here has a notable resemblance to the stratification strategy of refs. 3,4. However, the two methods have major differences as will be discussed later in this section.

Absolute binding free energy or  $\Delta G^\circ$  is the free-energy change associated with moving the ligand from the bulk to the binding pocket

(Supplementary Table 1). Within the formalism presented in this work,  $\Delta G^\circ$  is determined from the grid PMF  $G(\mathbf{x})$ , where  $\mathbf{x}$  is the position of the ligand mass center from the center of the binding pocket (Supplementary Table 1),  $G(\mathbf{x})$  is the PMF associated with the ligand position  $\mathbf{x}$ . In practice, we need to bin the three-dimensional space and define the PMF at every bin or grid point as:

$$G(\mathbf{x}) = -RT \ln p(\mathbf{x}) \quad (3)$$

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

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

$$\Delta G^\circ = -RT \ln \frac{\int_{\text{pocket}} e^{-\frac{G(\mathbf{x})}{RT}} dV}{\int_{\text{bulk}} e^{-\frac{G(\mathbf{x})}{RT}} dV} = -RT \ln \frac{\int_{\text{pocket}} e^{-\frac{\Delta G(\mathbf{x})}{RT}} dV}{\int_{\text{bulk}} e^{-\frac{\Delta G(\mathbf{x})}{RT}} dV} \quad (4)$$

in which the binding 'pocket' refers to all  $\mathbf{x} \in \mathbf{V}$  where the ligand is considered bound and 'bulk' refers to all  $\mathbf{x} \in \mathbf{V}$  where the ligand is not interacting with the protein.  $\mathbf{V}$  here is a subset of space with a single protein in standard concentration (that is, 1 M). As  $\Delta G(\mathbf{x})$  is the same everywhere in the bulk, we can simplify relation (4) as follows:

$$\Delta G^\circ = -RT \ln \frac{V_p}{e^{-\frac{\Delta G(\mathbf{x}_B)}{RT}} V_B} = -\Delta G(\mathbf{x}_B) - RT \ln \frac{V_p}{V_B} \quad (5)$$

where  $V_B$  is the bulk volume per protein associated with the standard concentration,  $\mathbf{x}_B$  is any grid point in the bulk and  $V_p$  is the binding pocket volume defined as:

$$V_p = \int_{\text{pocket}} e^{-\frac{\Delta G(\mathbf{x})}{RT}} dV \quad (6)$$

Defining  $\Delta G_v$  as the contribution of the difference between the volume of the binding pocket and the bulk to the binding free energy:

$$\Delta G_v = -RT \ln \frac{V_p}{V_B} \quad (7)$$

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

$$\Delta G^\circ = -\Delta G(\mathbf{x}_B) + \Delta G_v \quad (8)$$

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

$$V_B = \frac{1}{1M} \frac{\text{mol}}{N_A} = \frac{1}{N_A} \text{L} \approx 1,661 \text{ \AA}^3 \quad (9)$$

where  $N_A$  is Avogadro's constant and L is the unit of volume (litres). We can now rewrite  $\Delta G_V$  as:

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

in which  $\Delta G_B$  is the bulk volume contribution and  $\Delta G_P$  is the binding pocket contribution:

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

Determining both  $\Delta G(\mathbf{x}_B)$  and  $\Delta G_P$  requires finding the grid PMF  $\Delta G(\mathbf{x})$ .  $\Delta G(\mathbf{x}_B)$  is the PMF difference between the binding pocket center and the bulk and  $\Delta G_P$  also requires an estimate for  $\Delta G(\mathbf{x})$  within the binding pocket. We therefore do not need to find  $\Delta G(\mathbf{x})$  for all  $\mathbf{x}$  if we have a good estimate for  $\Delta G(\mathbf{x})$  within the binding pocket and in the bulk. Ideally,  $\Delta G(\mathbf{x})$  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  $d$ , that is, the distance between the mass centers of the ligand and protein.  $\Delta G(\mathbf{x})$  can be estimated for all sampled grid points  $\mathbf{x}$  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 three-dimensional 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 this section.  $\Delta G(\mathbf{x})$  can also be used to estimate  $\Delta G_P$  as defined in relation (11). There is often no need to strictly define the binding pocket as only low  $\Delta G(\mathbf{x})$  values have non-negligible contribution to  $V_P$  and thus even if we include all sampled grid points, only those close to the binding pocket center have non-negligible contributions.

A practical issue with determining  $\Delta G(\mathbf{x}_B)$  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<sup>3,4,24</sup>, we can circumvent extensive sampling of these degrees of freedom. Let us first focus on the orientation of the ligand ( $\Omega$ ), defined using the orientation quaternion formalism. We can restrain  $\Omega$  during the distance-based US simulations using a biasing potential ( $\frac{1}{2}k\Omega^2$  where a  $k$  is harmonic force constant) and later correct the free-energy difference based on the PMF associated with the  $\Omega$ , which is different in the bulk ( $F(\mathbf{x}_B, \Omega)$ ) and in the binding pocket ( $F(\mathbf{0}, \Omega)$ ). More generally, for any grid point  $\mathbf{x}$ , we may determine  $\Delta G(\mathbf{x})$  based on the PMF associated with the  $\Omega$  at  $\mathbf{x}$  ( $F(\mathbf{x}, \Omega)$ ) and  $\mathbf{0}$  ( $F(\mathbf{0}, \Omega)$ ):

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

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

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

where  $c$  is an arbitrary constant. We therefore have:

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

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

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

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

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

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

$$\Delta G(\mathbf{x}) = \Delta G_\Omega(\mathbf{x}) - \Delta U_\Omega(\mathbf{x}) \quad (17)$$

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

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

To determine the above ensemble averages, we need to determine the PMF along  $\Omega$  for the bound and unbound ligand and calculate the ensemble averages analytically using relation (16).  $\Delta G_\Omega(\mathbf{x}_B)$  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:

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

where we assume  $\Delta U_\Omega(\mathbf{x})$  is negligible for  $\mathbf{x}$  within the binding pocket.

In other words,  $\langle e^{-\frac{\frac{1}{2}k\Omega^2}{RT}} \rangle_{\mathbf{x}} \approx \langle e^{-\frac{\frac{1}{2}k\Omega^2}{RT}} \rangle_{\mathbf{0}}$  for  $\mathbf{x}$  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)):

$$\Delta G^\circ = -\Delta G_\Omega(\mathbf{x}_B) + \Delta U_\Omega(\mathbf{x}_B) + \Delta G_V \quad (20)$$

$F(\mathbf{x}_B, \Omega)$  can be calculated numerically from orientation angle distribution of a free ligand:  $F(\mathbf{x}_B, \Omega) = -RT \ln p(\Omega)$ , 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:

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

$\langle e^{-\frac{\frac{1}{2}k\Omega^2}{RT}} \rangle_{\text{bulk}}$  can then be calculated using relation (16) with numerically estimated  $F(\mathbf{x}_B, \Omega)$  and the  $k$  value used in the simulations.  $F(\mathbf{x}_B, \Omega) = -RT \ln p(\Omega)$  was numerically estimated by discretizing each of the 3 Euler angles with a bin width of  $1^\circ$  and a total of  $360 \times 360 \times 180$  bins to estimate  $p(\Omega)$  from  $p(\phi, \theta, \psi)$ .  $F(\mathbf{0}, \Omega)$  can be

determined approximately using orientation-based US simulations of bound ligand.  $F(\mathbf{0}, \Omega)$  can then be used to estimate  $\langle e^{-\frac{1}{2}k\Omega^2} \rangle_{\text{pocket}}$  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 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  $\Delta G(\mathbf{x})$  is calculated based on  $\Delta G_{\Omega,r}(\mathbf{x})$ , the grid PMF of a system whose  $\Omega$  and  $r$  are both restrained:

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

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

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

which results in:

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

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

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

where  $k'$  is the harmonic force constant associated with the  $r$  based on biasing potential ( $\frac{1}{2}k'r^2$ ). We also define  $U_r(\mathbf{x})$  similar to  $U_\Omega(\mathbf{x})$  in relation (15) except for using  $r$  instead of  $\Omega$ .  $U_\Omega^r(\mathbf{x})$  is also defined similar to  $U_\Omega(\mathbf{x})$  except for the additional restraint on  $r$ :

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

Finally, we have:

$$\Delta G(\mathbf{x}) = \Delta G_{\Omega,r}(\mathbf{x}) - \Delta U_r(\mathbf{x}) - \Delta U_\Omega^r(\mathbf{x}) + \Delta G_V \quad (27)$$

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

$$\Delta G^\circ = -\Delta G_{\Omega,r}(\mathbf{x}_B) + \Delta U_r(\mathbf{x}_B) + \Delta U_\Omega^r(\mathbf{x}_B) + \Delta G_V \quad (28)$$

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

$$V_p \approx \int_{\text{pocket}} e^{-\frac{\Delta G_{\Omega,r}(\mathbf{x})}{RT}} dV \quad (29)$$

Using relations (20) and (28), we can generalize the stratification strategy to include three restraints on arbitrary collective variables  $\alpha$ ,  $\beta$  and  $\gamma$ :

$$\Delta G^\circ = -\Delta G_{\alpha,\beta,\gamma}(\mathbf{x}_B) + \Delta U_\gamma(\mathbf{x}_B) + \Delta U_\beta^\gamma(\mathbf{x}_B) + \Delta U_\alpha^{\beta,\gamma}(\mathbf{x}_B) + \Delta G_V \quad (30)$$

where:

$$\Delta G_V \approx -RT \ln \int_{\text{pocket}} e^{-\frac{\Delta G_{\alpha,\beta,\gamma}(\mathbf{x})}{RT}} \frac{dV}{\tilde{A}^3} - \Delta G_B \quad (31)$$

### Isothermal titration calorimetry of hFGF1 with heparin hexasaccharide

ITC data were obtained using MicroCal iTC 200 (Malvern) with microcal origin software. 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 before loading. The protein-to-heparin ratio was maintained at 1:10 with the protein concentration being 100 μM and the heparin concentration being 1 mM. 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<sup>68</sup>. The standard binding free energy  $\Delta G^\circ$  was determined from dissociation constant via relation (2) at  $T = 25$  °C. 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  $\Delta G^\circ$ .

### All-atom MD simulations

For our bound state, we utilized the X-ray crystal structure of the dimeric hFGF1 combination with heparin hexasaccharide (PDB 2AXM; resolution, 3.0 Å)<sup>69</sup>, and for our apo state, we used the X-ray crystal structure of monomeric hFGF1 (PDB 1RGS; resolution, 1.1 Å)<sup>70</sup>. The NAMD 2.13 (ref. 71) was used to run MD simulations. Using a conjugate gradient, we energy-minimized the system for 10,000 steps. We next relaxed the systems using stepwise restrained MD simulations (for 1 ns) using CHARMM-GUI<sup>72</sup>. All production runs were done in an NPT (constant N, number of atoms; P, pressure; T, temperature) ensemble after the first NVT (constant N, number of atoms; V, volume; T, temperature) relaxation. Simulations were done at 300 K with a 2 fs time step and a 0.5 ps<sup>-1</sup> damping coefficient using a Langevin integrator. Nosé–Hoover–Langevin pistons were used to maintain 1 atm pressure<sup>72</sup>. Long-range electrostatic interactions were estimated using the particle mesh Ewald approach. The initial runs were done for 15 ns, followed by the production run on the Anton 2 supercomputer (Pittsburgh Supercomputing Center) for 4.8 μs with a 2.5 fs time step.

### MD simulations of free heparin hexasaccharide

Heparin hexasaccharide<sup>69</sup> was simulated in a rectangular water box without the protein. The system was set up as described previously in the 'All-atom MD simulations' section. 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.

### SMD simulations

The final conformations of the hFGF1–heparin<sup>73</sup>, apo hFGF1 (ref. 73) and free heparin hexasaccharide equilibrium simulations were used to generate starting conformations for the non-equilibrium pulling simulations. Four collective variables<sup>74</sup> were used for the SMD simulations<sup>75</sup>: (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 ( $\Omega$ ) defined using the orientation quanternion formalism; (3) RMSD of the protein ( $r_p$ ); (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 Å → 40 Å) with a force constant of 100 kcal (mol Å<sup>2</sup>)<sup>-1</sup>. The orientation angle was also restrained in this simulation with a force constant of 0.5 kcal (mol degree<sup>2</sup>)<sup>-1</sup> 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<sup>2</sup>)<sup>-1</sup>. Four RMSD-based SMD simulations were run for 10 ns each using a force constant of 50 kcal (mol Å<sup>2</sup>)<sup>-1</sup>: (1) to change the RMSD of the bound protein (0.5 Å → 2 Å) (the RMSD of heparin was restrained in this simulation with a force constant of 1 kcal (mol Å<sup>2</sup>)<sup>-1</sup>); (2) to change the RMSD of the bound heparin (1.5 Å → 4 Å); (3) to change the RMSD of the unbound protein (0.8 Å → 3.2 Å); (4) to change the RMSD of the free heparin (1.5 Å → 5.5 Å).

### BEUS simulations

BEUS<sup>53,76,77</sup>, which is a variation of the US simulation method, was performed to estimate grid PMF (Supplementary Fig. 1). Four independent sets of distance (*d*)-based BEUS simulations were performed, with no restraints, restraint on orientation angle of heparin with respect to the protein (*Ω*), restraint on RMSD of the ligand (*r<sub>l</sub>*) and RMSD of the protein (*r<sub>p</sub>*), and restraints on *Ω*, *r<sub>l</sub>* and *r<sub>p</sub>*. Two sets of BEUS simulations were also performed using the *Ω* collective variable, one with and one without a restraint on *r<sub>l</sub>* and *r<sub>p</sub>*. In addition, two sets of BEUS simulations were performed using the *r<sub>p</sub>* collective variable (bound protein with restraint on *r<sub>l</sub>*; unbound protein) and two sets were performed using the *r<sub>l</sub>* 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<sup>2</sup>)<sup>-1</sup> while the orientation was restrained as in SMD simulations using a force constant of 0.5 kcal (mol degree<sup>2</sup>)<sup>-1</sup>. 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<sup>2</sup>)<sup>-1</sup>. The force constant used for *r<sub>l</sub>* and *r<sub>p</sub>* in all cases was 1 kcal (mol Å<sup>2</sup>)<sup>-1</sup>. See Supplementary Fig. 1 for a schematic representation of these simulations.

### Free-energy calculations using non-parametric reweighting

Once the BEUS simulations described above were converged, a non-parametric reweighting method<sup>76,78</sup>, which is somewhat similar to the multi-state Bennett acceptance ratio method<sup>79</sup>, was used to construct the PMF. In this method<sup>76</sup>, 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<sub>i</sub>(r)*, where *i* = 1, ..., *N*, and **r** represents all atomic coordinates. Typically, *U<sub>i</sub>(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 {**r<sub>k</sub>**} (irrespective of which bias was used to generate each sample **r<sub>k</sub>**) and determine the weight of each sample (*w<sub>k</sub>*) as:

$$w_k = c / \sum_i e^{-\frac{U_i(\mathbf{r}_k) - F_i}{RT}} \quad (32)$$

where, *c* is the normalization constant such that  $\sum_k w_k = 1$  and both {*w<sub>k</sub>*} and perturbed free energies {*F<sub>i</sub>*} are determined iteratively using the above equation and the following:

$$e^{-\beta F_i} = \sum_k w_k e^{-\frac{U_i(\mathbf{r}_k)}{RT}} \quad (33)$$

Converged *w<sub>k</sub>* values can be used to construct any ensemble averages including any PMF (for example, *G(ζ)* PMF of the atomic system in the collective variable space(*ζ*)) in terms of not only 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:

$$G(\zeta_i) = -RT \ln \sum_k w_k \delta(\zeta(\mathbf{r}_k) - \zeta_i), \quad (34)$$

$$\delta(\zeta(\mathbf{r}_k) - \zeta_i) = \begin{cases} 1, & |\zeta(\mathbf{r}_k) - \zeta_i| < |\zeta(\mathbf{r}_k) - \zeta_j| \text{ for } j \neq i. \\ 0, & \text{otherwise} \end{cases} \quad (35)$$

To estimate the uncertainty of any of PMF calculations described above, one may use bootstrapping. Here, we have used a block Bayesian bootstrapping technique<sup>77</sup>, where 100 alternative datasets are resampled from the existing dataset 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 FEP simulations

We used the BFEE2 (ref. 27) package to estimate the absolute free energy of binding in silico for an alchemical or geometrical route with multiple subprocesses and geometric constraints. 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<sup>80</sup>, wherein the heparin hexasaccharide is annihilated in both the free and bound states. The final conformations of the hFGF1–heparin complex<sup>73</sup> and free heparin hexasaccharide equilibrium simulations (discussed previously in the ‘All-atom MD simulations’ section) were used to generate starting conformations for the bound hFGF1–heparin and free heparin FEP simulations respectively. For the alchemical route, four separate simulations are performed: (1) coupling the restraints of seven collective variables in the bound state; (2) decoupling the ligand alchemically in the bound state; (3) coupling the ligand alchemically in the unbound state; (4) decoupling the conformational restraints in the unbound state. The FEP simulations 1 and 3 were performed bidirectionally using 200 *λ*-windows (*λ* is the coupling parameter associated with the FEP that could vary between 0 and 1). Each *λ*-window included a 0.5 ns of equilibration and 5.0 ns of averaging for both the unbound and bound states, for a total of 2.3 μs (Supplementary Table 3). The decoupling FEP simulations 2 and 4 were also performed bidirectionally, each one for 51 ns. All FEP simulations were performed using the NAMD 2.13 (ref. 71) simulation package with the CHARMM36m all-atom additive force field, using the protocol discussed previously for the equilibrium simulations. We used the state-of-the-art BFEE2 (ref. 27) method to make input files and analyze the FEP simulations.

### Binding-free-energy calculations using geometrical route

The extended ABF technique with an umbrella integration estimator was used to calculate the free-energy change along the coarse variables required to characterize reversible heparin–hFGF1 binding<sup>3,24,27</sup>. We used the software BFEE2 (ref. 27) to generate the input files for these simulations. In the geometrical route, these collective variables are often subjected to restrictions, and the amount of reversible work required to impose each constraint is determined by a sequence of very accurate PMF simulations. The collective variables used here are the RMSDs of the two proteins’ backbone distances from the reference, native conformation, the three Euler angles (*θ*, *φ* and *ψ*) that describe

their relative orientation and the polar ( $\theta$ ) and azimuth angles ( $\varphi$ ) that describe their relative position<sup>27,81</sup>. The geometrical path consists of a sequence of separate PMF computations performed sequentially with the gradual inclusion of restrictions (RMSD,  $\theta$ ,  $\Phi$ ,  $\Psi$ ,  $\theta$  and  $\varphi$ ), as shown in Supplementary Table 2. Each geometric collective variable (RMSD,  $\theta$ ,  $\Phi$ ,  $\Psi$ ,  $\theta$ ,  $\varphi$  and  $r = (1/\beta) \ln(S^*P^*C^\circ)$ ;  $\beta = (k_B T)^{-1}$ , with  $k_B$  the Boltzmann constant and  $T$  the temperature;  $C^\circ$  denotes the standard concentration of 1 M.  $P^*$ , which stands for the separation term, and  $S^*$ , which stands for the surface term, indicate the percentage of a sphere with radius  $r^*$ , centered at the binding site of the reference protein, that is, accessible to its partner) simulation was run with 10 replicas per restriction, and each replica simulation included 20 ns (RMSD,  $\theta$ ,  $\Phi$ ,  $\Psi$ ,  $\theta$  and  $\varphi$ ) of simulation time ( $r$  collective variables simulations were run for 40 ns for each replica), for a total of 1.6  $\mu$ s. The BFEE2 (ref. 27) Gui was used to analyze the final ABF simulation data.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Datasets related to this article are deposited to the Zenodo repository<sup>82</sup>. Source data for Figs. 1–5 is available with this paper. Protein Data Bank (<https://www.rcsb.org/>) was used to collect the crystal structures 2AXM ref. 69 and 1RG8 ref. 70.

### Code availability

All scripts as well as the full source code for non-parametric reweighting can be obtained from Zenodo<sup>82</sup>.

### References

- Mobley, D. L. & Gilson, M. K. Predicting binding free energies: frontiers and benchmarks. *Annu. Rev. Biophys.* **46**, 531–558 (2017).
- Wan, S., Bhati, A. P., Zasada, S. J. & Coveney, P. V. Rapid, accurate, precise and reproducible ligand–protein binding free energy prediction: binding free energy prediction. *Interface Focus* **10**, 20200007 (2020).
- Woo, H. J. & Roux, B. Calculation of absolute protein–ligand binding free energy from computer simulations. *Proc. Natl Acad. Sci. USA* **102**, 6825–6830 (2005).
- Gumbart, J. C., Roux, B. & Chipot, C. Standard binding free energies from computer simulations: what is the best strategy? *J. Chem. Theory Comput.* **9**, 794–802 (2013).
- Siebenmorgen, T. & Zacharias, M. Computational prediction of protein–protein binding affinities. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **10**, e1448 (2020).
- Du, X. et al. Insights into protein–ligand interactions: mechanisms, models, and methods. *Int. J. Mol. Sci.* **17**, 144 (2016).
- 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.* **10**, 4069–4078 (2014).
- 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.* **125**, 10570–10579 (2003).
- Weiss, S. Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy. *Nat. Struct. Biol.* **7**, 724–729 (2000).
- Rossi, A. M. & Taylor, C. W. Analysis of protein–ligand interactions by fluorescence polarization. *Nat. Protoc.* **6**, 365–387 (2011).
- Huang, D. & Caflisch, A. Efficient evaluation of binding free energy using continuum electrostatics solvation. *J. Med. Chem.* **47**, 5791–5797 (2004).
- Rodinger, T., Howell, P. L. & Poms, R. Absolute free energy calculations by thermodynamic integration in four spatial dimensions. *J. Chem. Phys.* **123**, 34104 (2005).
- Ytreberg, F. M. & Zuckerman, D. M. Simple estimation of absolute free energies for biomolecules. *J. Chem. Phys.* **124**, 104105 (2006).
- Rodinger, T., Howell, P. L. & Poms, R. Calculation of absolute protein–ligand binding free energy using distributed replica sampling. *J. Chem. Phys.* **129**, 155102 (2008).
- 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.* **5**, 909–918 (2009).
- 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.* **6**, 2559–2565 (2010).
- General, I. J., Dragomirova, R. & Meirovitch, H. Absolute free energy of binding of avidin/biotin, revisited. *J. Phys. Chem. B* **116**, 6628–6636 (2012).
- Fu, H. et al. Accurate determination of protein:ligand standard binding free energies from molecular dynamics simulations. *Nat. Protoc.* **17**, 1114–1141 (2022).
- 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.* **48**, 2325–2335 (2005).
- 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.* **38**, 1887–1890 (2017).
- Lensink, M. F. & Wodak, S. J. Docking and scoring protein interactions: CAPRI 2009. *Proteins Struct. Funct. Bioinf.* **78**, 3073–3084 (2010).
- 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.* **120**, 9401–9409 (1998).
- Wang, C. et al. Calculating protein–ligand binding affinities with MMPBSA: method and error analysis. *J. Comput. Chem.* **37**, 2436–2446 (2016).
- Fu, H. et al. BFEE: a user-friendly graphical interface facilitating absolute binding free-energy calculations. *J. Chem. Inf. Model.* **58**, 556–560 (2018).
- Chipot, C. Frontiers in free-energy calculations of biological systems. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **4**, 71–89 (2014).
- Chodera, J. D. & Mobley, D. L. Entropy–enthalpy compensation: role and ramifications in biomolecular ligand recognition and design. *Annu. Rev. Biophys.* **42**, 121–142 (2013).
- Fu, H., Chen, H., Cai, W., Shao, X. & Chipot, C. BFEE2: automated, streamlined, and accurate absolute binding free-energy calculations. *J. Chem. Inf. Model.* **61**, 2116–2123 (2021).
- Ali, H. S., Chakravorty, A., Kalayan, J., de Visser, S. P. & Henchman, R. H. Energy–entropy method using multiscale cell correlation to calculate binding free energies in the SAMPL8 host–guest challenge. *J. Comput. Aided Mol. Des.* **35**, 911–921 (2021).
- Kollman, P. Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev.* **93**, 2395–2417 (1993).
- 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.* **72**, 1047–1069 (1997).
- 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.* **119**, 2707–2714 (1997).

32. Tuckerman, M. E. Free Energy Calculations: Theory and Applications in Chemistry and Biology Springer Series in Chemical Physics, 86 Edited by Christophe Chipot (Université Henri Poincaré Vandoeuvre-lès-Nancy, France) and Andrew Pohorille (University of California, San Francisco, USA). *J. Am. Chem. Soc.* **129**, 10963–10964 (2007).
33. Fratev, F. & Sirimulla, S. An improved free energy perturbation FEP+ sampling protocol for flexible ligand-binding domains. *Sci. Rep.* **9**, 16829 (2019).
34. Jorgensen, W. L. Free-energy calculations: a breakthrough for modeling organic chemistry in solutions. *Acc. Chem. Res.* **22**, 184–189 (1989).
35. Boresch, S., Tettinger, F., Leitgeb, M. & Karplus, M. Absolute binding free energies: a quantitative approach for their calculation. *J. Phys. Chem. B* **107**, 9535–9551 (2003).
36. Hermans, J. & Shankar, S. The free energy of xenon binding to myoglobin from molecular dynamics simulation. *Isr. J. Chem.* **27**, 225–227 (1986).
37. 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.* **71**, 670–681 (1996).
38. Fujitani, H. et al. Direct calculation of the binding free energies of FKBP ligands. *J. Chem. Phys.* **123**, 84108 (2005).
39. 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* **105**, 9795–9799 (2001).
40. Deng, Y. & Roux, B. Calculation of standard binding free energies: aromatic molecules in the T4 lysozyme L99A mutant. *J. Chem. Theory Comput.* **2**, 1255–1273 (2006).
41. Deng, Y. & Roux, B. Computations of standard binding free energies with molecular dynamics simulations. *J. Phys. Chem. B* **113**, 2234–2246 (2009).
42. Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **9**, 646–652 (2002).
43. 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.* **41**, 429–452 (2012).
44. Imadisetty, 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* **121**, 3644–3656 (2017).
45. Gunsteren, W. F. & Mark, A. E. Validation of molecular dynamics simulation. *J. Chem. Phys.* **108**, 6109–6116 (1989).
46. Gunsteren, W. F., Dolenc, J. & Mark, A. E. Molecular simulation as an aid to experimentalists. *Curr. Opin. Struct. Biol.* **18**, 149–153 (2008).
47. Zuckerman, D. M. & Chong, L. T. Weighted ensemble simulation: review of methodology, applications, and software. *Annu. Rev. Biophys.* **46**, 43–57 (2017).
48. Zwier, M. C. et al. WESTPA: an interoperable, highly scalable software package for weighted ensemble simulation and analysis. *J. Chem. Theory Comput.* **11**, 800–809 (2015).
49. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011–1021 (1992).
50. Souaille, M. & Roux, B. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput. Phys. Commun.* **135**, 40–57 (2001).
51. Luitz, M., Bomblies, R., Ostermeier, K. & Zacharias, M. Exploring biomolecular dynamics and interactions using advanced sampling methods. *J. Phys. Condens. Matter* **27**, 323101 (2015).
52. Kokubo, H., Tanaka, T. & Okamoto, Y. Ab Initio prediction of protein-ligand binding structures by replica-exchange umbrella sampling simulations. *J. Comput. Chem.* **32**, 2810–2821 (2011).
53. Moradi, M. & Tajkhorshid, E. Mechanistic picture for conformational transition of a membrane transporter at atomic resolution. *Proc. Natl Acad. Sci. USA* **110**, 18916–18921 (2013).
54. Kästner, J. & Thiel, W. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: ‘umbrella integration’. *J. Chem. Phys.* **123**, 144104 (2005).
55. Barducci, A., Bussi, G. & Parrinello, M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **100**, 20603 (2008).
56. Comer, J. et al. The adaptive biasing force method: everything you always wanted to know but were afraid to ask. *J. Phys. Chem. B* **119**, 1129–1151 (2015).
57. Gumbart, J. C., Roux, B. & Chipot, C. Efficient determination of protein-protein standard binding free energies from first principles. *J. Chem. Theory Comput.* **9**, 3789–3798 (2013).
58. Eswarakumar, V. P., Lax, I. & Schlessinger, J. Cellular signaling by fibroblast growth factor receptors. *Cytokine Growth Factor Rev.* **16**, 139–149 (2005).
59. Beenken, A. & Mohammadi, M. The FGF family: biology, pathophysiology and therapy. *Nat. Rev. Drug Discov.* <https://doi.org/10.1038/nrd2792> (2009).
60. Kuro-o, M. Endocrine FGFs and Klothos: emerging concepts. *Trends Endocrinol. Metab.* **19**, 239–245 (2008).
61. Ornitz, D. M. et al. FGF binding and FGF receptor activation by synthetic heparan-derived di- and trisaccharides. *Science* **268**, 432–436 (1995).
62. 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* **39**, 7153–7158 (2000).
63. Babik, S., Samsonov, S. A. & Pisabarro, M. T. Computational drill down on FGF1–heparin interactions through methodological evaluation. *Glycoconj. J.* **34**, 427–440 (2017).
64. 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.* **25**, 221–233 (2015).
65. Goetz, R. & Mohammadi, M. Exploring mechanisms of FGF signalling through the lens of structural biology. *Nat. Rev. Mol. Cell Biol.* <https://doi.org/10.1038/nrm3528> (2013).
66. Bojarski, K. K., Sieradzian, A. K. & Samsonov, S. A. Molecular dynamics insights into protein-glycosaminoglycan systems from microsecond-scale simulations. *Biopolymers* **110**, 23252 (2019).
67. Genheden, S. & Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **10**, 449–461 (2015).
68. 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.* **434**, 233–241 (2013).
69. Digabriele, A. D. et al. Structure of a heparin-linked biologically active dimer of fibroblast growth factor. *Nature* **393**, 812–817 (1998).
70. Bernett, M. J., Somasundaram, T. & Blaber, M. An atomic resolution structure for human fibroblast growth factor 1. *Proteins Struct. Funct. Genet.* **57**, 626–634 (2004).
71. Lee, J. 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.* **12**, 405–413 (2016).
72. 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* **125**, 12690–12697 (2021).

73. Fiorin, G., Klein, M. L. & Hémin, J. Using collective variables to drive molecular dynamics simulations. *Mol. Phys.* **111**, 3345–3362 (2013).
74. Izrailev, S., Stepaniants, S., Balsera, M., Oono, Y. & Schulten, K. Molecular dynamics study of unbinding of the avidin–biotin complex. *Biophys. J.* **72**, 1568–1581 (1997).
75. Moradi, M. & Tajkhorshid, E. Computational recipe for efficient description of large-scale conformational changes in biomolecular systems. *J. Chem. Theory Comput.* **10**, 2866–2880 (2014).
76. Moradi, M., Enkavi, G. & Tajkhorshid, E. Atomic-level characterization of transport cycle thermo-dynamics in the glycerol-3-phosphate:phosphate transporter. *Nat. Commun.* **6**, 8393 (2015).
77. Bartels, C. Analyzing biased Monte Carlo and molecular dynamics simulations. *Chem. Phys. Lett.* **331**, 446–454 (2000).
78. Shirts, M. R. & Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **129**, 124105 (2008).
79. 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.* **72**, 1047–1069 (1997).
80. Phillips, J. C. et al. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
81. Coderc De Lacam, E. G., Blazhynska, M., Chen, H., Gumbart, J. C. & Chipot, C. When the dust has settled: calculation of binding affinities from first principles for SARS-CoV-2 variants with quantitative accuracy. *J. Chem. Theory Comput.* **18**, 5890–5900 (2022).
82. Kumar, V. G., Polasa, A., Agrawal, S., Kumar, T. K. S. & Moradi, M. Binding affinity estimation from restrained umbrella sampling simulations. *Zenodo* <https://doi.org/10.5281/zenodo.7348705> (2022).

## Acknowledgements

This research is supported by National Science Foundation grants CHE 1945465 and OAC 1940188 (to M.M.) and the Arkansas Biosciences Institute (to M.M.). This work is also supported by the Department of Energy grant DE-FG02-01ER15161 (to M.M. and T.K.S.K.), the National Institutes of Health grants R15GM139140 (to M.M.), R01CA172631 (to T.K.S.K.), P30GM103450 (to T.K.S.K.) and P20GM139768 (to T.K.S.K.), and the Arkansas Integrative Metabolic Research Center at the University of Arkansas (to T.K.S.K.). 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 to M.M.),

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.

## Author contributions

M.M. designed the research. V.G.K. and A.P. 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., A.P., M.M., T.K.S.K. and S.A. wrote the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s43588-022-00389-9>.

**Correspondence and requests for materials** should be addressed to Mahmoud Moradi.

**Peer review information** *Nature Computational Science* thanks Jeffry Setiadi, Richard Henchman and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Kaitlin McCardle, in collaboration with the *Nature Computational Science* team.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- |     |           |
|-----|-----------|
| n/a | Confirmed |
|-----|-----------|
- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
  - A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
  - The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
  - A description of all covariates tested
  - A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
  - A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
  - For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
  - For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
  - For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
  - Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

- |                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|-----------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Data collection | The ITC thermograms were analysed using Microcal origin software (Version # 7E). BFEE2 (version BFEE 2.3.1) was used to to generate input files for Alchemical free energy perturbation (FEP) simulations and Binding free-energy calculations using geometrical route simulations. NAMD 2.13 (version 2.13) was used to collect the molecular dynamics simulation data. Protein data bank ( <a href="https://www.rcsb.org/">https://www.rcsb.org/</a> ) was used to collect the crystal structures 2AXM and 1RG8. |
| Data analysis   | The ITC thermograms were analysed using Microcal origin software (Version # 7E). BFEE2 (version BFEE 2.3.1) was used to analyze Alchemical free energy perturbation (FEP) simulations and Binding free-energy calculations using geometrical route simulations data.                                                                                                                                                                                                                                               |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Our simulations were based on the x-ray crystal structure of the dimeric complex with a heparin hexasaccharide (PDB:2AXM, resolution: 3.0 Å) and the model for the apo protein was based on the x-ray crystal structure of unbound monomeric hFGF1 (PDB: 1RG8, resolution: 1.1 Å). The list of figures that have the associated raw experimental data is Figure 4. There is no restriction on data availability from the ITC experiment.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	ITC experiments were performed in triplicate under the same experimental conditions with the same sample. The variance in the ITC data, obtained under the same experimental conditions, is negligible. The triplicate simulations simply show that the measured binding affinities are highly reproducible under the same experimental conditions and estimated uncertainties are negligible. Therefore, the ITC experiments, conducted in triplicate is sufficient and n>3 would not provide additional information. For computational data associated with the new algorithm, the number of replicas are chosen based on the complexity of the problem to reach convergence, which is shown in supplementary figures 2-5.
Data exclusions	No datapoint in the ITC results were omitted. All datapoints in the isothermogram were used in analysis.
Replication	The isothermograms, obtained under the same experimental conditions, representing the heparin hexasaccharide - FGF1 titration were quite reproducible both qualitatively and quantitatively. ITC experiments were independently performed in triplicates and all the replications were successful.
Randomization	No experimental groups were used in this study. Randomization therefore was not applicable.
Blinding	The ITC experiment is a well-established experimental technique. Therefore, it was deemed that the blinding exercise was not necessary. No other experiment was performed to require blinding.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging