Do all roads lead to Rome? Convergence issues in umbrella sampling simulations

Noora Aho,[*,†,‡] Gerrit Groenhof,[*,†] and Pavel Buslaev[*,†]

†*Nanoscience Center and Department of Chemistry, University of Jyväskylä, Finland*

‡*Theoretical Physics and Center for Biophysics, Saarland University, Saarbrücken, Germany*

E-mail: noora.aho@uni-saarland.de; gerrit.x.groenhof@jyu.fi; pavel.i.buslaev@jyu.fi

Abstract

Molecular dynamics (MD) simulations are widely applied to estimate absolute binding free energies of protein-ligand and protein-protein complexes. A routinely used method for binding free energy calculations with MD is umbrella sampling (US), which calculates the potential of mean force (PMF) along a reaction coordinate. In this work, we investigate the convergence of US along standard distance-based reaction coordinates for various protein-protein and protein-ligand complexes, following commonly used guidelines for the setup. We show that repeating the complete US workflow can lead to differences of 2-20 kcal/mol in computed binding free energies. We attribute those discrepancies to small differences in the binding pathways. We then demonstrate that adaptive-biasing approaches, which are constructed to sample multiple pathways in a single simulation, such as the accelerated weight histogram (AWH) method, can achieve convergence between independent simulations. To the best of our knowledge, this is the first attempt to systematically assess the shortcomings of the widely accepted protocols for US of protein-protein and protein-ligand binding affinities. We anticipate

therefore that our results will provide an incentive for a critical reassessment of the validity of PMFs computed with US, as well as adopt adaptive-biasing approaches for computing binding affinities.

Introduction

Molecular dynamics (MD) simulations have proven their usefulness in studying many kinds of biological processes, including intracellular dynamics,[1] protein folding,[2] and capturing the ligand-binding pathway.[3] Among the goals that MD can target is the accurate calculation of association/dissociation free energies of, for example, protein-protein, protein-ligand, protein-DNA, and ligand-membrane systems. For that, one would want to use MD to estimate absolute binding free energies (ABFE). While relative binding free energy (RBFE) calculations have already matured and are now almost routinely used in industrial workflows,[4–9] ABFE calculations are in general more complicated, require more resources, and are still under development.[10] The performance of methods for calculating ABFE is regularly assessed in the SAMPL challenge,[11,12] and currently, a typical error in ABFE calculations is twice as high as in RBFE (mean absolute error of $\sim$2 kcal/mol for the best ABFE method[12] against $<$1 kcal/mol for RBFE methods[5,7,13]).

In the SAMPL challenge, as for ABFE calculations in general, one of the most frequently used methods is umbrella sampling (US).[14] The idea of US is to create a set of reference configurations along a predefined reaction coordinate, for example by pulling a molecule from another molecule, and to run MD for these independent "umbrellas" while restraining the sampling to the region around the reference configurations along the chosen reaction coordinate. These simulations are independent and can hence be performed in parallel. After the US simulations have sufficiently converged, the umbrellas are unbiased, for example with the weighted histogram analysis method (WHAM),[15,16] to obtain the potential of mean force (PMF) from which the binding affinity is inferred. In what follows, we refer to the whole procedure, which includes the creation of the reference configurations along the reaction coordinate, the MD simulations of the individual umbrellas, and their transformation into a PMF, as a single 'repeat'. The popularity of the US is largely because of its availability in the popular MD software packages, tutorials, and automatic tools for setting up and postprocessing the simulations.

Owing to its wide applicability, the usage of the US is accompanied by a lot of methodological development. To provide some examples of the current methodological developments on US simulations for calculating the binding free energy, Woo and Roux[17] demonstrated that applying position restraints on the system during the US restricts the sampled configurational space, which needs to be corrected to obtain binding free energy. The approach was further developed by Doudou *et al.*[18] using linear restraints and Chipot *et al.*[19] using a set of geometrical restraints, the best combination of which was further scrutinized.[20,21] It has been shown that without using such restraints the resulting PMFs do not correspond well to the experimental values.[22]

As the reference configurations for US along a one-dimensional reaction coordinate are often obtained from one-dimensional pulling simulations, in which usually one part of the complex is pulled into the solvent, as generally advised,[23,24] the US procedure has been criticized for its unidirectional nature.[25,26] To overcome this issue, other methods such as metadynamics,[27] adaptive biasing force,[28] and adaptive weight histogram[25] have been developed. The advantage of these methods is that they iteratively alter the underlying potential energy surface and enhance the sampling along the selected reaction coordinate in all directions, meaning multiple binding/unbinding events occur during a single simulation.

Recent methodological developments have also focused on selecting optimal collective variables for running metadynamics or other adaptive algorithms. This has for instance been achieved by reweighting techniques[29] or machine learning approaches.[30] As a recent example, it has been demonstrated that for a widely studied test system, benzamidine in trypsin, the water networks can be explicitly taken into account as a part of collective variables, which leads to a much better agreement between simulations and experiments.[31]

Other methodological attempts focus on scaling the intermolecular interactions for different "umbrellas", thus reducing the potentially high barriers of the intermediate states while allowing to obtain the thermodynamic binding free energy.[32] The aforementioned list of methodological developments in ABFE calculations is far from complete, and we refer the interested readers to recent reviews, in which the advances in the field are covered.[33–36] Although various methods aiming at accurate calculations of ABFE have been introduced during the past decade, most of them are not yet implemented in popular software packages or have only been tested for toy models. Therefore, the US remains one of the most popular methods for computing ABFEs.

To illustrate how established the single-coordinate US is, we estimated which fraction of MD simulation research is actually using it. To compute the PMF from the individual umbrella sampling windows, the most commonly used approach is the weighted-histogram analysis method (WHAM).[15,16] Two popular implementations of WHAM used nowaways are the `g_wham` GROMACS tool by Hub *et al.*[37,38] and the WHAM by A. Grossfield.[39] To emphasize the popularity of PMF calculation by the US method along one-dimension coordinate, based on `https://scholar.google.com/` data, the aforementioned publications[15,16,37,39] related to the WHAM method are cited 700-800 times per year during the last five years. Compared to the number of total citations for the most popular MD softwares (AMBER,[40–42] CHARMM,[43,44] GROMACS,[38,45–47] NAMD[48–50]), being around 6000-11000 per year during the last five years, publications citing one of the PMF methods would correspond to approximately 7-11% of all MD papers (Figure S1). Some recent examples of the ABFE calculation using the standard US approach include estimation of the free energy related to a cancer antigen binding to a protein,[51] binding of DNA aptamers,[52,53] small molecules binding to membranes,[54,55] ligands binding to SARS-CoV-2,[56] and insulin dimer dissociation.[57] With so many resources dedicated to computing binding free energies with US for problems with such high medical relevance, it is imperative that this approach yields accurate predictions.

While US has been recommended as "one of the most accurate techniques for free-energy calculations [...] only limited by its elevated computational cost",[58] systematic validations are currently lacking. Moreover, the few studies aimed at assessing this validity for specific systems, suggest a dependency of the calculated free energies on simulation setup.[33,59] Despite this controversy, US along a one-dimensional distance-based coordinate has remained a widely accepted method of choice for computing binding affinities, presumably because of its simplicity. To understand if such optimism is justified, and in particular, to assess if the common US workflows are generally valid, we have systematically assessed the accuracy of the US sampling methods for protein-protein and protein-ligand systems.

Our results demonstrate that small deviations in the reference configurations selected for the umbrella windows for the exact same reaction coordinate can significantly affect the computed free energies, leading to 2-20 kcal/mol differences in the obtained PMF profiles. We thus conclude that performing a single US simulation along a single path is not sufficient for computing binding free energies and multiple pathways must be sampled instead.

Sampling multiple pathways can be conveniently achieved with adaptive-biasing approaches.[25,60–64] These methods rely on sampling multiple pathways simultaneously, and thus can be expected to provide better convergence and smaller errors. In these methods, the state of the system is coupled to a system parameter, which is a function of the atomic coordinates. This parameter evolves simultaneously with the system following the distribution of the assigned probability weight factor - bias. The bias is directly related to the free energy landscape of the parameter. During the adaptive biasing simulations, the distribution of assigned weights is constantly updated based on the simulation history in order to achieve the target distribution, usually uniform, of the parameter. Indeed, using the adaptive-biasing approach, namely, the accelerated weight histogram (AWH) method[25] by Hess and co-workers, led to converged PMFs. We additionally demonstrate that using a shared bias between different simulations is essential.

Thus, our results suggest that sampling multiple pathways is essential when computing binding free energies. In addition, we share important technical aspects of using adaptive biasing procedures for computing binding free energies. We anticipate that our results will help the computational community to select the most suitable method for performing and analyzing PMF computations. For further methodological development of computational techniques for evaluating ABFE, more reliable interaction functions,[65] and methods are required. This work is focused on the reliability of popular approaches for computing ABFE, rather than on reproducing experimental estimates, which, we hope, will help the community to further develop ABFE methodology.

Methods

We performed multiple repeats of non-equilibrium pulling followed by umbrella sampling to obtain PMF profiles for the dissociation of three protein-protein, one protein-peptide, and one protein-ligand complexes. To show that the results are independent of MD software, we performed simulations with GROMACS 2021.5 and 2022.4 versions[38] and OpenMM.[66] To demonstrate that the results do not depend on the force field used, we ran simulations with CHARMM36[67] and Amber ff99SB-ILDN force fields.[68] To sample multiple pathways within one simulation, we performed multiple simulations with the accelerated weight histogram (AWH) method[25] with GROMACS and the CHARMM36 force fied. Here, we provide the details of the simulation systems, simulation parameters, and procedures for pulling, US, and analysis.

Simulated systems

We calculated the PMF profiles of the dissociation for the following systems (Figure 1): (1) barnase-barstar (PDB ID: 1BRS[69]) (2) HdeA dimer (PDB ID: 1BG8[70]) (3) clpS protease adaptor with LLL tripeptide (PDB ID: 3G19[71]) (4) trypsin-benzamidine (PDB ID: 3PTB[72]) and (5) amyloid β-peptide $A\beta_{42}$ (PDB ID: 2BEG[73]). For all systems, we performed multiple repeats of US simulations. The AWH simulations were performed for systems 1 and

4. For systems 1-4, the interactions were modeled using the CHARMM36 all-atom force field (FF).[67] Benzamidine in system 4 was parameterized with CGenFF.[74,75] For system 5, the GROMOS96 53A6 parameter set[76] was used, as in the earlier work of Lemkul *et al.*[23] Additionally, to rule out bias of the force field, we repeated the simulations of system 3 using the Amber ff99SB-ILDN FF.[68] We also ran system 3 with CHARMM36 force field but with OpenMM software[66] to check that discrepancies between repeats are not software-specific. Systems 1-4 were constructed using the following procedure: (i) The structure for each protein system 1-4 was obtained from the Protein Data Bank[77] and placed in a periodic rectangular box; (ii) The C- and N-termini were kept charged, and the protonation states for the titratable residues were selected for pH=7: all Glu, Asp and His deprotonated, and Lys protonated; (iii) The proteins were oriented in the simulation box such that the reaction coordinate of the pulling simulation was aligned with the z-coordinate, as in the earlier work,[18,23] and the reaction coordinate was defined either as the vector connecting the center-of-masses (COMs) of the two proteins or protein and peptide (systems 1-3), or as the vector connecting the COM of heavy atoms of the ligand and the C_γ atom of Asp-189 (system 4) as was proposed in previous work;[18] (iv) Systems were then solvated with CHARMM TIP3P[78,79] or TIP3P[80] water molecules. Systems were neutralized by adding Na+ and Cl– ions at 0.15 M concentration. For system 5, the initial structure was an equilibrated snapshot from the tutorial of Lemkul,[81] based on their earlier publication.[23]

The sizes of the simulation boxes of our systems were: barnase barstar 9x9x15 nm^3 box with ∼38000 waters, HdeA dimer 8x8x15 nm^3 with ∼31000 waters, clpS protease adaptor 6x6x12 nm^3 with ∼13500 waters, and trypsin–benzamidine 9x9x12 nm^3 with ∼31000 waters. System 5, the amyloid β-peptide comprised a box of dimensions 11x10x12 nm^3 with ∼42000 SPC[82] waters. All input configurations, together with topology and run parameters, are provided as Supporting Information.

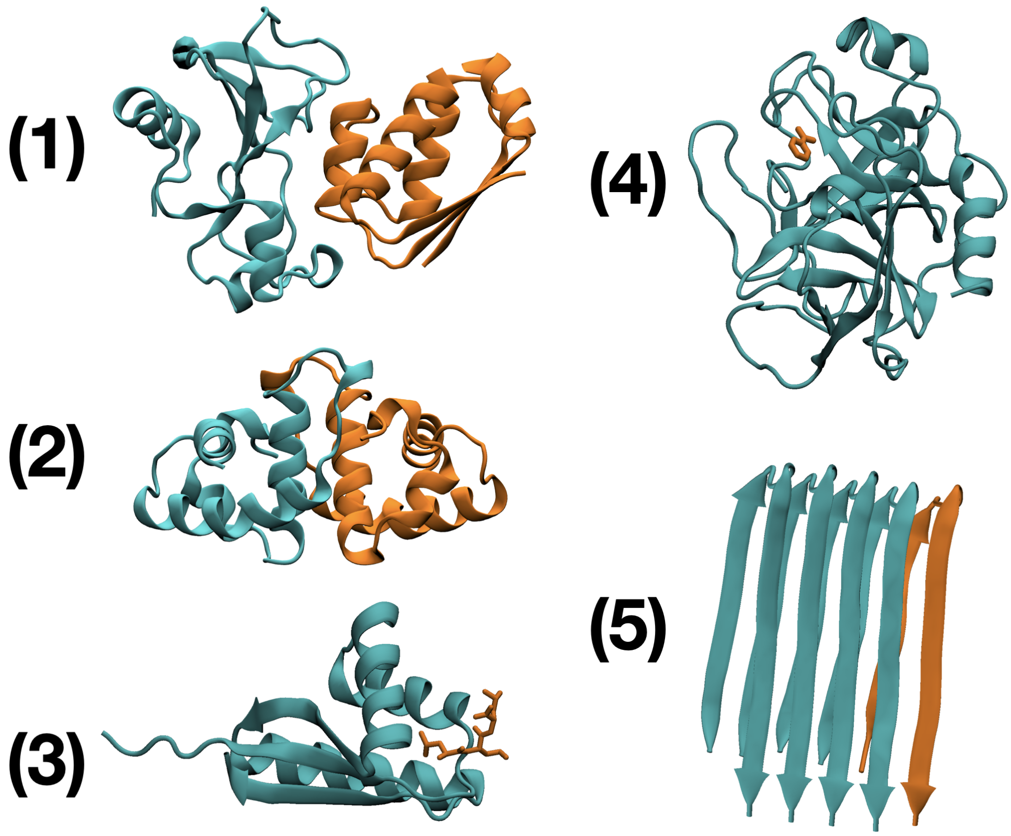

Figure 1: Structures of the simulated systems, for which the PMF of the dissociation was calculated using umbrella sampling. (1) barnase-barstar 1BRS[69] (2) HdeA dimer 1BG8[70] (3) clpS protease adaptor with LLL tripeptide 3G19[71] (4) trypsin–benzamidine 3PTB[72] and (5) amyloid β-peptide Aβ_{42} 2BEG.[73] The protein, peptide, or ligand dissociated from the protein is shown in orange.

Simulation details

To check the convergence of independent US repeats, we performed simulations of the five systems described in the previous section. The AWH simulations were performed for systems 1 and 4. In the following, we describe the GROMACS[38] and OpenMM[66] parameters used.

GROMACS

Molecular dynamics Since most of the simulations in this work were performed with the CHARMM36 force field, we only specify the parameters for those simulations. The input parameter files for simulations performed with the Amber ff99SB-ILDB and GROMOS96

53A6 force fields can be found in the supplementary archive with all the inputs for simulated systems.

For the simulations with the CHARMM36 force field, Coulomb interactions were computed using the smooth particle mesh Ewald (sPME)[83,84] method with a real-space cut-off of 1.2 nm and a grid spacing of 0.14 nm. Van der Waals interactions were modeled with the Lennard-Jones potential, which was smoothly switched to zero in the range from 1.0 to 1.2 nm. Constant temperature of 300 K was maintained with the v-rescale thermostat[85] using a time constant of 0.5 ps^{-1}. Constant pressure of 1 bar was maintained with the Parrinello-Rahman barostat[86] using a relaxation time of 2.0 ps. For AWH simulations of system 1, constant pressure was maintained with c-rescale barostat[87] with a relaxation time of 5.0 ps. The leapfrog integrator with a timestep of 2 fs was used, together with the LINCS[88] algorithm to constrain lengths of bonds to hydrogen atoms of the solute molecules, and the SETTLE[89] algorithm to constrain internal degrees of freedom of the water molecules. The aforementioned parameters were the same for pulling, umbrella sampling, and AWH simulations. For the additional simulations with the Amber ff99SB-ILDN and for system 5 with the GROMOS96 53A6 force fields, the parameters can be found in the input files provided in the SI.

Prior to umbrella sampling, the potential energy of systems 1-4 was minimized using the steepest descent algorithm, followed by equilibration of 50 ps in NVT and 50 ps in NPT ensembles. During the NVT and NPT equilibration, position restraints of 1000 kJ $mol^{-1}nm^{-2}$ were applied in all directions on the C_α atoms of the proteins and peptide, and on all heavy atoms of the ligand in system 4. During pulling, one of the protein chains was restrained and the other chain (can be protein, peptide, or lipid) was pulled away from it. We refer to the restrained protein chain as a fixed one, and to the one which was pulled away as the mobile one. During pulling and US simulations of systems 1-4 the following restraints were applied: 1000 kJ/mol/nm^2 position restraints in all directions for the C_α atoms of the fixed protein chain, and orthogonal position restraints in x-,y-directions with a force constant of

1000 kJ/mol/nm^2 for the C_α atoms of the mobile protein or peptide, or heavy atoms of the ligand. For system 5, following the procedure of Lemkul and Bevan,[23] position restraints of 1000 kJ/mol/nm^2 in all directions were only applied to the C_α atoms of the subunit next to the one dissociated from the aggregate. To test the effect of position restraints in x,y-directions, we also performed additional simulations for the trypsin-benzamidine (system 4) using different force constants $k_{xy} = 4184$ kJ/mol/nm^2 and $k_{xy} = 0$ kJ/mol/nm^2

Pulling and umbrella sampling. As the aim was to perform multiple repeats for umbrella sampling in order to compare the PMF profiles, we first performed multiple repeats of the non-equilibrium pulling to get initial structures for umbrella windows along the dissociation coordinate. Starting from the equilibrated system, the mobile protein, peptide, or ligand was pulled away from the fixed protein along the z-axis. Multiple repeats (5 repeats for systems 1-4 and 10 repeats for system 5) of pulling for each system were performed, all starting from the same initial structure but with different initial velocities. For system 1, barstar was pulled away while barnase was kept fixed, for system 2 the two monomers of the HdeA dimer are identical, so one of the monomers was kept fixed and the other one was pulled. For systems 3 and 4, the peptide or ligand was pulled away while the protein was kept fixed. For system 5, the subunit A was pulled away from the rest of the protein. In Figure 1 the dissociated parts are shown in orange.

The pulling was carried out as follows: for systems 1-3 slowly over 100 ns, using a pull rate of 0.000033 nm/ps and a spring constant of 10000 kJ/mol/nm^2. For system 4, the original work[18] did not provide details on the generation of initial windows for the US, so we performed the pulling over 1 ns, using a pull rate of 0.005 nm/ps and a spring constant of 1000 kJ/mol/nm^2. For system 5, we used the values as in the original work,[23] namely a much faster pulling over 500 ps with a pull rate of 0.01 nm/ps and a spring constant of 1000 kJ/mol/nm^2.

The starting configurations for the US windows were obtained as snapshots from the individual pulling trajectories. For systems 1 and 2, 26 umbrella windows were used with a 0.05 nm spacing along the reaction coordinate for the first 1 nm of COM separation of the proteins and a 0.1 nm spacing for the next 0.5 nm. For system 3, 21 umbrella windows with a 0.05 nm spacing were used for 1 nm COM separation between the peptide and protein. For system 4, 25 umbrella windows with a 0.1 nm spacing were used, where the separation was measured as the distance from the COM of heavy atoms of the ligand and the C_γ atom of Asp-189. Lastly for system 5, 31 umbrella windows with a 0.1 nm spacing for the first 1.5 nm separation and a 0.2 nm spacing for the next 2.0 nm separation, as suggested in the original research of the system.[23]

For the individual umbrella sampling simulations, a harmonic restraining potential was used. For systems 1 and 2, a higher force constant of 10000 kJ/mol/nm^2 was used for the denser windows of the first 1 nm, and a force constant of 1000 kJ/mol/nm^2 for the remaining. For system 3, a force constant of 10000 kJ/mol/nm^2 was used for all US windows. For systems 4 and 5, again following the original works,[18,23] we used force constants of 4184 kJ/mol/nm^2 = 10 kcal/mol/Å^2 and 1000 kJ/mol/nm^2, respectively. The lengths of umbrella sampling simulations for systems 1, 2, 4, and 5 were 10 ns per US window in each repeat, resulting in a total of 200-300 ns per system. For system 4, simulating 10 ns per window was ten times longer than in the original work,[18] and for system 5, the same as in the original work.[23] For system 3, each umbrella window was simulated for 50 ns.

To estimate the PMF along the reaction coordinate from the US simulations, the weighted histogram analysis method (WHAM) implemented in GROMACS as `gmx wham`[37] was used. For the error analysis of the PMF profiles from US simulations, the conventional bootstrapping method with 100 bootstraps was used.

Accelerated weight histogram method In addition to multiple repeats of the US simulations, we also performed multiple repeats of accelerated weight histogram method (AWH) simulations for systems 1 and 4. For all the AWH simulations, a constant target distribution was used. The estimated initial error was set to 40 kJ/mol and the input diffusion constant was set to 2.0×10^{-5} nm^2/ps. The number of steps between the sampling of the coordinate value, as well as the number of coordinate samples used for each AWH update, were set to 10. For system 1 the interval for the z-distance was set to 2.3-3.2 nm and for system 4 to 0.53-2.3 nm. For the AWH with a single bias potential, the simulations were started from the bound state and were run for 1600 ns (system 1) and 800 ns (system 4) per repeat. In addition, the AWH simulations were repeated using four bias-sharing walkers. The starting configurations for the walkers were obtained at regular intervals along the distance from the pulling simulations, and the same starting configurations were used for all repeats. Each walker was sampled for 400 ns (system 1) and 200 ns (system 4). For the AWH simulations, the same reaction coordinate along z and position restraints in x-, and y-directions were used as in the US runs.

OpenMM

To check that the discrepancies between independent umbrella sampling repeats are not software specific, we also performed simulations with OpenMM[66] for clpS protease adaptor. The input files and scripts used to run OpenMM simulations are available at the gitlab page.

Pulling and umbrella sampling The input structure used for pulling simulations with OpenMM was taken after NPT equilibration run with GROMACS. All the simulations were performed in the NVT ensemble at 300 K using Langevin integrator with a friction coefficient of 1 ps^{-1} and step size of 0.002 ps. Coulomb interactions were computed using the PME[83,84] method. Van der Waals interactions were modeled with the Lennard-Jones potential with a cutoff of 1.2 nm. Constraints were used for bonds to hydrogens.

The pulling was run for over 100 ns, using a pull rate of 0.000033 nm/ps and a spring constant of 1000 kJ/mol/nm^2. Frames were saved every picosecond. For umbrella sampling simulations the distance between protein chains was restrained with the force constant of

$10000\ \mathrm{kJ/mol/nm^2}$. In all the simulations C_α atoms of the static protein chain were restrained to their initial positions with the force constant of 1000 $\mathrm{kJ/mol/nm^2}$. C_α atoms of the moving protein chain were restrained to their initial positions with the force constant of 1000 $\mathrm{kJ/mol/nm^2}$ in $x-$ and $y-$ directions, as proposed by Doudou et al.[18] To estimate the PMF along the reaction coordinate from the US simulations, the WHAM implementation from Grossfield lab version 2.0.11[39] was used.

Results and Discussion

Here, we present the results of multiple repeats of umbrella sampling for the dissociation of protein-protein, protein-peptide, and protein-ligand complexes. We show that the obtained PMF profiles for each system differ by about 2-20 kcal/mol between individual repeats, indicating that the common procedure of relying on one repeat of US leads to unreliable and untrustable results. A similar issue was recently discussed for MD in general by Peter Coveney and co-authors,[90] but here we focus specifically on US simulations. We demonstrate that very small differences in reference structures for the umbrellas can lead to large differences in the PMF profiles. Adaptive-biasing methods, like AWH,[25] rely on the sampling of multiple pathways in one simulation, and to understand if the inclusion of multiple pathways resolves the convergence issues, we also assess the performance of such simulation techniques for protein-protein and protein-ligand systems.

Multiple independent repeats of the US simulations

In Figure 2, we show the computed PMF profiles for the dissociation of the five systems described in the Methods section and illustrated in Figure 1 using the US procedure. For all simulated systems, the PMF profiles differ for individual repeats. Differences in the PMF depth (ΔW_R) between repeats lie in the range of $\sim$2-20 kcal/mol (Table 1), which is much higher than the PMF errors calculated with the bootstrapping method usually

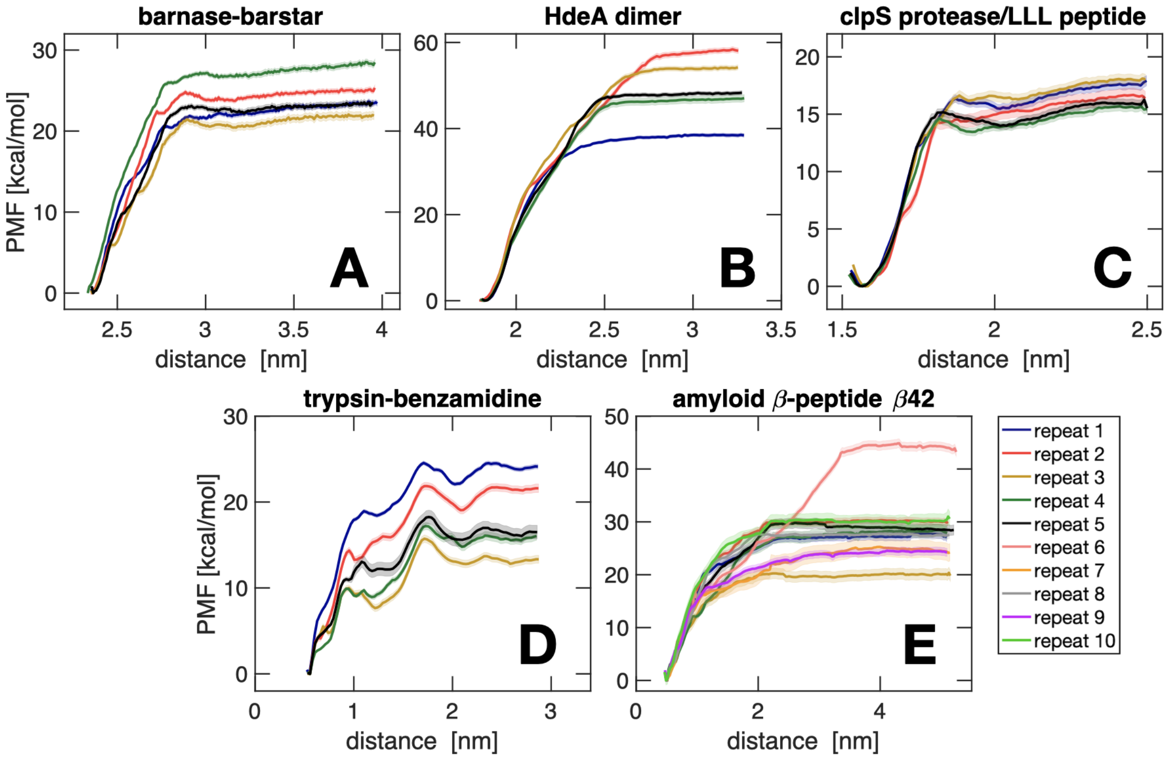

Figure 2: PMF profiles for multiple repeats of umbrella sampling for the simulated systems, where different colors correspond to different repeats. The different repeats of the US were generated by performing multiple pulling simulations starting from the same equilibrated structure, from which the initial snapshots for US windows were taken. The error estimation from bootstrapping is shown with a lighter color.

employed for the WHAM,[37] which is in the order of only 0.5-2 kcal/mol. Large differences in the ΔW_R between repeats can be problematic because the binding free energy between different ligands to a binding site may vary by about the same amount,[33] making it difficult to differentiate between ligands with US. To rule out potential effects of the force field or software, we additionally calculated the dissociation PMFs for the clpS protease adaptor complexed with LLL tripeptide (system 3) using Amber99sb-ildn FF[68] with GROMACS[38] and CHARMM36[67] with OpenMM[66] software. The differences between the US repeats were also observed for those simulations (Figure S2).

Repeats, leading to different computed values, often indicate a lack of convergence of the umbrellas. However, for all simulated systems there is a sufficient overlap between adjacent

Table 1: Estimated PMF depths ΔW_R values for the repeats of the US (systems 1-5) and the multi-walker AWH (systems 1 and 4) simulations. ΔW_R was calculated from the PMF profiles presented in Figure 2 as the mean of the flat region of the PMF (cut-offs of flat region 3.0, 2.75, 1.7, 1.9 and 3.5 nm for US systems 1-5, respectively, and 3.0 and 1.2 for AWH systems 1 and 4, respectively.)

system	ΔW_R from US [kcal/mol]	ΔW_R from AWH [kcal/mol]
(1) barnase-barstar	-22.5/-24.4/-21.1/-27.3/-22.8	13.8/13.3/13.4
(2) HdeA dimer	-38.3/-57.4/-53.9/-46.6/-48.0	–
(3) clpS protease adaptor	-15.7/-14.5/-15.9/-14.0/-14.5	–
(4) trypsin-benzamidine	-23.7/-21.0/-13.2/-15.7/-16.4	21.3/21.8/24.0/23.1/22.4
(5) amyloid β-peptide	-27.2/-30.2/-20.0/-28.1/-28.9	–
	-44.3/-25.0/-27.6/-24.3/-30.0	–

umbrella histograms (Figures S3-S7), which is a key requirement for computing PMFs with WHAM. Additionally, we find that the umbrellas in each repeat are converged in time, as indicated by the convergence of PMF profiles with simulation length (0-2.5, 0-5.0, 0-7.5, 0-10.0 ns for systems 1,2,4,5 and 0-10, 0-20, 0-30, 0-40, 0-50 ns for system 3) (Figures S8-S12). These observations thus suggest that the differences between the PMF profiles of individual repeats are not due to a lack of convergence in the US simulations of the individual repeats.

Instead, following You *et al.*,[59] we attribute the discrepancies between the PMF profiles of the individual repeats to differences in the pathways along the reaction coordinate. In the US method, those pathways are defined by the reference configurations for the harmonic restraints in each US window. For the HdeA dimer (system 2) and the amyloid β-peptide A β_{42} (system 5) differences in pathways were observed for different repeats, as shown in Figures S18 and S19 of SI. However, for all other systems, the root mean square deviations (RMSDs) between reference structures for the umbrella windows of different repeats were around 0.1 nm, indicating that the pathways were very similar (Figure S20). Moreover, exchanging the reference structures between repeats while keeping the initial coordinates the same, reveals a strong dependency of the PMF on those reference structures (Figure 3). Since the sequence of reference structures defines the pathway for the US, these results suggest that the PMF is determined to a large extent by the pathway.

Because the HdeA dimer is a partially disordered protein, for which dimer formation and monomer folding occur simultaneously,[91,92] pulling can be expected to yield multiple pathways along the same reaction coordinate. In general, a one-dimensional reaction coordinate is not a suitable choice for systems where binding is accompanied by large conformational rearrangements. Some success for such systems has been achieved by using the total number of contacts between residues that define the state of the system, as the reaction coordinate,[92,93] but the analysis of such methods lies beyond the scope of this work.

As for the amyloid β-peptide A β_{42} system, the initial configurations for the umbrella windows were obtained with shorter pulling simulations (0.5 ns) than for all other systems (1-100 ns), and both pulling and US simulations were performed without any orthogonal restraints. This is because our main goal for this system was to reproduce and assess the US protocol published by Lemkul and co-workers more than 10 years ago.[23] This protocol is also used in widely spread GROMACS tutorials,[81] which is the starting point for many users in the area of MD simulations and US. With much faster pulling, the projection of initial structures onto the reaction coordinate is not uniformly distributed (see Figure S17), compared to systems that were pulled more slowly (Figures S13-S16). When pulled fast, the system can sample only a small portion of configuration space, which leads to larger differences between pathways, as compared to slower pulling.(Figure S20, S21). However, differences between pathways remain.

These differences can be further reduced by applying orthogonal restraints, which reduces the RMSD between reference structures for umbrella windows to 0.5 nm for fast pulling and to 0.1 nm for slow pulling (Figure S21). Slow pulling with orthogonal restraints leads to higher similarity between pathways, therefore avoiding one of the potential reasons for differences in PMF profiles. The importance of applying orthogonal restraints was recently discussed by Blazhynska and co-workers,[22] who compared PMF profiles obtained from simulations with geometrical restraints (which according to Doudou and co-workers are similar to orthogonal restraints[18]) to those obtained from free simulations. They observed differences of several kcal/mol between individual repeats of simulations with no restraints, and none of those repeats converged to experimental values. The authors, however, did not report performing multiple repeats of restrained simulations, though the value reported for a single restrained simulation was in good agreement with experimental values.

We observe that even when pathways are structurally similar (*i.e.*, RMSD below 0.1 nm), differences between PMF profiles remain, as seen for example for system 4 (Figure 2D). To take into account the effect of orthogonal restraints, Doudou and co-workers proposed to use two correction terms ΔG_V and ΔG_R, which reflect the change in the free energy when restraints are removed for the unbound and bound states, respectively.[18] To check if the differences between individual repeats of the US simulations can be mitigated with such corrections, we calculated the correction terms for trypsin-benzamidine as in Doudou's paper and for HdeA dimer, as an example system with clearly different pathways. For both systems, the computed correction terms were almost identical for all repeats (Table S1), suggesting that the differences between the PMFs are not due to the orthogonal restraints.

Another possible reason for non-converged ABFE calculations was discussed by Ansari and co-workers, who demonstrated that water networks can play a key role in the binding process of a ligand.[31] In the case of the trypsin-benzamidine system, a water molecule in the binding pocket initiates the ligand release process.[31] In the unbound state, there is a water network replacing the ligand in the binding pocket.[94] However, while also in our simulations, water networks form during the pulling simulations, the structures of these networks differ significantly between the repeats of the US simulations (Figure S22). In addition, within a single repeat the network changes in simulations of the single umbrella windows (Figure S23). The average number of water molecules around residues Y228 and D189 fluctuates

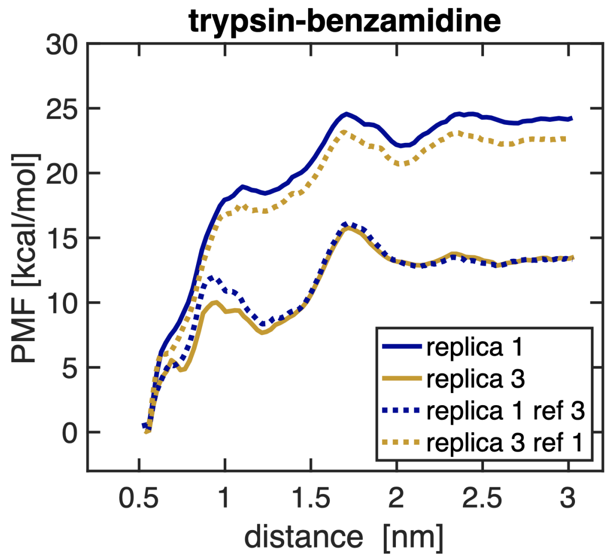

Figure 3: Effect of initial structures used for the US windows on the PMF profiles. Two additional repeats of the US simulations were performed for trypsin-benzamidine (system 4). In those repeats, we took reference configurations, which are used as the origin of position restraints for umbrella windows, from repeats 1 and 3 and the initial configurations from repeats 3 and 1, respectively. The initial configuration did not affect the computed PMF profiles.

between the US repeats (Figure S24), pointing to a lack of structural convergence in the binding pocket, which, according to Ansari *et al.*, can influence the resulting PMF profiles.[31]

While the wetting/dewetting transition of the binding pocket often plays a key role in ligand binding,[95,96] it is not the only determinant defining the protein-ligand and protein-protein dissociation.[97] Our results, unfortunately, demonstrate that for large biomolecular systems like protein-protein or protein-ligand complexes, a straightforward application of the umbrella sampling protocol for a single reaction coordinate, as widely used by the community, may not always provide reliable binding free energies by itself. Similar results were also obtained for solute permeation across membrane channels, for which solute pulling in two directions led to major hysteresis effects.[98,99]

As a conclusion, despite starting the pulling from the same equilibrated structure and applying orthogonal restraints, each pulling simulation results in a slightly different pathway along the reaction coordinate. Such differences in the pathways, even if small, can result in differences of at least a few, or even tens of, kcal/mol in the resulting PMF. Therefore, a single set of US using a simple distance-based reaction coordinate is not sufficient for obtaining a reliable estimate for protein-protein, protein-peptide, or protein-ligand affinities.

Sampling multiple pathways

While performing a single repeat of US along a one-dimensional coordinate has emerged as the established approach for computing PMFs, our results suggest that this may lead to an incorrect estimation of binding affinity as there can be other pathways that lead to different PMFs. To estimate this effect, multiple repeats are needed. However, when the differences between PMFs from multiple repeats are large, obtaining meaningful results is challenging. While averaging PMF profiles is one possibility, as introduced by Niskikawa *et al.*,[100] this would assume equal weights for all paths sampled in the repeats. Because of the differences between the PMFs computed for multiple repeats of US along the same 1D reaction coordinate, we assume that not all non-equilibrium work has dissipated in our simulations. Therefore, computing the average with equal weights will also overestimate the free energy difference since the total work exceeds the free energy:[101]

$$\overline{W} \geq \Delta G$$

with $\overline{W}$ the total work performed, and ΔG the free energy difference between the states.

As opposed to equal weights, Jarzynsky demonstrated that free energy can be estimated from nonequilibrium-pulling simulations by computing the exponentially averaged work values over the repeats.[101,102] Non-equilibrium techniques do sample multiple pathways and resemble the multiple-US-repeats discussed in this work. However, the US in principle samples an equilibrium distribution, whereas the repeats in the Jarzynsky approach are always out-of-equilibrium. Therefore, using exponential averaging may not be directly applicable to the different PMFs of multiple US-repeats. The third possibility is to compute the Boltzmann-weighted average of the computed PMF profiles. This weighting approach, however, requires a sufficient overlap of phase spaces between independent US repeats, which most probably is not the case due to the large difference in the computed free energy differences.

While none of the aforementioned averaging procedures is ideal, we nevertheless applied them to estimate free energies from our simulations. The difference between the approaches is 4.7 kcal/mol, with averaging over the results from the repeats with equal weights providing the highest estimate (18.2 kcal/mol), while the lowest estimate (13.4 kcal/mol) was obtained by averaging with the Boltzmann weights. The exponential averaging yields a value of 14.3 kcal/mol (Figure 4). The results of different averaging approaches for the other systems can be found in Supplementary Table S2, and the corresponding Figures S25-28.

Additionally, we repeated 1000 rounds of WHAM using umbrella windows along the reaction coordinate selected randomly from the independent repeats of the US runs. This "mixing" of the umbrella windows, or cross-WHAM, results in new pathways, which are combined from pieces of the original pathways. Figure 4 shows that cross-WHAM can lead to lower and higher ΔW_R values than for original repeats. We used the cross-WHAM replicas to estimate the error of each averaging approach. We randomly selected 5 replicas out of 1000 generated and computed averages for those 5 replicas. We repeated this procedure 10000 times and estimated the errors of various averaging approaches as the standard deviation of 10000 computed averages. The errors were 1.3, 1.8, and 1.7 kcal/mol for equal weight, exponential, and Boltzmann averaging respectively. These errors were at least 0.9 kcal/mol higher than WHAM bootstrapping error (0.3-0.8 kcal/mol) for the trypsin-benzamidine system. Similar results for the errors were obtained for other systems (Table S2).

In principle, free energy differences associated with different pathways should be averaged with some weights, and those weights should depend on the overlap between the conformational ensembles sampled in the US simulation along the particular pathway. While such reweighing might be possible, it will require the development of a new formalism and software. The simple reweighing techniques discussed here, led to differences in the estimated

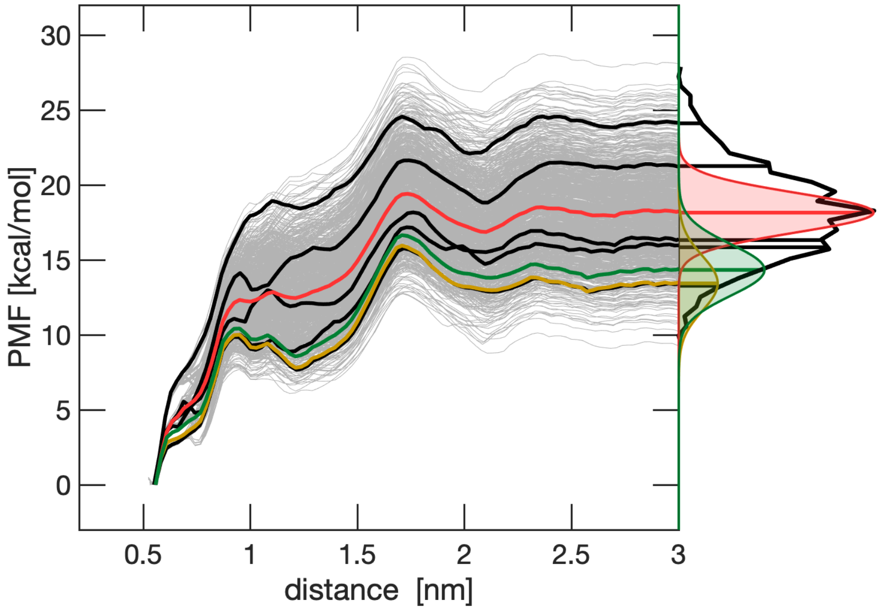

Figure 4: Gray lines show the 1000 PMFs from cross-WHAM combining randomly selected umbrella windows from different US repeats of the trypsin-benzamidine system. Black lines show the PMF profiles for the original sets of US, from Figure 2. Red, green, and yellow lines show the averages of these five PMFs, obtained using equal, exponential, and Boltzmann averaging, respectively. On the right-hand side of the figure, the black line shows the distribution of ΔW_R, where each ΔW_R is calculated as the average of the flat part of the PMF, here for distances >2.0 nm. The red, green, and yellow lines show Gaussian distributions with a σ equal to the standard deviation of the distribution of the averages, computed for 10000 randomly selected sets of 5 cross-WHAM repeats.

free energies of 4.7 kcal/mol, which is below the 12 kcal/mol maximal difference between repeats. However, the differences between the various reweighing techniques are higher than the error for the individual reweighting (≈ 1.5 kcal/mol). Without a physical argument in favor of a specific reweighting approach, it remains unclear how to extract meaningful free energy estimates from multiple US repeats.

In contrast, adaptive-biasing methods, which are discussed below, sample multiple pathways on the fly according to Boltzmann's distribution while providing sufficient overlap of phase space between neighboring states. We therefore, applied such an approach, namely the accelerated weight histogram (AWH) method by Lindahl *et al.*,[25] to compute the free energy for protein-protein and protein-ligand binding.

Adaptive-biasing methods

In adaptive-biasing methods, a system parameter, which is a function of the atomic coordinates, is introduced. Typically, the evolution of this parameter is associated with a transition between different states of the system, for example, a distance between chains for binding, or a reaction coordinate for a chemical transformation. Because the barriers between the states can be high, sampling all relevant states can be infeasible. Sampling can be improved by adding a biasing potential to the system. In adaptive methods, the bias is applied to the system parameter and dynamically updated until a predefined distribution of the parameter is obtained. Because the parameter is coupled to the coordinates, the biasing potential forces the system to sample the relevant states and the underlying free energy profile can be directly obtained from the bias.[25,60–64] In contrast to US simulations, in which sampling is restrained along fixed points on a predefined path that connects the end states, no such path is defined in adaptive-biasing methods. Instead, by applying the bias, a single trajectory samples the full range of the system parameter along multiple pathways. To further speed up sampling, multiple trajectories can be coupled to the same system parameter and run in parallel (multiple-walker simulations).[103] Using the AWH method[25] as a representative of a multi-walker adaptive-biasing approach, we computed the dissociation-free energies of barnase-barnstar and benzamidine-trypsin systems.

Figure 5 shows the PMF profiles computed from AWH simulations for the barnase-barstar complex (system 1), and the trypsin-benzamidine complex (system 4). While with a single walker (Figure 5A, C) the differences between repeats are smaller than for the US, but still in the order of 5 kcal/mol, this difference reduces to 1-3 kcal/mol when multiple walkers are applied (Figure 5B, D). The convergence of the PMF profiles as a function of time for AWH demonstrates a similar trend (Figures S33-S36). For single-walker simulations, 400 ns is not sufficient to achieve convergence for both trypsin-benzamidine and barnase-

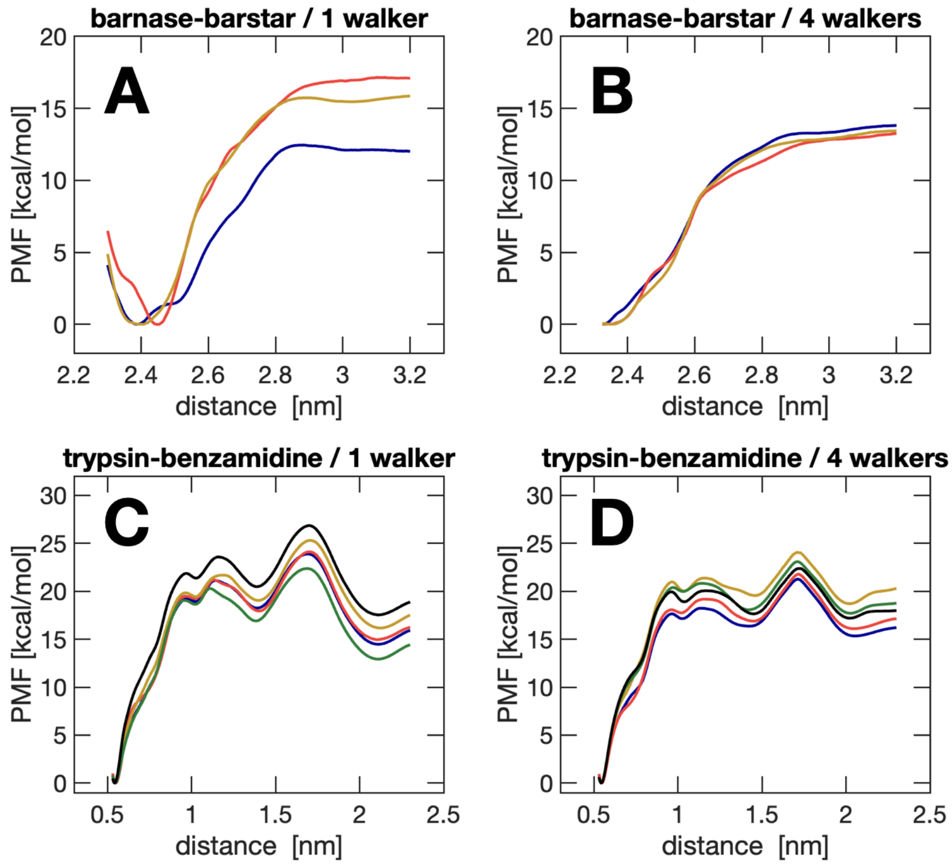

Figure 5: PMF profiles for repeats of AWH simulations. (A) barnase-barstar (system 1) with 1 walker, 1600 ns per replica (B) barnase-barstar with four walkers, 400 ns per replica (C) trypsin-benzamidine (system 4) with 1 walker, 800 ns per replica, and (D) trypsin-benzamidine with four walkers, 200 ns per replica.

barstar complexes, as the profiles continue to evolve (Figures S34, S36). In contrast, with four walkers the PMF profiles for both trypsin-benzamidine and barnase-barstar complexes, converge within 100 ns (Figures S33, S35), demonstrating the higher efficiency of multiple-walker simulations compared to single-walker ones. Similar trends were demonstrated for pulling solutes across membrane channels.[98,99]

An additional advantage of the multiple walkers is that the bound state is sampled more extensively as compared to a single walker simulation (Figures S30 and S32). In the single walker trajectory of the barnase-barnstar complex (Figure S30), the bound state is undersampled and the system does not return to the bound state during the simulation, which affects the PMF profile (Figure S34). In contrast, we observe that in the multiple walker AHW simulations, at least one of the four walkers samples the bound state extensively (Figures S29 and S31). Thus, based on previous and our results, we claim that adaptive-biasing approaches are more robust and reliable than US simulations.

Conclusions

We have presented a systematic assessment of the computation of absolute binding free energies for protein-protein and protein-ligand systems by means of the umbrella sampling technique, currently one of the most popular techniques for calculating the ABFE. Our analysis demonstrates that independent repeats of US simulations can lead to different PMF profiles. We attribute these discrepancies to the differences between the pathways. We also demonstrate that adaptive biasing approaches, by sampling multiple pathways in one simulation, can lead to more consistent PMF profiles. We anticipate that our results will motivate the computational community to (i) shift towards more reliable methods (e.g. adaptive-biasing methods) for ABFE computation; and (ii) validate the approach chosen for the calculation by scrutinizing its accuracy, convergence, and reproducibility.

Acknowledgement

The authors thank Dmitry Morozov, Emmi Pohjolainen, Justin Lemkul, Berk Hess, and Rebecca Wade for fruitful discussions. This work was supported by the Academy of Finland (Grants 311031, 342908, 332743). The simulations were performed on resources provided by the CSC-IT Center for Science, Finland.

Supporting Information Available

The Supporting Information is available free of charge at ACS Publications website and contains the following information:

- Input files and parameters for presented MD (https://zenodo.org/records/10220083)
- Additional analysis (Do_all_roads_lead_to_Rome_SI.pdf)
  - Details on the citations for PMF and MD methods
  - US with OpenMM / Amber FF
  - Umbrella histograms, convergence of individual US repeats
  - Analysis of pulling simulations
  - Contributions to the free energy of binding from PMF
  - Trypsin-benzamidine pathways
  - Cross-WHAM analysis
  - Analysis of AWH simulations

References

(1) Singharoy, A.; Maffeo, C.; Delgado-Magnero, K. H.; Swainsbury, D. J.; Sener, M.; Kleinekathöfer, U.; Vant, J. W.; Nguyen, J.; Hitchcock, A.; Isralewitz, B., et al. Atoms to phenotypes: molecular design principles of cellular energy metabolism. *Cell* **2019**, *179*, 1098–1111.

(2) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How fast-folding proteins fold. *Science* **2011**, *334*, 517–520.

(3) Dror, R. O.; Pan, A. C.; Arlow, D. H.; Borhani, D. W.; Maragakis, P.; Shan, Y.; Xu, H.; Shaw, D. E. Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proceedings of the National Academy of Sciences* **2011**, *108*, 13118–13123.

(4) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J., et al. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *Journal of the American Chemical Society* **2015**, *137*, 2695–2703.

(5) Abel, R.; Wang, L.; Harder, E. D.; Berne, B.; Friesner, R. A. Advancing drug discovery through enhanced free energy calculations. *Accounts of chemical research* **2017**, *50*, 1625–1632.

(6) Raman, E. P.; Paul, T. J.; Hayes, R. L.; Brooks III, C. L. Automated, accurate, and scalable relative protein–ligand binding free-energy calculations using lambda dynamics. *Journal of chemical theory and computation* **2020**, *16*, 7895–7914.

(7) Gapsys, V.; Pérez-Benito, L.; Aldeghi, M.; Seeliger, D.; Van Vlijmen, H.; Tresadern, G.; De Groot, B. L. Large scale relative protein ligand binding affinities using non-equilibrium alchemy. *Chemical Science* **2020**, *11*, 1140–1152.

(8) Gapsys, V.; Hahn, D. F.; Tresadern, G.; Mobley, D. L.; Rampp, M.; de Groot, B. L. Pre-exascale computing of protein–ligand binding free energies with open source software for drug design. *Journal of chemical information and modeling* **2022**, *62*, 1172–1177.

(9) Cournia, Z.; Allen, B.; Sherman, W. Relative binding free energy calculations in drug discovery: recent advances and practical considerations. *Journal of chemical information and modeling* **2017**, *57*, 2911–2937.

(10) Fu, H.; Chipot, C.; Shao, X.; Cai, W. Standard Binding Free-Energy Calculations: How Far Are We from Automation? *The Journal of Physical Chemistry B* **2023**,

(11) Amezcua, M.; El Khoury, L.; Mobley, D. L. SAMPL7 Host–Guest Challenge Overview: Assessing the reliability of polarizable and non-polarizable methods for binding free energy calculations. *Journal of computer-aided molecular design* **2021**, *35*, 1–35.

(12) Amezcua, M.; Setiadi, J.; Ge, Y.; Mobley, D. L. An overview of the SAMPL8 host–guest binding challenge. *Journal of Computer-Aided Molecular Design* **2022**, *36*, 707–734.

(13) Ross, G. A.; Lu, C.; Scarabelli, G.; Albanese, S. K.; Houang, E.; Abel, R.; Harder, E. D.; Wang, L. The maximal and current accuracy of rigorous protein-ligand binding free energy calculations. *Communications Chemistry* **2023**, *6*, 222.

(14) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **1977**, *23*, 187–199.

(15) Ferrenberg, A. M.; Swendsen, R. H. Optimized monte carlo data analysis. *Computers in Physics* **1989**, *3*, 101–104.

(16) 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. *Journal of computational chemistry* **1992**, *13*, 1011–1021.

(17) Woo, H.-J.; Roux, B. Calculation of absolute protein–ligand binding free energy from computer simulations. *Proceedings of the National Academy of Sciences* **2005**, *102*, 6825–6830.

(18) Doudou, S.; Burton, N. A.; Henchman, R. H. Standard free energy of binding from a

one-dimensional potential of mean force. *Journal of chemical theory and computation* **2009**, *5*, 909–918.

(19) Fu, H.; Cai, W.; Hénin, J.; Roux, B.; Chipot, C. New coarse variables for the accurate determination of standard binding free energies. *Journal of chemical theory and computation* **2017**, *13*, 5173–5178.

(20) Markthaler, D.; Jakobtorweihen, S.; Hansen, N. Lessons learned from the calculation of one-dimensional potentials of mean force [article v1. 0]. *Living Journal of Computational Molecular Science* **2019**, *1*, 11073–11073.

(21) Govind Kumar, V.; Polasa, A.; Agrawal, S.; Kumar, T. K. S.; Moradi, M. Binding affinity estimation from restrained umbrella sampling simulations. *Nature Computational Science* **2023**, *3*, 59–70.

(22) Blazhynska, M.; Goulard Coderc de Lacam, E.; Chen, H.; Roux, B.; Chipot, C. Hazardous shortcuts in standard binding free energy calculations. *The Journal of Physical Chemistry Letters* **2022**, *13*, 6250–6258.

(23) Lemkul, J. A.; Bevan, D. R. Assessing the stability of Alzheimer's amyloid protofibrils using molecular dynamics. *The Journal of Physical Chemistry B* **2010**, *114*, 1652–1660.

(24) Justin, A. L. From proteins to perturbed Hamiltonians: a suite of tutorials for the GROMACS-2018 molecular simulation package. *Living J Comp Mol Sci* **2018**, *1*, 5068–5120.

(25) Lindahl, V.; Lidmar, J.; Hess, B. Accelerated weight histogram method for exploring free energy landscapes. *The Journal of chemical physics* **2014**, *141*, 044110.

(26) Periole, X.; Zeppelin, T.; Schiøtt, B. Dimer interface of the human serotonin transporter and effect of the membrane composition. *Scientific reports* **2018**, *8*, 5080.

(27) Laio, A.; Gervasio, F. L. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports on Progress in Physics* **2008**, *71*, 126601.

(28) Lelièvre, T.; Rousset, M.; Stoltz, G. Computation of free energy profiles with parallel adaptive dynamics. *The Journal of chemical physics* **2007**, *126*, 134111.

(29) Invernizzi, M.; Parrinello, M. Rethinking metadynamics: From bias potentials to probability distributions. *The journal of physical chemistry letters* **2020**, *11*, 2731–2736.

(30) Bonati, L.; Piccini, G.; Parrinello, M. Deep learning the slow modes for rare events sampling. *Proceedings of the National Academy of Sciences* **2021**, *118*, e2113533118.

(31) Ansari, N.; Rizzi, V.; Parrinello, M. Water regulates the residence time of Benzamidine in Trypsin. *Nature Communications* **2022**, *13*, 5438.

(32) Reif, M. M.; Zacharias, M. Improving the Potential of Mean Force and Nonequilibrium Pulling Simulations by Simultaneous Alchemical Modifications. *Journal of Chemical Theory and Computation* **2022**, *18*, 3873–3893.

(33) Limongelli, V. Ligand binding free energy and kinetics calculation in 2020. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2020**, *10*, e1455.

(34) Hénin, J.; Lelièvre, T.; Shirts, M. R.; Valsson, O.; Delemotte, L. Enhanced sampling methods for molecular dynamics simulations. *arXiv preprint arXiv:2202.04164* **2022**,

(35) Chen, H.; Chipot, C. Enhancing sampling with free-energy calculations. *Current Opinion in Structural Biology* **2022**, *77*, 102497.

(36) Ahmad, K.; Rizzi, A.; Capelli, R.; Mandelli, D.; Lyu, W.; Carloni, P. Enhanced-sampling simulations for the estimation of ligand binding kinetics: current status and perspective. *Frontiers in molecular biosciences* **2022**, *9*.

(37) Hub, J. S.; De Groot, B. L.; Van Der Spoel, D. g_wham A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *Journal of chemical theory and computation* **2010**, *6*, 3713–3720.

(38) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25.

(39) Grossfield, A. WHAM: the weighted histogram analysis method. 2012.

(40) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham III, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications* **1995**, *91*, 1–41.

(41) Case, D. A.; Cheatham III, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz Jr, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *Journal of computational chemistry* **2005**, *26*, 1668–1688.

(42) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber biomolecular simulation package. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2013**, *3*, 198–210.

(43) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S. a.; Karplus, M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry* **1983**, *4*, 187–217.

(44) Brooks, B. R.; Brooks III, C. L.; Mackerell Jr, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S., et al. CHARMM: the biomolecular simulation program. *Journal of computational chemistry* **2009**, *30*, 1545–1614.

(45) Berendsen, H. J.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer physics communications* **1995**, *91*, 43–56.

(46) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. GROMACS: fast, flexible, and free. *Journal of computational chemistry* **2005**, *26*, 1701–1718.

(47) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation* **2008**, *4*, 435–447.

(48) Nelson, M. T.; Humphrey, W.; Gursoy, A.; Dalke, A.; Kalé, L. V.; Skeel, R. D.; Schulten, K. NAMD: a parallel, object-oriented molecular dynamics program. *The International Journal of Supercomputer Applications and High Performance Computing* **1996**, *10*, 251–268.

(49) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *Journal of computational chemistry* **2005**, *26*, 1781–1802.

(50) Phillips, J. C.; Hardy, D. J.; Maia, J. D.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W., et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *The Journal of chemical physics* **2020**, *153*, 044130.

(51) Bhardwaj, V. K.; Purohit, R. A lesson for the maestro of the replication fork: targeting the protein-binding interface of proliferating cell nuclear antigen for anticancer therapy. *Journal of Cellular Biochemistry* **2022**, *123*, 1091–1102.

(52) Mathivanan, J.; Bai, Z.; Chen, A.; Sheng, J. Design, Synthesis, and Characterization of

a Novel 2–5-Linked Amikacin-Binding Aptamer: An Experimental and MD Simulation Study. *ACS Chemical Biology* **2022**, *17*, 3478–3488.

(53) Ramasanoff, R. R.; Sokolov, P. A. The binding model of adenosine-specific DNA aptamer: Umbrella sampling study. *Journal of Molecular Graphics and Modelling* **2023**, *118*, 108338.

(54) Pirhadi, E.; Vanegas, J. M.; Farin, M.; Schertzer, J. W.; Yong, X. Effect of local stress on accurate modeling of bacterial outer membranes using all-atom molecular dynamics. *Journal of Chemical Theory and Computation* **2022**,

(55) Sousa, C. F.; Kamal, M. A.; Richter, R.; Elamaldeniya, K.; Hartmann, R. W.; Empting, M.; Lehr, C.-M.; Kalinina, O. V. Modeling the Effect of Hydrophobicity on the Passive Permeation of Solutes across a Bacterial Model Membrane. *Journal of Chemical Information and Modeling* **2022**, *62*, 5023–5033.

(56) Tam, N. M.; Nguyen, T. H.; Ngan, V. T.; Tung, N. T.; Ngo, S. T. Unbinding ligands from SARS-CoV-2 Mpro via umbrella sampling simulations. *Royal Society Open Science* **2022**, *9*, 211480.

(57) Chaudhary, Y.; Bhimalapuram, P. Insulin aspart dimer dissociation in water. *The Journal of Chemical Physics* **2022**, *156*, 105106.

(58) De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of molecular dynamics and related methods in drug discovery. *Journal of medicinal chemistry* **2016**, *59*, 4035–4061.

(59) You, W.; Tang, Z.; Chang, C.-E. A. Potential mean force from umbrella sampling simulations: What can we learn and what is missed? *Journal of chemical theory and computation* **2019**, *15*, 2433–2443.

(60) Dickson, B. M.; Legoll, F.; Lelievre, T.; Stoltz, G.; Fleurat-Lessard, P. Free energy calculations: An efficient adaptive biasing potential method. *The Journal of Physical Chemistry B* **2010**, *114*, 5823–5830.

(61) Kim, J.; Straub, J. E.; Keyes, T. Statistical-temperature Monte Carlo and molecular dynamics algorithms. *Physical review letters* **2006**, *97*, 050601.

(62) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proceedings of the national academy of sciences* **2002**, *99*, 12562–12566.

(63) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Physical review letters* **2008**, *100*, 020603.

(64) Bartels, C.; Karplus, M. Multidimensional adaptive umbrella sampling: Applications to main chain and side chain peptide conformations. *Journal of Computational Chemistry* **1997**, *18*, 1450–1462.

(65) Piana, S.; Robustelli, P.; Tan, D.; Chen, S.; Shaw, D. E. Development of a force field for the simulation of single-chain proteins and protein–protein complexes. *Journal of chemical theory and computation* **2020**, *16*, 2494–2507.

(66) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D., et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS computational biology* **2017**, *13*, e1005659.

(67) Huang, J.; MacKerell Jr, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of computational chemistry* **2013**, *34*, 2135–2145.

(68) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.;

Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function, and Bioinformatics* **2010**, *78*, 1950–1958.

(69) Buckle, A. M.; Schreiber, G.; Fersht, A. R. Protein-protein recognition: Crystal structural analysis of a barnase-barstar complex at 2.0-. ANG. resolution. *Biochemistry* **1994**, *33*, 8878–8889.

(70) Yang, F.; Gustafson, K. R.; Boyd, M. R.; Wlodawer, A. Crystal structure of Escherichia coli HdeA. *Nature structural biology* **1998**, *5*, 763–764.

(71) Román-Hernández, G.; Grant, R. A.; Sauer, R. T.; Baker, T. A. Molecular basis of substrate selection by the N-end rule adaptor protein ClpS. *Proceedings of the National Academy of Sciences* **2009**, *106*, 8888–8893.

(72) Marquart, M.; Walter, J.; Deisenhofer, J.; Bode, W.; Huber, R. The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Crystallographica Section B: Structural Science* **1983**, *39*, 480–490.

(73) Lührs, T.; Ritter, C.; Adrian, M.; Riek-Loher, D.; Bohrmann, B.; Döbeli, H.; Schubert, D.; Riek, R. 3D structure of Alzheimer's amyloid-β (1–42) fibrils. *Proceedings of the National Academy of Sciences* **2005**, *102*, 17342–17347.

(74) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I., et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of computational chemistry* **2010**, *31*, 671–690.

(75) Gutiérrez, I. S.; Lin, F.-Y.; Vanommeslaeghe, K.; Lemkul, J. A.; Armacost, K. A.; Brooks III, C. L.; MacKerell Jr, A. D. Parametrization of halogen bonds in the CHARMM general force field: Improved treatment of ligand–protein interactions. *Bioorganic & medicinal chemistry* **2016**, *24*, 4812–4825.

(76) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *Journal of computational chemistry* **2004**, *25*, 1656–1676.

(77) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The protein data bank. *Nucleic acids research* **2000**, *28*, 235–242.

(78) Durell, S. R.; Brooks, B. R.; Ben-Naim, A. Solvent-induced forces between two hydrophilic groups. *The Journal of Physical Chemistry* **1994**, *98*, 2198–2202.

(79) Neria, E.; Fischer, S.; Karplus, M. Simulation of activation free energies in molecular systems. *The Journal of chemical physics* **1996**, *105*, 1902–1921.

(80) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **1983**, *79*, 926–935.

(81) Lemkul, J. A. From proteins to perturbed hamiltonians: A suite of tutorials for the GROMACS-2018 molecular simulation package [article v1. 0]. *Living Journal of Computational Molecular Science* **2018**, *1*, 5068.

(82) Berendsen, H. J.; Postma, J. P.; van Gunsteren, W. F.; Hermans, J. Interaction models for water in relation to protein hydration. Intermolecular forces: proceedings of the fourteenth Jerusalem symposium on quantum chemistry and biochemistry held in jerusalem, israel, april 13–16, 1981. 1981; pp 331–342.

(83) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *The Journal of chemical physics* **1993**, *98*, 10089–10092.

(84) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *The Journal of chemical physics* **1995**, *103*, 8577–8593.

(85) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of chemical physics* **2007**, *126*, 014101.

(86) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics* **1981**, *52*, 7182–7190.

(87) Bernetti, M.; Bussi, G. Pressure control using stochastic cell rescaling. *The Journal of Chemical Physics* **2020**, *153*, 114107.

(88) Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G. LINCS: a linear constraint solver for molecular simulations. *Journal of computational chemistry* **1997**, *18*, 1463–1472.

(89) Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of computational chemistry* **1992**, *13*, 952–962.

(90) Wan, S.; Bhati, A. P.; Wade, A. D.; Coveney, P. V. Ensemble-Based Approaches Ensure Reliability and Reproducibility. *Journal of Chemical Information and Modeling* **2023**,

(91) Ahlstrom, L. S.; Dickson, A.; Brooks III, C. L. Binding and folding of the small bacterial chaperone HdeA. *The Journal of Physical Chemistry B* **2013**, *117*, 13219–13225.

(92) Ahlstrom, L. S.; Law, S. M.; Dickson, A.; Brooks III, C. L. Multiscale modeling of a conditionally disordered pH-sensing chaperone. *Journal of molecular biology* **2015**, *427*, 1670–1680.

(93) Best, R. B.; Hummer, G.; Eaton, W. A. Native contacts determine protein folding mechanisms in atomistic simulations. *Proceedings of the National Academy of Sciences* **2013**, *110*, 17874–17879.

(94) Schiebel, J.; Gaspari, R.; Wulsdorf, T.; Ngo, K.; Sohn, C.; Schrader, T. E.; Cavalli, A.; Ostermann, A.; Heine, A.; Klebe, G. Intriguing role of water in protein-ligand binding studied by neutron crystallography on trypsin complexes. *Nature communications* **2018**, *9*, 3559.

(95) Setny, P.; Wang, Z.; Cheng, L.-T.; Li, B.; McCammon, J.; Dzubiella, J. Dewetting-controlled binding of ligands to hydrophobic pockets. *Physical review letters* **2009**, *103*, 187801.

(96) Baron, R.; Setny, P.; McCammon, J. A. Water in cavity- ligand recognition. *Journal of the American Chemical Society* **2010**, *132*, 12091–12097.

(97) Pan, A. C.; Borhani, D. W.; Dror, R. O.; Shaw, D. E. Molecular determinants of drug–receptor binding kinetics. *Drug discovery today* **2013**, *18*, 667–673.

(98) Golla, V. K.; Prajapati, J. D.; Joshi, M.; Kleinekathofer, U. Exploration of free energy surfaces across a membrane channel using metadynamics and umbrella sampling. *Journal of Chemical Theory and Computation* **2020**, *16*, 2751–2765.

(99) Lapierre, J.; Hub, J. S. Converging PMF Calculations of Antibiotic Permeation across an Outer Membrane Porin with Subkilocalorie per Mole Accuracy. *Journal of Chemical Information and Modeling* **2023**, *63*, 5319–5330, PMID: 37560945.

(100) Nishikawa, N.; Han, K.; Wu, X.; Tofoleanu, F.; Brooks, B. R. Comparison of the umbrella sampling and the double decoupling method in binding free energy predictions for SAMPL6 octa-acid host–guest challenges. *Journal of computer-aided molecular design* **2018**, *32*, 1075–1086.

(101) Jarzynski, C. Nonequilibrium equality for free energy differences. *Physical Review Letters* **1997**, *78*, 2690.

(102) Park, S.; Schulten, K. Calculating potentials of mean force from steered molecular dynamics simulations. *The Journal of chemical physics* **2004**, *120*, 5946–5961.

(103) Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti, C.; Parrinello, M. Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics. *The journal of physical chemistry B* **2006**, *110*, 3533–3539.