JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

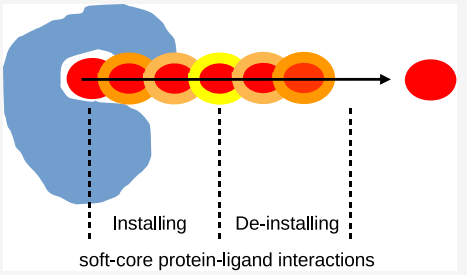
# Improving the Potential of Mean Force and Nonequilibrium Pulling Simulations by Simultaneous Alchemical Modifications

Maria M. Reif* and Martin Zacharias

Cite This: J. Chem. Theory Comput. 2022, 18, 3873−3893 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** We present an approach combining alchemical modifications and physical-pathway methods to calculate absolute binding free energies. The employed physical-pathway method is either a stratified umbrella sampling to calculate a potential of mean force or nonequilibrium pulling. We devised two basic approaches: the simultaneous approach (S-approach), where, along the physical unbinding pathway, an alchemical transformation of ligand−protein interactions is installed and deinstalled, and the prior-plus-simultaneous approach (PPS-approach), where, prior to the physical-pathway simulation, an alchemical transformation of ligand−protein interactions is installed in the binding site and deinstalled during the physical-pathway simulation. Using a mutant of T4 lysozyme with a benzene ligand as an example, we show that installation and deinstallation of soft-core interactions concurrent with physical ligand unbinding (S-approach) allow successful potential of mean force calculations and nonequilibrium pulling simulations despite the problems posed by the occluded nature of the lysozyme binding pocket. Good agreement between the potential of the mean-force-based S-approach and double decoupling simulations as well as a remarkable efficiency and accuracy of the nonequilibrium-pulling-based S-approach is found. The latter turned out to be more compute-efficient than the potential of mean force calculation by approximately 70%. Furthermore, we illustrate the merits of reducing ligand−protein interactions prior to potential of mean force calculations using the murine double minute homologue protein MDM2 with a p53-derived peptide ligand (PPS-approach). Here, the problem of breaking strong interactions in the binding pocket is transferred to a prior alchemical transformation that reduces the free-energy barrier between the bound and unbound state in the potential of mean force. Besides, disentangling physical ligand displacement from the deinstallation of ligand−protein interactions was seen to allow a more uniform sampling of distance histograms in the umbrella sampling. In the future, physical ligand unbinding combined with simultaneous alchemical modifications may prove useful in the calculation of protein−protein binding free energies, where sampling problems posed by multiple, possibly sticky interactions and potential steric clashes can thus be reduced.






## 1. INTRODUCTION

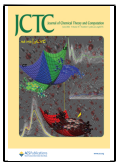
On a molecular level, a cornerstone of processes involving living organisms and inanimate matter is the reversible noncovalent association of molecules. The stability of a bound complex consisting of two associated molecular species can be characterized by the reversible work required to dissociate the complex again, the unbinding free energy, or, equivalently, its sign-reversed counterpart, the binding free energy $\Delta G_{bnd}$.

A quantification of $\Delta G_{bnd}$ helps to appreciate the strength of intermolecular interactions and may be done experimentally by isothermal titration calorimetry (ITC). These experiments, however, have some drawbacks such as high expenses due to large sample requirement, low sensitivity, low throughput, and inability to treat high-affinity complexes (binding free energies lower than ca. −70 kJ·mol$^{-1}$, implying dissociation constants $K_d$ in the picomolar range). Nevertheless, ITC is a key tool to characterize the thermodynamics of molecular recognition. For example, in drug discovery, it is used to determine the binding affinity of drug−receptor complexes, and thus, it guides hit selection and lead optimization tasks. Besides, it is the pivotal means to validate computational estimates of binding affinity.

Computational tools, in comparison to ITC, do not require physical samples; their sensitivity can be easily rationalized and optimized; and they are amenable to high-throughput applications and can also deal with high-affinity complexes. There is a plethora of different computational tools available, which, on the broadest level, may be divided into molecular docking[1,2] and molecular-simulation methods.[3−6] Recent overviews are provided in refs 4, 7, and 8.

Molecular docking, while useful to rank different ligands according to their binding free energy to a given protein, is short of molecular-simulation methods when it comes to the

Received: November 29, 2021
Published: June 2, 2022







https://doi.org/10.1021/acs.jctc.1c01194
J. Chem. Theory Comput. 2022, 18, 3873−3893

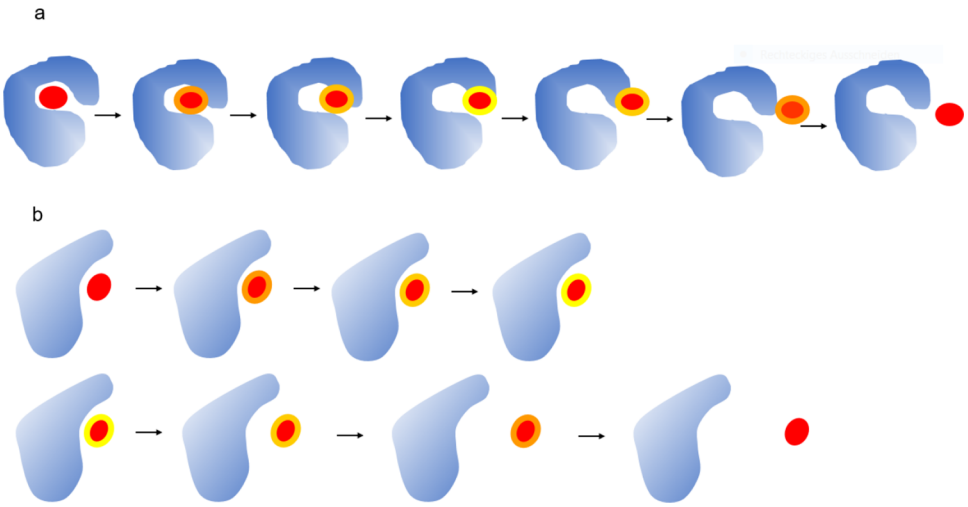



**Figure 1.** Ways to combine alchemical and physical-pathway methods applied in this work to calculate the binding free energy of a ligand (red oval) to a protein (blue entity). (a) Simultaneous approach (S-approach). During a PMF calculation, an alchemical modification is gradually installed and deinstalled so that both end states are physical. (b) Prior-plus-simultaneous approach (PPS-approach). Prior to a PMF calculation, an alchemical modification that eases the dissociation of the ligand from the binding pocket is performed in the bound state (top panel). In the successive PMF calculation, the alchemical modification is deinstalled again (bottom panel).

accurate estimation of absolute binding free energies. Generally not considering an ensemble of microstates of the species involved in the binding reaction (including an explicit representation of the solvent), it lacks a sound statistical−mechanical basis. On the other hand, molecular-simulation methods generate a Boltzmann-weighted ensemble of microstates and are in principle able to accurately capture the free energy, entropy, and enthalpy of (bio)molecular noncovalent binding reactions.

Among molecular-simulation methods, one distinguishes end-point methods, physical-pathway methods, and alchemical methods. End-point methods such as the linear interaction energy method (LIE),[9−11] the molecular-mechanics Poisson−Boltzmann surface area method (MM/PBSA),[12,13] and the related molecular-mechanics generalized-Born surface area method (MM/GBSA)[11,13,14] consider only the end points of the binding process, *i.e.*, the bound complex and its unbound components, and based on their thermodynamic features (*e.g.*, interaction energies, solvation free energies, and possibly entropic contributions), a binding free energy is calculated. Physical-pathway and alchemical methods do not involve any empirical approximations and are hence able to provide a statistically mechanically sound estimate of the binding free energy. Since they involve the simulation of intermediate states, they are, however, more compute-intense than end-point methods.

Physical-pathway methods access the binding free energy for example by counting populations of bound and unbound configurations in a long simulation or, currently more commonly done, by constructing a free-energy profile along the physical binding pathway, *i.e.*, usually along a reaction coordinate describing the distance between the two associating molecules. The latter can for instance be achieved on the basis of umbrella sampling (US),[15,16] which relies on a series of biasing potentials applied to successively increasing values of the intermolecular distance to enforce piecewise overlapping sampling of this distance along the entire range of the reaction coordinate between the bound and the solvated state of the binding partners. The difference of the free energies pertaining to the bound and the unbound state is related to the binding free energy, and the free-energy profile is related to the potential of mean force (PMF)[17] of the binding/unbinding process along the chosen reaction coordinate. Physical-pathway methods have the advantage that both binding partners can be biomacromolecules (*i.e.*, encompass hundreds of solute atoms) and that the obtained PMF can allow a mechanistic understanding of the molecular association in terms of thermodynamic and partially also kinetic features,[18] which is of great relevance because the obtained knowledge may be exploited to foster our understanding of cellular processes[19] or to fuel innovations in drug development.[20] A disadvantage of physical-pathway methods is that special care has to be taken when the binding/unbinding path involves high free-energy barriers, *e.g.*, due to burial of the binding pocket,[21] or when structural changes occur during unbinding.[22] Furthermore, large computational boxes are required to adequately describe the noninteracting unbound state.

Alchemical methods access the binding free energy through a thermodynamic cycle, where the small binding partner (ligand) is decoupled from the system once in the solvent and once in the bound complex. In the widely used double decoupling (DD) method,[23,24] the ligand is gradually, *i.e.*, along a series of unphysical intermediate states described by an interpolation of the two end-state Hamiltonians pertaining to the fully interacting ligand molecule and a ″dummy″ ligand exempt from interactions with the environment, mutated into the latter, non-interacting species. The difference of the ligand decoupling free energies in the pure solvent environment and the complex is related to the binding free energy. Alchemical methods have the advantage that they can be easily carried out in the case of buried binding pockets. Disadvantages are the restriction to relatively small ligands (not larger than ca. 50 atoms) and that special care has to be taken in the case of charged ligand molecules.[25−27]

Recently, the potential of combining alchemical and physical-pathway methods has been realized. Wu *et al.*[28] developed the alchemical transfer method (ATM), which is based on a thermodynamic cycle involving a ″symmetric



https://doi.org/10.1021/acs.jctc.1c01194
J. Chem. Theory Comput. 2022, 18, 3873−3893

intermediate″ between the bound and unbound state that features the ligand being present half in solvent and half in the complex, where ″half″ implies a coupling parameter value of $\lambda = 0.5$ pertaining to a $\lambda$-dependent hybrid alchemical potential. Due to having only half of the normal strength of interaction, this state is exempt from severe steric clashes between the ligand and the environment. The hybrid alchemical potential also depends on a mapping vector that achieves a rigid translation of all ligand atoms from the binding pocket into the solvent. An asymmetric inclusion of soft-core interactions is used, meaning that a singularity can only be treated at one end point. A disadvantage of the method is that, at each time step, it calculates the energies and forces twice, namely, once in the binding pocket and once translated into the solvent.

Cruz *et al.*[29] combined alchemical and pathway methods in that prior to a PMF calculation of ligand unbinding, problematic atoms (*e.g.*, due to steric clashes) are decoupled or transformed into soft-core atoms in an alchemical free-energy calculation *via* thermodynamic integration. With the resulting modified ligand−protein interaction potential, it is easier to carry out umbrella-sampling-based PMF calculations. Once the ligand is immersed in the bulk solvent, the initial, unmodified interaction potential is reinstalled. The method is referred to as ″AlchemPMF″.[29]

In this work, we also combine alchemical and physical-pathway methods. This can be done in a simultaneous approach (S-approach) where, during a PMF calculation, a simultaneous alchemical modification is performed in that the end states are physical (Figure 1a). This means that, along the physical pathway, an alchemical modification is gradually installed and deinstalled again. The crucial difference to the AlchemPMF method of Cruz *et al.*[29] is that we carry out the alchemical modification along with the umbrella sampling; *i.e.,* the S-approach combines a transformation in the alchemical space with the physical translocation of the ligand and thus circumvents the need to do alchemical free-energy calculations at the end points of the physical pathway. A second possibility is a prior-plus-simultaneous approach (PPS-approach). This approach bears resemblance to the AlchemPMF approach[29] in that the installation of the modified protein−ligand interaction potential is performed in the bound state prior to the physical unbinding process. This alchemical modification eases the dissociation of the ligand from the binding pocket (Figure 1b, top panel). However, again, the crucial difference to the AlchemPMF approach[29] is that the alchemical modification is deinstalled again (Figure 1b, bottom panel) during the successive PMF calculation rather than afterward. In both our S- and PPS-approach, as well as in the work of Cruz *et al.*,[29] thermodynamic integration (TI) and the weighted histogram analysis method (WHAM)[30,31] are used as free-energy estimators for the alchemical- and pathway-related free-energy contributions, respectively.

Finally, we also combine alchemical and physical-pathway methods in the simultaneous approach using multiple nonequilibrium pulling simulations along with the Jarzynski identity as a free-energy estimator. Thus, during the pulling simulations, not only the ligand−protein distance is increased, but also ligand−protein intermolecular interactions are gradually modified and then restored so that the end state corresponds to physical interactions. The Jarzynski identity has been used before to determine absolute protein−ligand binding free energies[32−37] or docking scores[38] from nonequilibrium pulling simulations. Overall, it was seen to provide a reliable ranking of ligand binding free energies, although deviations from experimental values may in certain cases still fail chemical accuracy criteria.[37] The Jarzynski identity has also been used to determine relative[39] or absolute protein−ligand binding free energies from nonequilibrium alchemical changes[40−42] and was seen to give reliable results comparable to equilibrium free-energy calculations.[41]

Combining alchemical and physical-pathway methods has the advantages of facilitating and increasing the applicability of physical-pathway methods. By virtue of reducing intermolecular interactions, it may hence (i) be easier to pull apart the binding partners since energy barriers due to strong intermolecular interactions are reduced and (ii) be possible to pull a ligand out of an occluded binding pocket for which the choice of an unbinding reaction coordinate is nontrivial. The reduction of interactions referred to above can respectively be achieved by (i) decreasing the magnitude of partial charges and the well depth of the Lennard−Jones potential, with the aim to weaken intermolecular interactions, and (ii) introducing soft cores in the electrostatic and Lennard−Jones potential, with the aim to reduce steric clashes.

The first objective of this study is to illustrate the power of combining alchemical and physical-pathway methods in these two scenarios, with the latter method being an US-based PMF calculation. For the latter scenario, we use a benzene ligand in the L99A mutant of T4 lysozyme. This system has an occluded binding pocket. The L99A mutant of T4 lysozyme is often used as a model system for the application of binding free energy calculations,[21,41−48] although these simulations can pose challenging conformational sampling problems due to structural changes near the binding site (loop movement between the unbound and bound state, helix rearrangement, and side chain reorientation depending on the ligand size).[42,46,49] We show that it is possible to determine the binding free energy based on a PMF calculation where the US is combined with a simultaneous softening of protein−ligand interactions (S-approach; Figure 1a). For the former scenario, we use a peptide ligand (p53-derived peptide) bound to the MDM2 protein. This system describes a high-affinity protein−protein interaction of high relevance to potential oncological therapies because it is representative of the complex between the tumor-suppressor protein p53 and its negative regulator protein MDM2.[50,51] It features strong nonpolar peptide−protein interactions,[52] and we illustrate how prior weakening of peptide−protein interactions can alleviate a successive PMF calculation (PPS-approach; Figure 1b).

The second objective is to illustrate the power of combining alchemical and physical-pathway methods where the physical-pathway method consists of nonequilibrium pulling simulations. Here, we show that softening ligand−protein interactions allows successful application of the Jarzynski identity even in the difficult case of an occluded binding pocket. Formally, this corresponds to the application of the S-approach in the course of nonequilibrium pulling rather than in the course of US as done during the calculation of a reversible work profile.

The paper is organized as follows: In Section 2, we explain our approach to introduce alchemical modifications in an US-based PMF calculation (Section 2.1) or in nonequilibrium pulling simulations (Section 2.2). The simulated systems, simulation, and analysis details are presented in Section 3. Results are shown in Section 4, and finally, concluding remarks are given in Section 5.



https://doi.org/10.1021/acs.jctc.1c01194
J. Chem. Theory Comput. 2022, 18, 3873−3893

## 2. METHODS

### 2.1. Alchemically Modified Potential of Mean Force.

A PMF expresses free energy as a function of a reaction coordinate. In the context of biomolecular association, the reaction coordinate is usually the distance $\xi$ between the binding partners. A common method to calculate PMFs is to perform US simulations. This involves $N$ biased simulations in which the system is restrained to a particular value $\xi_{o,i}$, $i = 1, ...N$ of the reaction coordinate. The free-energy profile $\Delta G(\xi)$ along the reaction coordinate can then be constructed for instance with WHAM.[30,31] In the PMF calculations performed here, the reaction coordinate was the distance $\xi$ along a Cartesian coordinate axis. Each of the $N$ simulations involved a harmonic restraining potential

$$V(\xi; \xi_{o,i}) = \frac{1}{2}k(\xi - \xi_{o,i})^2 \tag{1}$$

to a distance $\xi_{o,i}$, $i = 1, ...N$ with force constant $k$. In addition, the interaction potential between a subset of the atoms was modified in an alchemical $\lambda$-coupling parameter formalism. Two types of alchemical modification were conceived of: either an installation and deinstallation of ligand−protein soft-core interactions or a reduction and restoration of selected ligand−protein intermolecular interactions along the reaction coordinate. The entire procedure can be carried out as either the S- or the PPS-approach (Figure 1). Note that the inclusion of an alchemical modification during the US alters the free-energy profile. This is reflected in a change of the corresponding distance probability distributions (histograms) per US window. Since the shape of the individual histograms is altered, the histogram overlap is also altered, which may be beneficial for the concatenation of the free-energy profile (Section 3.2).

*2.1.1. Modification of Soft-Core Interactions.* In the employed GROMOS simulation package,[53] a $\lambda$-dependent van der Waals interaction potential $V^{vdw}(r_{ij}, \lambda)$ depends on a coupling parameter $\lambda$ that varies from 0 to 1 along the examined range of the reaction coordinate such that soft cores[54,55] can be switched on or off in a user-defined manner[53]

$$V^{vdw}(r_{ij}, \lambda) = V^{LJ}(r_{ij}, \Lambda_{IJ}^{sft,vdw}(\lambda)) \tag{2}$$

where $r_{ij}$ is the distance between atoms $i$ and $j$.

$$\Lambda_{IJ}^{sft,vdw}(\lambda) = \sum_{k=0}^{k=4} a_{k,IJ}\lambda^k \tag{3}$$

is a fourth-order polynomial of $\lambda$ with coefficients $a_{k,IJ}$ dependent on interacting energy groups $I$ (to which atom $i$ belongs) and $J$ (to which atom $j$ belongs), and

$$V^{LJ}(r_{ij}, \Lambda_{IJ}^{sft,vdw}(\lambda)) = \left[\frac{C_{12,ij}}{\alpha C_{126,ij}(\Lambda_{IJ}^{sft,vdw}(\lambda))^2 + r_{ij}^6} - C_{6,ij}\right] \times (\alpha C_{126,ij}(\Lambda_{IJ}^{sft,vdw}(\lambda))^2 + r_{ij}^6)^{-1} \tag{4}$$

is a soft-core Lennard−Jones potential,[54,55] with $C_{12,ij}$ and $C_{6,ij}$ being the Lennard−Jones parameters for the interaction between atoms $i$ and $j$ and $C_{126,ij}$ defined as $C_{126,ij} = 0$ for $C_{6,ij} = 0$ and $C_{126,ij} = C_{6,ij}^{-1}C_{12,ij}$ otherwise. The softness parameter $\alpha$ determines the magnitude of the distance-independent offset in the denominator. Here, $\alpha = 0.5$.

Similarly, for electrostatic interactions, the interaction potential is[53]

$$V^{ele}(r_{ij}, \lambda) = V^{RF}(r_{ij}, \Lambda_{IJ}^{sft,ele}(\lambda)) \tag{5}$$

where

$$\Lambda_{IJ}^{sft,ele}(\lambda) = \sum_{k=0}^{k=4} a_{k,IJ}\lambda^k \tag{6}$$

and a soft-core electrostatic interaction potential with reaction-field correction[56] was used:

$$V^{RF}(r_{ij}, \Lambda_{IJ}^{sft,ele}(\lambda)) = \frac{q_i q_j}{4\pi\epsilon_o}\left[\frac{1}{(\alpha(\Lambda_{IJ}^{sft,ele}(\lambda))^2 + r_{ij}^2)^{1/2}} - \frac{\frac{1}{2}C_{RF}r_{ij}^2}{(\alpha(\Lambda_{IJ}^{sft,ele}(\lambda))^2 + R_{RF}^2)^{3/2}} - \frac{1 - \frac{1}{2}C_{RF}}{R_{RF}}\right] \tag{7}$$

where $q_i$ is the partial charge of atom $i$, $\epsilon_o$ is the vacuum permittivity, $C_{RF} = \frac{2 - 2\epsilon_{RF}}{1 + 2\epsilon_{RF}}$, $\epsilon_{RF}$ is the reaction-field permittivity, and $R_{RF}$ is the reaction-field cutoff distance. Here, $\alpha = 0.5$ nm$^2$. Taken together, the $\lambda$-dependent potential energy becomes

$$V(\mathbf{r}^{3N}, \lambda) = \sum_{i,j}[V^{LJ}(r_{ij}, \Lambda_{IJ}^{sft,vdw}(\lambda)) + V^{RF}(r_{ij}, \Lambda_{IJ}^{sft,ele}(\lambda))] \tag{8}$$

where the sum runs over all perturbed pairs $i$, $j$ and $\mathbf{r}^{3N}$ denotes the positions of all $N$ atoms in the system.

In the present work, a modification of soft-core interactions as described above was applied in the context of the S-approach (Figure 1a). The inclusion of this modification during the US entails an additional free-energy contribution, here denoted $\Delta G_{sft}$, that has to be considered in the calculation of the binding free-energy (Section 3.4). This additional free-energy contribution vanishes only in very special cases (*e.g.*, by construction due to a symmetric PMF in conjunction with symmetric $\Lambda_{IJ}^{sft,vdw}(\lambda)$ and $\Lambda_{IJ}^{sft,ele}(\lambda)$ functions). Otherwise, it can be calculated, *e.g.*, *via* TI:

$$\Delta G_{sft} = \int_0^1 d\lambda' \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_{\lambda'} \tag{9}$$

where the $\lambda$-dependent terms in the Hamiltonian are given by eqs 4 and 7, $\langle\ \rangle_{\lambda'}$ denotes ensemble averaging at a value $\lambda = \lambda'$, and the simulations at the individual $\lambda$-values include the US potentials used in the PMF calculation.

The functions $\Lambda_{IJ}^{sft,vdw}(\lambda)$ and $\Lambda_{IJ}^{sft,ele}(\lambda)$ were chosen such that they vanish for $\lambda = 0$ and $\lambda = 1$ and that they have a maximum at $\lambda_m$, $0 < \lambda_m < 1$. This means that

(i) At the end points ($\lambda = 0,1$), interactions between energy groups $I$ and $J$ are described with the "normal" Lennard−Jones and electrostatic interaction potentials, *i.e.*, the $\Lambda$-dependent terms in eqs 4 and 7 vanish.

(ii) At intermediate points, interactions between energy groups $I$ and $J$ are softened according to the softness



https://doi.org/10.1021/acs.jctc.1c01194
J. Chem. Theory Comput. 2022, 18, 3873−3893

parameter $\alpha$. The maximum softness is reached at $\lambda = \lambda_m$.

The functions $\Lambda_{IJ}^{\mathrm{sft,\ vdw}}(\lambda)$ and $\Lambda_{IJ}^{\mathrm{sft,\ ele}}(\lambda)$ have to be chosen such that $\lambda_m$ corresponds to a region of the reaction coordinate where computational benefits arise from the inclusion of maximum softness. This may be the case where the friction between energy groups $I$ and $J$ representing the protein and the ligand, respectively, is high due to steric clashes and/or molecular rearrangement processes.

*2.1.2. Modification of Interactions without Softness Involvement.* An alternative alchemical modification that was applied in the present work during an US-based PMF calculation is the reduction and restoration of selected ligand−protein intermolecular interactions along the reaction coordinate. The alchemical modification followed a standard single-perturbation topology approach,[53,57] with the initial ("$A$") and final ("$B$") state differing in atom partial charges and Lennard−Jones parameters. The end state in the perturbation topology has reduced Lennard−Jones potential well depths and reduced partial charges, whereas the initial state has physical interactions. Note that these are states defined concerning the perturbation topology and do not necessarily correspond to the initial and final states of the alchemical transformation carried out along the unbinding pathway, where the particular choice of Λ-function can actually exchange the states (see below).

Akin to the formalism discussed in Section 2.1.1, albeit now without soft cores, *i.e.*, $\alpha = 0$, the course of the alchemical modification was here determined by functions[53]

$$\Lambda_{IJ}^{\mathrm{vdw}}(\lambda) = \sum_{k=0}^{k=4} a_{k,IJ}\lambda^k \tag{10}$$

and

$$\Lambda_{IJ}^{\mathrm{ele}}(\lambda) = \sum_{k=0}^{k=4} a_{k,IJ}\lambda^k \tag{11}$$

Here, the parameters of the interaction potentials differ between the initial state and the final state, *i.e.*,

$$V^{\mathrm{LJ}}(r_{ij}, X) = \frac{C_{12,ij}^{X}}{r_{ij}^{12}} - \frac{C_{6,ij}^{X}}{r_{ij}^{6}} \tag{12}$$

and

$$V^{\mathrm{RF}}(r_{ij}, X) = \frac{q_i^X q_j^X}{4\pi\epsilon_o}\left[\frac{1}{r_{ij}} - \frac{\frac{1}{2}C_{\mathrm{RF}}r_{ij}^2}{R_{\mathrm{RF}}^3} - \frac{1-\frac{1}{2}C_{\mathrm{RF}}}{R_{\mathrm{RF}}}\right] \tag{13}$$

where $X = A$ or $B$ and, taken together, the $\lambda$-dependent potential energy becomes

$$V(\mathbf{r}^{3N}, \lambda) = \sum_{i,j}[\Lambda_{IJ}^{\mathrm{vdw}}(\lambda)V^{\mathrm{LJ}}(r_{ij}, B) + (1-\Lambda_{IJ}^{\mathrm{vdw}}(\lambda))V^{\mathrm{LJ}}(r_{ij}, A)$$

$$+\Lambda_{IJ}^{\mathrm{ele}}(\lambda)V^{\mathrm{RF}}(r_{ij}, B) + (1-\Lambda_{IJ}^{\mathrm{ele}}(\lambda))V^{\mathrm{RF}}(r_{ij}, A)], \tag{14}$$

where the sum runs over all perturbed pairs $i$, $j$ and $\mathbf{r}^{3N}$ denotes the positions of all $N$ atoms in the system.

In the present work, the reduction and restoration of selected ligand−protein intermolecular interactions were applied in the context of the PPS-approach (Figure 1a); *i.e.*, in the initial state of the PMF calculation (and the associated alchemical modification), the ligand has already reduced interactions with the protein. Physical interactions are then gradually restored toward the final, fully interacting state. The corresponding free-energy contribution is here denoted $\Delta G_{\mathrm{al,pmf}}$, and along with the free energy pertaining to the prior alchemical transformation in the binding pocket, $\Delta G_{\mathrm{al,prior}}$, it has to be considered in the calculation of the binding free-energy (Section 3.4).

The functions $\Lambda_{IJ}^{\mathrm{vdw}}(\lambda)$ and $\Lambda_{IJ}^{\mathrm{ele}}(\lambda)$ were chosen such that they have a maximum for $\lambda = 0$ with $\Lambda_{IJ}^{\mathrm{vdw}}(\lambda = 0) = \Lambda_{IJ}^{\mathrm{ele}}(\lambda = 0) = 1$ and vanish for $\lambda = 1$. This means that:

(i) In the final state ($\lambda = 1$), interactions between energy groups $I$ and $J$ are described with the "normal" Lennard−Jones and electrostatic interaction potentials (state $A$).
(ii) At all other $\lambda$-values, interactions between energy groups $I$ and $J$ are reduced.
(iii) In the initial state ($\lambda = 0$), interactions between energy groups $I$ and $J$ are described with minimal Lennard−Jones and electrostatic interaction potentials (state $B$).

The functions $\Lambda_{IJ}^{\mathrm{vdw}}(\lambda)$ and $\Lambda_{IJ}^{\mathrm{ele}}(\lambda)$ are chosen such that they have a maximum at $\lambda = 0$ that declines very slowly so that along a wide range of the reaction coordinate in the vicinity of the binding site, the interactions between energy groups $I$ and $J$, representing the protein and the ligand, respectively, are close to their minimum. This is why a fourth-power parabola was used.

**2.2. Alchemically Modified Nonequilibrium Pulling Simulations.** As an alternative to using an US-based PMF calculation as the physical-pathway method, we also carried out multiple nonequilibrium pulling simulations. Here, the nonequilibrium work of unbinding is calculated by integration over a coupling parameter $\lambda$ that determines the change of a distance restraining potential along the reaction coordinate,

$$H(\lambda) = \frac{1}{2}k(\xi - \xi_o(\lambda))^2, \tag{15}$$

where $\xi_o(\lambda) = \xi_{o,A} + \lambda(\xi_{o,B} - \xi_{o,A})$ is the target distance per $\lambda$-point and subscripts $A$ and $B$ denote initial and final states, respectively. In contrast to TI, $\lambda$ varies at each time step of the simulation, $\frac{\mathrm{d}\lambda}{\mathrm{d}t} = \frac{1}{t_{\mathrm{sim}}}$, where $t_{\mathrm{sim}}$ is the length of the nonequilibrium pulling simulation.

In analogy to TI, one calculates the $\lambda$-derivative of the Hamiltonian:

$$\frac{\partial H}{\partial \lambda} = -k(\xi - \xi_o(\lambda))(\xi_{o,B} - \xi_{o,A}) \tag{16}$$

The resulting work is an integral over the simulation time:

$$W = \int_0^{t_{\mathrm{sim}}} \mathrm{d}t\, \frac{\partial H}{\partial \lambda}\frac{\mathrm{d}\lambda}{\mathrm{d}t} \tag{17}$$

According to the Jarzynski identity,[58,59] an equilibrium free-energy difference $\Delta G$ is then expressed as an exponential average over multiple nonequilibrium works $W_i$:

$$\Delta G = -\beta^{-1}\overline{\exp[-\beta W_i]}, \tag{18}$$

where $\beta = (k_{\mathrm{B}}T)^{-1}$, $k_{\mathrm{B}}$ is the Boltzmann constant, $T$ the absolute temperature, and $\overline{\phantom{x}}$ denotes arithmetic averaging. In practice, this method is computationally inefficient because a large number of works $W_i$ have to be accumulated to give a precise and accurate result. The underlying reason is that the average is dominated by small work values that occur very


https://doi.org/10.1021/acs.jctc.1c01194
J. Chem. Theory Comput. 2022, 18, 3873−3893

rarely in most processes of interest in condensed-phase biomolecular systems.[60−62] However, during the calculation of $W_i$, the interaction potential between two suitable energy groups $I$ and $J$ may be modified similar to the way outlined above (Section 2.1) such that small values of $W_i$ are sampled with higher probability. Hence, we expect that a statistically precise and accurate $\Delta G$ can be obtained from fewer $W_i$ calculations.

In the present work, we combined nonequilibrium pulling simulations with simultaneous modification of soft-core interactions (Section 2.1.1) using the S-approach.

## 3. COMPUTATIONAL DETAILS

**3.1. System Setup.** Two different systems, each hydrated using the SPC water model,[63] were simulated: the T4 lysozyme mutant L99A with a benzene ligand (in the following abbreviated "T4L") and the MDM2-protein p53-peptide complex (in the following abbreviated "MDM").

Initial coordinates for system T4L were based on the X-ray structure with PDB ID 4W52.[64] Protonation states were set according to pH 7. The protein was oriented such that the interface between helices 4 (residues 83−91) and 7 (residues 115−123) is approximately perpendicular to the $y$-axis and the vector connecting the benzene center of mass (com) and the com of atoms 981, 987, and 993 (CA atoms of residues A98, A99, and I100) is approximately aligned with the $y$-axis (Figure 2a). The system was solvated in a replicated array of pre-equilibrated cubic water boxes of 5.5 nm edge length. Excess water was discarded such that a final rectangular box with dimensions of 7.0 nm × 8.75 nm × 7.0 nm resulted. The system energy was minimized with the steepest-descent algorithm. Eight chloride ions were added to the solvent to neutralize the protein net charge, and the system was thermalized and equilibrated. In five successive simulation runs of 20 ps length at a constant volume, the system temperature was raised to 300 K, and position restraints that were initially applied to all protein and ligand atoms (force constant 25,000 kJ·mol$^{-1}$ · nm$^{-2}$) were gradually released. In a sixth simulation step of 6 ns length, barostatting was introduced. Throughout, a rototranslational constraint was applied to the protein to prevent a change of orientation in the rectangular box.

Initial coordinates for the PMF calculation of system T4L were generated by successive simulations of 100 ps length in which a harmonic restraint was applied to a stepwise increasing $y$-coordinate distance between the benzene com and the com of atoms 981, 987, and 993 (CA atoms of residues A98, A99, and I100). The target distances were set to 0.59, 0.67, 0.75, ..., 2.51, and 2.59 nm, respectively. Here, softness was already applied to make possible the simulations were the benzene molecule overlaps with protein atoms. This is detailed further in Section 3.3. Note that the chosen unbinding pathway of the benzene molecule corresponds to a well-known possible egress route seen in enhanced sampling simulations.[49,65]

Initial coordinates for system MDM were based on the X-ray structure with PDB ID 1T4F.[52]

Protonation states were set according to pH 7. The protein was oriented such that the helix axis of the peptide ligand points approximately along the $x$-axis (Figure 3a,b). The peptide ligand derives from the p53 protein and consists of the nine residues RFMDYWEGL. The hydrophobic side chains of residues F2, W6, and L9 point into the binding groove and form strong hydrophobic interactions with the protein (Figure 3a and ref 52). The system was solvated in a replicated array of pre-equilibrated cubic water boxes of 5.5 nm edge length. Excess water was discarded such that a final rectangular box with dimensions of 5.8 nm × 5.8 nm × 8.5 nm resulted. The system energy was minimized with the steepest-descent algorithm. Three chloride ions were added to the solvent to neutralize the protein net charge, and the system was thermalized and equilibrated. In five successive simulation runs of 20 ps length at a constant volume, the system temperature was raised to 300 K, and position restraints that were initially applied to all protein and ligand atoms (force constant 25,000 kJ·mol$^{-1}$ · nm$^{-2}$) were gradually released. In a sixth simulation step of 5 ns length, barostatting and dihedral angle restraints on the backbone $\phi$ and $\psi$ dihedral angles of the

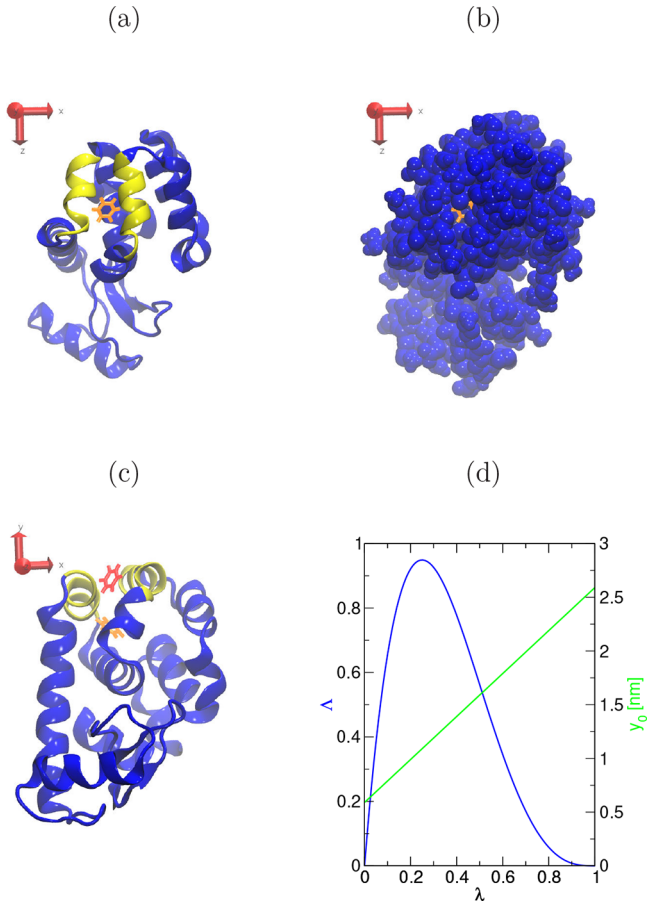



**Figure 2.** Equilibrated structure of benzene bound to the L99A mutant of T4 lysozyme. The benzene molecule is depicted in orange. (a) View where the $y$-axis points out of the plane. The protein is shown in cartoon representation, with the two helices 4 and 7 above the binding pocket highlighted in yellow (helix 4: residues 83−91; helix 7: residues 115−123). In the computational system, a rototranslational constraint is applied to the protein in the shown orientation where the benzene egress pathway through the helix−helix interface is parallel to the $y$-axis. (b) The protein is shown in space-filling bead representation to illustrate that the benzene molecule resides in an occluded binding pocket. (c) View where the $y$-axis points upward. Representative snapshot for $\lambda$ = 0.32 along the benzene−protein unbinding PMF, with the benzene molecule, now residing at the helix−helix interface, shown in red. For comparison, the benzene molecule as found in an equilibrated snapshot of the bound state is shown in orange. (d) Function $\Lambda(\lambda)$ (eqs 3 and 6) and corresponding target distances $y_0(\lambda)$ used in US or nonequilibrium pulling simulations.



https://doi.org/10.1021/acs.jctc.1c01194
J. Chem. Theory Comput. 2022, 18, 3873−3893

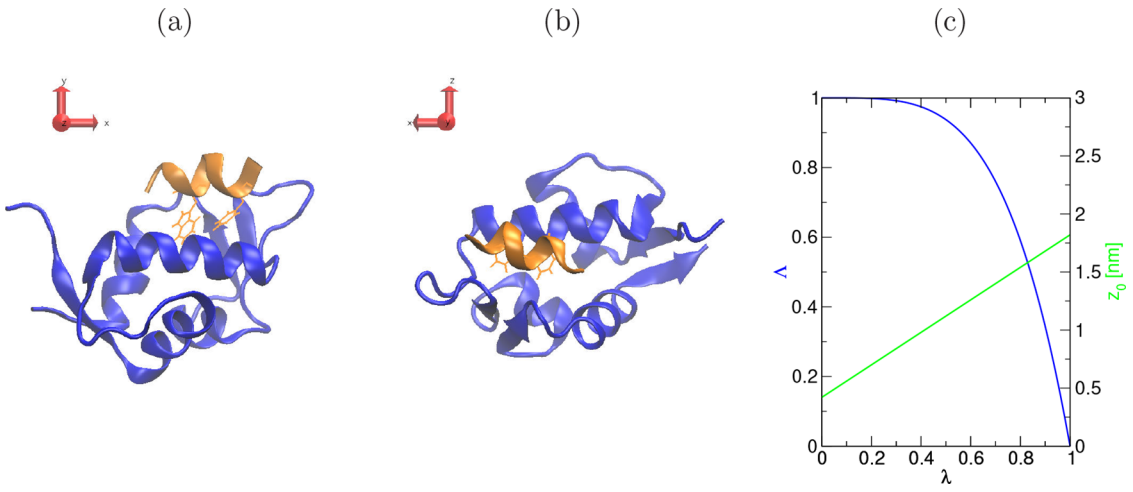



**Figure 3.** Equilibrated structure of the p53-derived peptide bound to the MDM2 protein. The peptide is shown in orange, with residues F2 and W6 highlighted in stick representation. (a) View where the *z*-axis points out of the plane. In the computational system, a rototranslational constraint is applied to the protein in the shown orientation where the hydrophobic peptide side chains approximately point along the *y*-axis. (b) View where the *z*-axis points upward. (c) Function $\Lambda(\lambda)$ (eqs 10 and 11) and corresponding target distances $z_o(\lambda)$ used in US simulations.

peptide were introduced. A harmonic potential with a force constant of 100 kJ·mol$^{-1}$ · deg$^{-2}$, linearization beyond a deviation of 40°, and reference values based on the minimized protein−peptide complex were used. The latter were 3.4, −41.6, −38.4, −36.9, −43.9, −29.6, −23.0, and − 27.4° for the $\psi$ dihedral angles and −52.7, −60.4, −61.7, −78.3, −63.3, −57.6, −67.6, and −63.1° for the $\phi$ dihedral angles of the peptide. Throughout, a rototranslational constraint was applied to the protein to prevent a change of orientation in the rectangular box.

Initial coordinates for the PMF calculation of system MDM were generated by successive simulations of 100 ps length in which a harmonic restraint was applied to a stepwise increasing *z* coordinate distance between the com of atoms 315, 412, 801, and 823 (CA atoms of protein residues L32, M40, K76, and Y78) and the com of atoms 993, 1002, and 1020 (CA atoms of peptide residues D4, Y5, and W6). The target distances were set to 0.42, 0.49, 0.56, ..., 1.75, and 1.82 nm, respectively. Here, the same protein−ligand interaction modifications as in the PMF production runs were already applied. This is detailed further in Section 3.3. A further set of distances without any interaction modifications was prepared at 1.89, 1.96, ..., 2.73, and 2.80 nm.

**3.2. Simulation Settings.** All simulations were carried out with the GROMOS11 software[53] and the GROMOS 54A8 force-field version[66] in rectangular boxes under periodic boundary conditions.

The benzene ligand in system T4L was described with partial charges from ref 67 (charges ±0.12980 *e* on CH groups). These differ slightly from charges in a phenylalanine residue (±0.14 *e* on CH groups). Temperature was maintained at a reference value of 300 K by weak coupling[68] to a heat bath using a coupling time of 0.1 ps. Distinct heat baths were used for solute and solvent (water molecules and ions) environments. Whenever required, pressure was maintained at 1 atm by weak coupling[68] to a piston using a coupling time of 0.5 ps and an isothermal compressibility of 4.575 × 10$^{-4}$ nm$^3$ kJ$^{-1}$ mol. Barostatting was performed with isotropic scaling of spatial dimensions. The equations of motion were integrated using the leap-frog scheme[69] with a time step of 2 fs. Bond lengths were constrained by the SHAKE algorithm,[70] with a relative geometric tolerance of 10$^{-4}$. Further constraints and/or restraints applied to solute molecules are explained in Section 3.1.

Electrostatic interactions were calculated using a molecule-based cutoff truncation with a Barker−Watts reaction-field correction.[56] The cutoff distance was set to 1.4 nm and the solvent relative dielectric permittivity entering in the reaction-field terms was set to 61, as appropriate for the SPC water model.[71] The van der Waals interactions were calculated using the Lennard−Jones potential, truncated at a distance of 1.4 nm. The neighbor pairlist was constructed every five time steps. Based on this pairlist, interactions at a distance smaller than 0.8 nm were computed at every time step, while interactions at a distance between 0.8 and 1.4 nm were only computed at pairlist update and kept unaltered otherwise.[72] Throughout, coordinates were written to file every 3 ps, and energies and Hamiltonian $\lambda$-derivatives were written to file every picosecond.

**3.3. Molecular Dynamics Simulations.** For the PMF calculation of system T4L, harmonic distance restraints centered at *y*-coordinate distances between the benzene com and the com of atoms 981, 987, and 993 (CA atoms of residues A98, A99, and I100) were applied. The force constant was set to 1500 kJ·mol$^{-1}$ · nm$^{-2}$. The target distances were set to 0.59, 0.67, 0.75, ..., 2.51, and 2.59 nm, respectively, and correspond to $\lambda$-values of 0.0, 0.04, 0.08, ..., 0.96, and 1.0, respectively, in the simultaneous alchemical transformation modifying the softness (of van der Waals and electrostatic interactions; eqs 4 and 7) between the protein and the benzene molecule (corresponding to energy groups *I* and *J*, respectively, in eqs 2 and 5). For the interaction between these two energy groups, the polynomials in eqs 3 and 6 are given by coefficients $a_0 = 0$, $a_1 = 9$, $a_2 = -27$, $a_3 = 27$, and $a_4 = -9$, while all other interactions were kept unaltered. Additional points at $\lambda = 0.10$ (*y*-distance 0.79 nm) and $\lambda = 0.58$ (*y*-distance 1.75 nm) were simulated to improve the histogram overlap in the PMF calculation. Throughout, orthogonal restraints along the *x*- and *z*-directions at distances of 0.02 and 0.14 nm, respectively, were used. These values are mean values determined from the last equilibration step. A force constant of 1500 kJ·mol$^{-1}$ · nm$^{-2}$ was used for the distance restraints. In addition, a


https://doi.org/10.1021/acs.jctc.1c01194
J. Chem. Theory Comput. 2022, 18, 3873−3893

harmonic distance restraint was applied between the two helices formed by residues 83 to 91 (here noted as helix 4; referred to as helix 4 in ref 73 or helix D in ref 65) as well as residues 115−123 (here noted as helix 7; referred to as helix 6 in ref 73 or helix G in ref 65) situated above the binding pocket (com of atoms 844, 856, 864, and 872 in helix 4 (CA atoms of residues K85, P86, V87 and Y88) and atoms 1153, 1161, 1170, 1187 in helix 7 (CA atoms of residues S117, L118, R119, and M120)) with a target distance of 1.05 nm (mean value determined from the last equilibration step) and a force constant of 1500 kJ·mol$^{-1}$ · nm$^{-2}$ to prevent conformational changes in the pocket closing. That is, we wanted the pocket to stay in a closed conformation to illustrate the benefits of our method and to enhance the computational efficiency by avoiding the sampling of the potential pocket opening or closure in the PMF. Thus, with the interface of the two helices restrained, the benzene ligand is buried in the protein interior (Figure 2b). At $\lambda$-values of around 0.32, it passes the helix−helix interface (Figure 2c), and the polynomial coefficients above were chosen to lead to maximum softness around this region of high steric hindrance (Figure 2d). US simulations were carried out for 70 ns per window, and the simulation period between 10 and 70 ns was used for analysis.

DD simulations were performed to compare the absolute binding free energy. This involved two sets of TI simulations performed in the solvent or in the binding site, where the interactions between the benzene ligand and the environment were gradually switched off.

The free-energy contribution associated with the annihilation of intraligand electrostatic and van der Waals interactions is, due to the rigidity of the benzene molecule, the same in both environments. The TI simulations comprised 21 equidistant points with coupling parameters $\lambda$ varying linearly from 0.0 to 1.0. For the simulation in the solvent, six additional points were inserted around the minimum of the curve (Figure S2). After a short equilibration period of 100 ps (successively per window, starting with the final structure of the bound-state equilibration), production runs of 14 ns length were carried out. For the in-protein TI simulation, a rototranslational constraint and a restraint on the closing of the binding pocket (helix 4−helix 7 interface) were used as described above in the context of the US simulations for the PMF calculation, and a harmonic distance restraint was applied between the benzene com and the com of atoms 981, 987, and 993 (CA atoms of residues A98, A99, and I100) with a target distance of 0.60 nm (mean value determined from the last equilibration step) and a force constant of 300 kJ·mol$^{-1}$ · nm$^{-2}$ to keep the ligand inside the binding pocket during the decoupling procedure. We also used an alternative scheme where a different initial structure (taken after 40 ns US in the bound state) was used, the benzene−environment electrostatic and van der Waals interactions were vanished in two successive TI simulations during which the helix 4−helix 7 interface restraint was present, and a harmonic distance restraint was applied between the benzene com and the com of atoms 973, 987, 1105, and 1161 (CA atoms of residues C97, A99, A112, and L118) with a target distance of 0.14 nm (mean value determined from the last equilibration step) and a force constant of 300 kJ·mol$^{-1}$ · nm$^{-2}$. Production runs of 8 ns for the electrostatic free-energy contribution and of 14 ns for the van der Waals free-energy contribution were carried out. The in-protein simulations were conducted in the same box size as the PMF calculation, and for the in-solvent simulations, a box containing 814 solvent molecules and 1 benzene solute was used.

For the nonequilibrium pulling simulations of system T4L, two setups with pulling velocities of 0.2 or 0.4 m·s$^{-1}$ were performed where the benzene molecule was pulled away from the binding pocket *via* a harmonic distance restraint (force constant 3000 kJ·mol$^{-1}$ · nm$^{-2}$) between the $y$-coordinate of the benzene com and the com of atoms 981, 987, and 993 (CA atoms of residues A98, A99, and I100) of continuously varying target value $y_o(\lambda)$, where $\lambda$ is a coupling parameter varying from 0 to 1 during a pulling simulation (eq 15). The distance restraint target value varied from an initial value of $y_o(\lambda = 0) = y_{o,A} = 0.59$ nm to $y_o(\lambda = 1) = y_{o,B} = 2.59$ nm during a simulation of 10 or 5 ns length (corresponding to the two different pulling velocities), with

$$y_o(\lambda) = (1 - \lambda)y_{o,A} + \lambda y_{o,B}. \tag{19}$$

Seventy pulling simulations were performed for each of the two pulling velocities. In these simulations, restraints along the $x$- and $z$-coordinates, as well as on the helix 4−helix 7 interface, were applied in analogy to the US simulations (see above), albeit with a different force constant of 3000 kJ·mol$^{-1}$ · nm$^{-2}$. Benzene−protein electrostatic and van der Waals interactions were likewise altered, meaning that, during each pulling simulation, the benzene−protein interactions were gradually softened and then desoftened again. Initial, uncorrelated coordinates were extracted every 50 ps from an equilibrium simulation of the T4L system. The time interval was chosen because it exceeds the rotational reorientation time of the ligand in the binding pocket.

For comparison, we also performed 70 nonequilibrium pulling simulations without softness modification and an otherwise identical setup to the pulling simulations with a pulling velocity of 0.2 m·s$^{-1}$.

System MDM required two types of simulations: (i) TI simulations to weaken the protein−peptide interactions in the bound state and (ii) US simulations during which the normal protein−peptide interactions were reinstalled. For the former, in a stepwise process involving 21 equidistant points with coupling parameters $\lambda$ as well as $\Lambda$ varying linearly from 0.0 to 1.0 ($\Lambda(\lambda) = \lambda$), the electrostatic interactions between peptide side chains F2 and W6 and the protein were removed, and the corresponding Lennard−Jones interactions were reduced to 10% of the original Lennard−Jones potential well depth. After a short equilibration time of 100 ps, production runs of 30 ns length were carried out. For the PMF calculations of system MDM, harmonic distance restraints centered at $z$-coordinate distances between the com of atoms 315, 412, 801, and 823 (CA atoms of protein residues L32, M40, K76, and Y78) and the com of atoms 993, 1002, and 1020 (CA atoms of peptide residues D4, Y5, and W6) were applied. The force constant was set to 1500 kJ·mol$^{-1}$ · nm$^{-2}$. The target distances were set to 0.42, 0.49, 0.56, ..., 1.75, and 1.82 nm, respectively, and correspond to $\lambda$ values of 0.0, 0.04, 0.08, ..., 0.96, and 1.0, respectively, in the simultaneous alchemical transformation reinstalling the normal van der Waals and electrostatic interactions between the protein and the F2 and W6 side chains of the peptide. For the interaction between these two energy groups, the polynomials in eqs 10 and 11 are given by coefficients $a_0 = 1$, $a_1 = 0$, $a_2 = 0$, $a_3 = 0$, and $a_4 = -1$. This fourth-order polynomial with a maximum at $\lambda = 0$ ensures that protein−peptide interactions are maximally modified near the


https://doi.org/10.1021/acs.jctc.1c01194
J. Chem. Theory Comput. 2022, 18, 3873−3893

binding site and gradually restored during the peptide unbinding process (Figure 3c). No softness was employed, *i.e.*, $\alpha = 0$, and all other interactions in the system were kept unaltered.

To prevent the peptide from tilting its helix axis with respect to the protein (and thus alleviate sampling issues), additional harmonic distance restraints were applied along the *x*- and *y*-directions between the com of atoms 315, 412, 801, and 823 (CA atoms of protein residues L32, M40, K76, and Y78) and the com of atoms 950 and 967 (CA atoms of peptide residues R1 and F2). The target distances were set to 0.97 and 0.80 nm, respectively, which correspond to mean values determined from the last equilibration step. The force constant for the distance restraints was set to 1500 kJ·mol$^{-1}$ · nm$^{-2}$. US simulations were carried out for 50 ns per window, and the period between 10 and 50 ns was used for analysis. A normal PMF calculation without alchemical modifications but otherwise identical setup was performed for comparison.

**3.4. Analysis of Simulations.** *3.4.1. T4L System.* In each window of the US simulations of system T4L, the distance coordinate $\xi$ was written to file every 3 ps, and the WHAM package of Grossfield[74] was used to calculate the free-energy profile $\Delta G(\xi)$ along the distance coordinate. Errors on the PMF were calculated from bootstrapping.[74] The PMFs were anchored to zero in the bound minimum. A raw binding free energy $\Delta G_{\mathrm{bnd,\,raw}}$ was obtained in the T4L case as

$$\Delta G_{\mathrm{bnd,raw}} = -\Delta G(\xi_{\mathrm{u}}) - \Delta G_{\mathrm{sft}} \tag{20}$$

where $\xi_{\mathrm{u}}$ is the distance in the unbound regime of the last PMF window exempt from softness modification and

$$\Delta G_{\mathrm{sft}} = \int_0^1 \mathrm{d}\lambda \left\langle \frac{\partial H(\lambda')}{\partial \lambda'} \right\rangle_\lambda \tag{21}$$

is the reversible work of softness modification along the unbinding path. Angular brackets denote ensemble averaging. Note that in the definition of $\Delta G_{\mathrm{bnd,raw}}$ (eq 20), we do not average over a certain length of the unbound regime because except for the very last PMF window (where $\lambda = 1$ and $\Lambda(\lambda = 1) = 0$; Figure 2d), the unbound state does not correspond to the physical ligand. Since only protein−ligand interactions are affected by softness modification and since these interactions can be considered vanishing in the unbound state, such averaging may actually still be justified and is not expected to affect the result if the length of the unbound regime is chosen in a region where the PMF is constant and the length of the unbound regime is properly accounted for by a corresponding standard-state correction for the unbound state. The standard binding free energy is

$$\Delta G^{o}_{\mathrm{bnd}} = \Delta G_{\mathrm{bnd,raw}} + \Delta G_{\mathrm{ores,b}} - \Delta G^{o}_{\mathrm{ores,u}} \tag{22}$$

where $\Delta G_{\mathrm{ores,b}}$ and $\Delta G^{o}_{\mathrm{ores,u}}$ are free-energy contributions for releasing the orthogonal ($x$, $z$) restraints in the bound and unbound state, respectively. The latter also converts the binding free energy to the standard volume in the unbound state. Note that $\Delta G^{o}_{\mathrm{bnd}}$ in eq 22 pertains to the protein in the presence of the helix 4−helix 7 interface restraint (Section 3.3). If the restraint energy is sufficiently small, the associated free-energy contribution can in principle be determined using a single application of the Zwanzig perturbation equation[75] to unrestrained simulations of the apo- and holo-proteins, which would finally allow a comparison of the computed binding free energy to the experimental value. In the present study, this approach was not feasible because in a long unrestrained simulation of the apo-protein, we observed a translation of helix 7 away from the native position that increased the restraint distance to up to 2−2.5 nm. This helix shift is most likely an artifact because the apo structure of lysozyme mutant L99A should be rather similar to the benzene-bound structure[64] except for small conformational changes in the helix formed by residues 107−114 (PDB ID 7L37).[76] Note that the latter helix (labeled ″F″ in refs 64, 76, and 77) is rather flexible and has been observed to play a role in the transition to a rare ″excited state″ of the apo lysozyme mutant L99A where it combines with the successive helix 7 into a single long helix after tens of microseconds.[77] This, however, was not observed here. Due to the above artifact and the sampling effort that is required due to the flexible nature of the involved loops and helices,[46] we decided not to calculate the restraint-free energy contribution in the scope of the present work but to focus on the comparison of methods to calculate the binding free energy to the restrained protein.

$\Delta G_{\mathrm{ores,b}}$ is evaluated from the Zwanzig perturbation equation:[75]

$$\Delta G_{\mathrm{ores,b}} = +\beta^{-1} \ln \langle e^{-\beta V_{\mathrm{ores}}(\mathbf{r}^{3N})} \rangle_0 \tag{23}$$

where a 10 ns simulation of the protein with the helix 4−helix 7 interface restraint (as above; Section 3.3) and with the benzene in the bound state with the PMF restraint along the *y*-direction (*i.e.* the reaction coordinate direction) but without orthogonal restraints (indicated by the subscript ″0″ of the angular brackets) and with coordinates written to file every 3 ps was used to calculate the exponential ensemble average of the restraining potential $V_{\mathrm{ores}}(\mathbf{r}^{3N})$ along the orthogonal *x*- and *z*-directions,

$$V_{\mathrm{ores}}(\mathbf{r}^{3N}) = \frac{1}{2} k[(x - x_o)^2 + (z - z_o)^2] \tag{24}$$

where $k = 1500$ kJ·mol$^{-1}$ · nm$^{-2}$, $x_o = 0.02$, and $z_o = 0.14$ nm as described above (Section 3.3). Note that the + sign in eq 23 is due to the definition of $\Delta G_{\mathrm{ores,b}}$ as the free energy of releasing rather than installing the restraints. $\Delta G^{o}_{\mathrm{ores,u}}$ is evaluated as[43,78,79]

$$\Delta G^{\mathrm{o}}_{\mathrm{ores,u}} = -\beta^{-1} \ln \frac{V_o}{\widetilde{V}}, \tag{25}$$

where $V_o = 1.66$ nm$^3$ is the volume available to the ligand in the unbound state at a reference concentration of 1 M, and

$$\widetilde{V} = \int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{d}y \int_{-\infty}^{\infty} \mathrm{d}z \; e^{-\beta V_{\mathrm{ores}}(\mathbf{r}^{3N}) - (1/2) k\beta (y - y_o)^2} \tag{26}$$

where the first term in the exponential, $V_{\mathrm{ores}}(\mathbf{r}^{3N})$, denotes the restraining potential energy associated with orthogonal restraints (eq 24) and the second term in the exponential refers to the restraint along the reaction coordinate (*i.e.*, here the *y*-direction). With eq 24, the volume integration in eq 26 evaluates to

$$\widetilde{V} = \left( \frac{\pi}{\frac{1}{2} k\beta} \right)^{3/2} \tag{27}$$



https://doi.org/10.1021/acs.jctc.1c01194
J. Chem. Theory Comput. 2022, 18, 3873−3893

For the DD method, the raw free energy $\Delta G_{\mathrm{van},i}$ of vanishing intermolecular interactions between the benzene molecule and the environment in the bound ($i$ = b) or unbound ($i$ = u) state

$$\Delta G_{\mathrm{van},i} = \int_0^1 \mathrm{d}\lambda \left\langle \frac{\partial H(\lambda')}{\partial \lambda'} \right\rangle_\lambda \tag{28}$$

was calculated numerically with the trapezoidal rule. Errors on these quantities were calculated from integrating errors on the ensemble averages, the latter obtained by block averaging.[80] The standard binding free energy is

$$\Delta G^o_{\mathrm{bnd}} = \Delta G_{\mathrm{van,u}} - \Delta G_{\mathrm{van,b}} + \Delta G_{\mathrm{res,b}} - \Delta G^o_{\mathrm{res,u}} \tag{29}$$

$\Delta G_{\mathrm{res,b}}$ is the free energy of releasing the restraint on the ligand position in the binding pocket in the simulations of the bound state. It is calculated with the Zwanzig perturbation equation:[75]

$$\Delta G_{\mathrm{res,b}} = +\beta^{-1}\ln\langle e^{-\beta V_{\mathrm{res}}(\mathbf{r}^{3N})}\rangle_0 \tag{30}$$

A 10 ns simulation of the protein with the unrestrained ligand in the binding pocket (indicated by the subscript "0" of the angular brackets) and with the helix 4−helix 7 interface restraint and with coordinates written to file every 3 ps was used to calculate the exponential ensemble average of the restraining potential $V_{\mathrm{res}}(\mathbf{r}^{3N})$,

$$V_{\mathrm{res}}(\mathbf{r}^{3N}) = \frac{1}{2}k(r - r_o)^2 \tag{31}$$

with $k$ = 300 kJ·mol$^{-1}$ · nm$^{-2}$, $r_o$ = 0.60 or 0.14 nm, and the definition of $r_o$ as described above (Section 3.3). Note that the + sign in eq 30 is due to the definition of $\Delta G_{\mathrm{res,b}}$ as the free energy of releasing rather than installing the restraint. $\Delta G^o_{\mathrm{res,u}}$ is the free energy of releasing the equivalent restraint on the solvated ligand and can, by virtue of the isotropic environment in the solvated state, be calculated analytically as[43,78,79]

$$\Delta G^o_{\mathrm{res,u}} = -\beta^{-1}\ln\frac{V_o}{\widetilde{V}} \tag{32}$$

where $V_o$ was defined above and[79]

$$\widetilde{V} = 4\pi\int_0^\infty \mathrm{d}r\, r^2 e^{-\beta V_{\mathrm{res}}(\mathbf{r}^{3N})} \tag{33}$$

The Jarzynski identity[58,59]

$$\Delta G_{\mathrm{bnd,raw}} = +k_{\mathrm{B}}T\ln\overline{e^{-\beta W_{\mathrm{tot}}}} \tag{34}$$

where the total nonequilibrium work $W_{\mathrm{tot}}$ along the pulling pathway between the binding pocket and the solvated regime,

$$W_{\mathrm{tot}} = W_{\mathrm{disres}} + W_{\mathrm{sft}} \tag{35}$$

a composite of work contributions due to the distance restraint ($W_{\mathrm{disres}}$) and the softness modification ($W_{\mathrm{sft}}$), was used to calculate a raw binding free energy $\Delta G_{\mathrm{bnd,\,raw}}$. Each of the two contributions $W_{\mathrm{disres}}$ and $W_{\mathrm{sft}}$ can be calculated according to eq 17, where the $\lambda$-dependent Hamiltonian contributions refer to either the $\lambda$-dependent distance-restraining potential (eq 15) for the calculation of $W_{\mathrm{disres}}$ or the $\lambda$-dependent soft-core van der Waals and electrostatic interactions (eq 8) for the calculation of $W_{\mathrm{sft}}$. The raw binding free energy $\Delta G_{\mathrm{bnd,raw}}$ was converted to a standard binding free energy through

$$\Delta G^o_{\mathrm{bnd}} = \Delta G_{\mathrm{bnd,raw}} + \Delta G_{\mathrm{cor,b}} - \Delta G^o_{\mathrm{res,u}} \tag{36}$$

where $\Delta G_{\mathrm{cor,b}}$ corrects for the initial-coordinate ensemble not entirely representing the volume available to the bound ligand in an equilibrated ensemble,

$$\Delta G_{\mathrm{cor,b}} = -\beta^{-1}\ln\frac{V_{\mathrm{eq}}}{V_{\mathrm{ini}}} \tag{37}$$

where $V_{\mathrm{eq}}$ is the cubic volume sampled by the bound ligand in an equilibrated ensemble, calculated as a product of sampled ranges along the $x$-, $y$-, and $z$-directions, and $V_{\mathrm{ini}}$ is the equivalent cubic volume represented by the initial-coordinate ensemble for the 70 nonequilibrium pulling simulations. $\Delta G^o_{\mathrm{res,u}}$ corrects for the presence of (orthogonal) $x$- and $z$-direction restraints and the $y$-direction pulling restraint in the unbound state. It is calculated as in eq 32,

$$\Delta G^o_{\mathrm{res,u}} = -\beta^{-1}\ln\frac{V_o}{\widetilde{V}} \tag{38}$$

where $V_o$ was defined above and

$$\widetilde{V} = \left(\frac{\pi}{\frac{1}{2}k\beta}\right)^{3/2} \tag{39}$$

akin to the definition of $\widetilde{V}$ for the unbound state in the PMF calculations (eq 27). For the nonequilibrium pulling simulations, the force constant was 3000 kJ·mol$^{-1}$ · nm$^{-2}$.

Uncorrelation of the obtained values $W_{\mathrm{sft}}$ and $W_{\mathrm{disres}}$ per pulling simulation was tested by the Pearson correlation coefficient $R$ and the Kendall rank coefficient $\tau$ considering the total of $N_w$ = 70 work pairs. Uncorrelation allows a combination of all possible $N_w$ values $W_{\mathrm{sft}}$ and $N_w$ values $W_{\mathrm{disres}}$ to get a total of $N_w^2$ values $W_{\mathrm{tot}}$ to be used in eq 34. This approach is referred to as statistical boosting.[81]

A statistical error on $\Delta G^o_{\mathrm{bnd}}$ evaluated from the nonequilibrium pulling simulations with the Jarzynski identity (eq 34) is reported as the standard deviation of 100 free energy values obtained from drawing 100 random samples, with each sample consisting of $N_e$ work values, from the total number of work values. $N_e$ was chosen as the value, where the standard deviation approximately reaches a plateau, $N_e$ = 45 (Figure S4). An alternative to eq 34 for the analysis of nonequilibrium pulling simulations is to use a cumulant expansion,[82] which expresses the free energy $\Delta G$ in terms of the mean ($\overline{W}$) and the variance ($\delta W^2$) of the sampled work values,

$$\Delta G = \overline{W} - \frac{1}{2}\beta\delta W^2. \tag{40}$$

However, this approach was not successful in the present case, which is likely due to non-Gaussianity of the underlying work distributions (Table S1). Failure of this approach has been observed before for biomolecular ligand unbinding processes,[32] and it has been suggested that it is only meaningful in the near-equilibrium regime and in the presence of sufficient data to accurately estimate the variance.[60]

Backbone root-mean-square deviation (rmsd) and secondary structure (dssp[83]) analyses[84] were performed to assess the stability of T4L helices 4 and 7. Prior to the rmsd calculation, a rototranslational fit of all backbone atoms to the equilibrated structure was carried out (although this may actually not be required due to the protein rototranslational constraint used in the simulations).

*3.4.2. MDM System.* In each window of the US simulations of system MDM, the distance coordinate $\xi$ was written to file



https://doi.org/10.1021/acs.jctc.1c01194
J. Chem. Theory Comput. 2022, 18, 3873−3893

every 3 ps, and WHAM[74] was used to calculate the free-energy profile $\Delta G(\xi)$ along the distance coordinate. Errors on the PMF were calculated from bootstrapping as implemented in the WHAM package of Grossfield.[74] The PMFs were anchored to zero in the bound minimum. For the MDM system, two PMFs were computed. One of the PMF calculations (labeled "MDM-MOD") featured simultaneous alchemical modifications, whereas the second PMF calculation (labeled "MDM-0") was done without any alchemical modifications to serve as comparison. A raw binding free energy $\Delta G_{\mathrm{bnd,\,raw}}$ was obtained in the MDM-MOD case as

$$\Delta G_{\mathrm{bnd,raw}} = -\Delta G(\xi_{\mathrm{u}}) - \Delta G_{\mathrm{al,pmf}} - \Delta G_{\mathrm{al,prior}} \tag{41}$$

where $\xi_{\mathrm{u}}$ is the distance in the unbound regime exempt from alchemical modification,

$$\Delta G_{\mathrm{al,pmf}} = \int_0^1 \mathrm{d}\lambda \left\langle \frac{\partial H(\lambda')}{\partial \lambda'} \right\rangle_\lambda \tag{42}$$

is the reversible work of alchemical modifications along the unbinding path, and

$$\Delta G_{\mathrm{al,prior}} = \int_0^1 \mathrm{d}\lambda \left\langle \frac{\partial H(\lambda')}{\partial \lambda'} \right\rangle_\lambda \tag{43}$$

is the reversible work of alchemical modifications within the binding pocket prior to the PMF calculation. Angular brackets denote ensemble averaging. For comparison, the raw binding free energy $\Delta G_{\mathrm{bnd,raw}}$ in the MDM-0 case is

$$\Delta G_{\mathrm{bnd,raw}} = -\Delta G(\xi_{\mathrm{u}}) \tag{44}$$

*i.e.,* only involves the reversible work of pulling the peptide away from the protein. Note that $\Delta G_{\mathrm{bnd,raw}}$ in both eqs 41 and 44 can be directly compared but do not refer to standard absolute peptide–protein binding free energies comparable to the experiment since they are affected by conformational helix restraints and directional restraints during the pulling. The free energy associated with the restraints can in principle be evaluated from free-energy perturbation[75] or, possibly more suitable in the case of conformational restraints, a PMF in terms of distance root-mean-square deviation[85] as done in previous work.[22]

## 4. RESULTS AND DISCUSSION

### 4.1. Alchemically Modified PMF Calculations.

*4.1.1. T4L System: Modifying the Softness of Protein–Ligand Interactions within a S-Approach.* Considering the occluded nature of its binding pocket, the T4L system was used to illustrate the power of modifying the softness level of ligand–protein interactions during a PMF calculation (Section 2.1.1). This corresponds to the S-approach (Figure 1a), applied along a series of US simulations. The Λ-function used for the PMF calculations of system T4L has a maximum at $\lambda$ = 0.25 (Figure 2d) and vanishes at the end points $\lambda$ = 0 and $\lambda$ = 1. This means that the bound and unbound states correspond to physical ligand–protein interactions, while intermediate states along the reaction coordinate of the PMF calculation feature softened ligand–protein interactions. The softening of interactions is especially important near the interface of lysozyme helices 4 and 7 (Figure 2c). In the vicinity of this region, the Λ-function has its maximum. Thus, prior physical insight into the system, *e.g.,* by visual inspection, is required to determine the shape of this function. In an alternative simulation approach without application softness, one has to explicitly sample the opening and reclosing of this interface or other binding channels during PMF calculations to calculate a binding free energy.[21] Long brute-force simulations[73] or simulations with enhanced sampling[49,65] have also been performed to study benzene dissociation pathways. These approaches are computationally expensive and require exhaustive sampling of conformational changes in the surroundings of the binding site. In the present work, a distance restraint on the helix 4–helix 7 interface was used to keep the binding pocket closed and hence to avoid conformational sampling issues associated with conformational differences between the open and closed state[46] (Section 3.2).

For $\lambda$-values around 0.16–0.32 (distances around 0.91–1.23 nm), the physical (recalculated) protein–benzene van der Waals interaction energy would start to diverge toward very high positive values due to steric clashes (Figure 4). This divergence is removed by softening the interactions such that the protein–benzene van der Waals interaction energies even remain negative while the benzene molecule leaves the buried binding pocket (Figure 4). Due to the nonpolar nature of the ligand, the effect is less pronounced for electrostatic interactions, where the physical (recalculated) values do not exceed 20 kJ·mol$^{-1}$ and the softened interaction energies are almost vanishing. In total, the protein–benzene interaction energy remains negative and in the moderate two-digit range (−65 to −20 kJ·mol$^{-1}$) until the benzene molecule has left the vicinity of the protein (Figure 4). Slight increases in the softened interaction energy are observed until about a distance of 1.2 nm (van der Waals and total interaction energy) or 0.8 nm (electrostatic interaction energy). The corresponding free-energy profile reflects these trends in terms of a steep initial incline before a first plateau is reached. In this plateau region

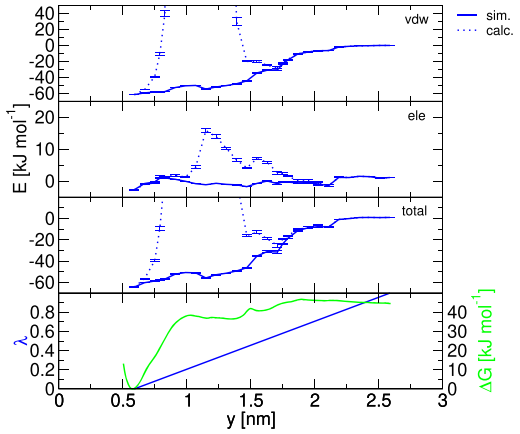



**Figure 4.** Protein–benzene interaction energies in the T4L system along the reaction coordinate $y$ of the PMF calculation. The van der Waals ("vdw") and electrostatic ("ele") components as well as the total ("total") interaction energies are shown. The solid lines depict the interaction energies as occurring during the simulation ("sim."), *i.e.,* with the softened interaction potentials, while the dotted lines depict interaction energies recalculated with the physical force-field ("calc"), *i.e.,* with nonsoftened interactions. Error bars refer to statistical errors from block averaging.[80] In the bottom panel, the variation of the $\lambda$-parameter for the simultaneous alchemical modification and the free-energy profile (see also Figure 5a) are shown.



https://doi.org/10.1021/acs.jctc.1c01194
J. Chem. Theory Comput. 2022, 18, 3873–3893

(ca. 1.1−1.4 nm), unfavorable steric clashes, as evidenced by the recalculated interaction energies, are slowly decreasing again, while the actual van der Waals interaction energies are, after a minimum at 1.16 nm (corresponding to the regime close to maximum softness), again increasing. Beyond a distance of 1.7 nm, there is virtually no effect of softened protein−benzene interactions because due to the large separation, the benzene molecule is interacting very little with the protein and is almost in the unbound state.

The free-energy profile of benzene−protein unbinding converges rather quickly (40−60 ns; Figure 5a). It is artificial because it corresponds to an alchemically modified ligand in the intermediate states; *i.e.*, its slopes and plateaus reflect a combination of physical interaction phenomena in the protein (steric clashes with the helix−helix interface around 1.1−1.3 nm) and the nature of the Λ-function that determines the modification of these interactions. Note that reweighting the free-energy profile with the Zwanzig perturbation equation,[75] as done in previous work in the context of mutations,[22] to the physical (nonsoftened) interaction-energy function is not possible here because the extremely high recalculated van der Waals interaction energies along the pulling pathway (Figure 4) prohibit a meaningful use of perturbation theory; *i.e.*, the perturbation from the physical to the soft ligand is too large in the vicinity of the closed helix−helix interface.

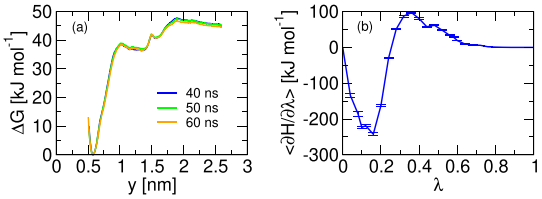



**Figure 5.** PMF and TI curves to calculate the binding free energy for the T4L system. (a) PMF along the benzene−protein distance, anchored to zero in the bound state. The depicted curves do not include the free energy of softness modification. Error bars on the PMF values are not shown for clarity. They are at most 0.6 kJ·mol$^{-1}$ per point and were evaluated from bootstrapping.[74] (b) TI curve for softness modification. Error bars refer to statistical errors from block averaging.[80]

A binding free energy can be calculated when the free energy associated with the alchemical modification along the unbinding process, $\Delta G_{\mathrm{sft}}$, is added to the free energy derived from the PMF well depth (eq 20). This was done with TI over ensemble averages $\left\langle \frac{\partial H}{\partial \lambda} \right\rangle$ which gives the free energy of installing and deinstalling soft-core benzene−protein interactions ($\Delta G_{\mathrm{sft}}$; eq 9 and Figure 5b). It evaluates to $\Delta G_{\mathrm{sft}} = -16.9$ kJ·mol$^{-1}$ and when combined with the well depth of the PMF (−44.5 kJ·mol$^{-1}$; first term in eq 20) gives $\Delta G_{\mathrm{bnd,raw}} = -27.6$ kJ·mol$^{-1}$ (eq 20 and Table 1). Adding corrections for orthogonal restraining, a standard binding free energy of $\Delta G^{o}_{\mathrm{bnd}} = -11.6$ kJ·mol$^{-1}$ ensues. The experimental value for the benzene-T4 lysozyme L99A standard binding free energy is −21.7 kJ·mol$^{-1}$.[86] It cannot be compared to the calculated values because we do not include the free energy associated with releasing the restraint on the helix 4−helix 7 interface. The magnitude of $\Delta G_{\mathrm{sft}}$ depends on the extent of softening used in the simulation, *i.e.*, on the magnitude of $\alpha$ (eqs 4 and 7) and the shape of the employed Λ-function, as well as on the

**Table 1. Quantities Involved in the Calculation of the Standard Binding Free Energy $\Delta G^{o}_{\mathrm{bnd}}$ (eqs 22, 29, or 36) for the T4L System[i]**

| | | | | | |
|---|---|---|---|---|---|
| PMF | $\Delta G_{\mathrm{bnd,raw}}$ [kJ·mol$^{-1}$] | $\Delta G_{\mathrm{ores,b}}$ [kJ·mol$^{-1}$] | $\Delta G^{o}_{\mathrm{ores,u}}$ [kJ·mol$^{-1}$] | | $\Delta G^{o}_{\mathrm{bnd}}$ [kJ·mol$^{-1}$] |
| | −27.6 ± 1.6[a] | −2.4 | −18.4 | | −11.6 ± 1.6 |
| DD[d] | $\Delta G_{\mathrm{van,u}}$ [kJ·mol$^{-1}$] | $\Delta G_{\mathrm{van,b}}$ [kJ·mol$^{-1}$] | $\Delta G_{\mathrm{res,b}}$ [kJ·mol$^{-1}$] | $\Delta G^{o}_{\mathrm{res,u}}$ [kJ·mol$^{-1}$] | $\Delta G^{o}_{\mathrm{bnd}}$ [kJ·mol$^{-1}$] |
| | 5.4 ± 0.5 | 16.5 ± 1.9 | −0.1 | −1.4 | −9.8 ± 2.4 |
| DD[e] | $\Delta G_{\mathrm{van,u}}$ [kJ·mol$^{-1}$] | $\Delta G_{\mathrm{van,b}}$ [kJ·mol$^{-1}$] | $\Delta G_{\mathrm{res,b}}$ [kJ·mol$^{-1}$] | $\Delta G^{o}_{\mathrm{res,u}}$ [kJ·mol$^{-1}$] | $\Delta G^{o}_{\mathrm{bnd}}$ [kJ·mol$^{-1}$] |
| | 5.4 ± 0.5 | 21.9 ± 1.9 | −0.7 | −8.7 | −8.5 ± 2.4 |
| Jarzynski[f] | $\Delta G_{\mathrm{bnd,raw}}$ [kJ·mol$^{-1}$] | $\Delta G_{\mathrm{cor,b}}$ [kJ·mol$^{-1}$] | $\Delta G^{o}_{\mathrm{res,u}}$ [kJ·mol$^{-1}$] | | $\Delta G^{o}_{\mathrm{bnd}}$ [kJ·mol$^{-1}$] |
| | −30.1 ± 4.2[b] | −1.1 | −21.0 | | −10.2 ± 4.2 |
| Jarzynski[g] | $\Delta G_{\mathrm{bnd,raw}}$ [kJ·mol$^{-1}$] | $\Delta G_{\mathrm{cor,b}}$ [kJ·mol$^{-1}$] | $\Delta G^{o}_{\mathrm{res,u}}$ [kJ·mol$^{-1}$] | | $\Delta G^{o}_{\mathrm{bnd}}$ [kJ·mol$^{-1}$] |
| | −24.6 ± 4.5[c] | −1.1 | −21.0 | | −4.7 ± 4.5 |
| Jarzynski[h] | $\Delta G_{\mathrm{bnd,raw}}$ [kJ·mol$^{-1}$] | $\Delta G_{\mathrm{cor,b}}$ [kJ·mol$^{-1}$] | $\Delta G^{o}_{\mathrm{res,u}}$ [kJ·mol$^{-1}$] | | $\Delta G^{o}_{\mathrm{bnd}}$ [kJ·mol$^{-1}$] |
| | −51.0 ± 4.5 | −1.1 | −21.0 | | −31.1 ± 4.5 |

[a]The contribution from the PMF (Figure 5a; first term in eq 20) is $\Delta G(\xi_{\mathrm{u}}) = 44.5 \pm 0.2$ kJ·mol$^{-1}$ (error from adding bootstrapping errors in the bound and unbound state) and the contribution $\Delta G_{\mathrm{sft}}$ (Figure 5b and eq 9) is $\Delta G_{\mathrm{sft}} = -16.9 \pm 1.4$ kJ·mol$^{-1}$ (error from integrating block-averaged errors). [b]With statistical boosting, a value $\Delta G_{\mathrm{bnd,raw}} = -27.0$ kJ·mol$^{-1}$ is obtained, which results in $\Delta G^{o}_{\mathrm{bnd}} = -7.1$ kJ·mol$^{-1}$. [c]With statistical boosting, a value $\Delta G_{\mathrm{bnd,raw}} = -22.6$ kJ·mol$^{-1}$ is obtained, which results in $\Delta G^{o}_{\mathrm{bnd}} = -2.7$ kJ·mol$^{-1}$. [d]Restraining scheme with a distance of 0.60 nm (Section 3.3). [e]Restraining scheme with a distance of 0.14 nm and separate TI calculations for the electrostatic (2.1 ± 0.5 kJ·mol$^{-1}$) and van der Waals (19.8 ± 1.4 kJ·mol$^{-1}$) free-energy contribution (Section 3.3). [f]With softness modification and pulling rate 0.2 m·s$^{-1}$. [g]With softness modification and pulling rate 0.4 m·s$^{-1}$. [h]No softness modification and pulling rate 0.2 m·s$^{-1}$. [i]For the PMF calculation, the raw binding free energy $\Delta G_{\mathrm{bnd,\ raw}}$ (eq 20) and the free energies associated with the release of orthogonal restraints in the bound state and the associated standard-state correction in the unbound state, $\Delta G_{\mathrm{ores,b}}$ and $\Delta G^{o}_{\mathrm{ores,u}}$ (eqs 23 and 25), respectively, are reported. The raw binding free energy contains a distance-US contribution corresponding to the free energy captured by the PMF (first term in eq 20), as well as a free-energy contribution due to the softness modification ($\Delta G_{\mathrm{sft}}$; eq 9). For the DD calculation, the free energies of decoupling ligand−environment intermolecular interactions in the unbound and bound state, $\Delta G_{\mathrm{van,u}}$ and $\Delta G_{\mathrm{van,b}}$ (eq 28), respectively, and the free energies associated with the release of harmonic distance restraints in the bound and unbound state, $\Delta G_{\mathrm{res,b}}$ and $\Delta G^{o}_{\mathrm{res,u}}$ (eqs 30 and 32), respectively, are reported. For the nonequilibrium pulling simulations with the Jarzynski identity, the raw binding free energy $\Delta G_{\mathrm{bnd,raw}}$ (eq 34) and the free energies associated with limited volume sampling in the bound state and the associated standard-state correction in the unbound state, $\Delta G_{\mathrm{cor,b}}$ and $\Delta G^{o}_{\mathrm{res,u}}$ (eqs 37 and 38), respectively, are reported. For comparison, results of the Jarzynski identity applied to nonequilibrium pulling simulations exempt of simultaneous softness modification are reported as well.



https://doi.org/10.1021/acs.jctc.1c01194
J. Chem. Theory Comput. 2022, 18, 3873−3893

extent of steric hindrance encountered by the ligand during the softening process. Figure 5b shows an initially negative $\left\langle \frac{\partial H}{\partial \lambda} \right\rangle$, decreasing up to $\lambda = 0.16$ and then increasing again, passing through the zero at $\lambda = 0.25$. This also corresponds to a turning point in $\left\langle \frac{\partial H}{\partial \lambda} \right\rangle$ and the maximum in the Λ-function (Figure 2d). In this initial range up to $\lambda = 0.25$, softness increases and it is favorable to install soft-core interactions because thereby steric benzene−protein clashes are removed. Physical benzene−protein interactions are restored for $\lambda > 0.25$, and in this regime positive, *i.e.*, unfavorable, free-energy contributions of the softness modification occur. Nevertheless, overall, it is favorable to carry out softening and desoftening of benzene−protein interactions ($\Delta G_{\mathrm{sft}} < 0$). Despite the usefulness of the softening approach in the T4L system, there is an optimum level of softness to be used beyond which application of the approach becomes impractical because a too high magnitude of $\Delta G_{\mathrm{sft}}$ may be very difficult to determine by numerical integration and require additional points to be simulated to achieve accurate evaluation by numerical integration or other approaches like FEP[75] or MBAR.[87]

Note that the TI approach used to calculate $\Delta G_{\mathrm{sft}}$ may be similarly applied to calculate the free-energy contribution associated with the distance restraint (Figure S1) as an alternative to the WHAM procedure for PMF analysis (Section 3.4). Interpreting the target distances of the successive umbrella windows as $\lambda$-dependent distances $\xi_o(\lambda)$, one may formulate a $\left\langle \frac{\partial H}{\partial \lambda} \right\rangle$ corresponding to the perturbed distance-restraining potential $H(\lambda) = \frac{1}{2}k(\xi - \xi_o(\lambda))^2$. Integration of the TI curve yields a free energy of 43.8 kJ·mol$^{-1}$, in very good agreement with the well depth of the PMF (44.5 kJ·mol$^{-1}$; Figure 5 and Figure S1). The running integral of the TI curve is virtually identical with the PMF from the WHAM procedure (Figure S1). Addition of the free-energy derivatives corresponding to the distance-restraint and softness contributions yields a TI curve whose integral similarly gives a quantity corresponding to $\Delta G_{\mathrm{bnd,raw}}$, −26.9 kJ·mol$^{-1}$, very close to the value $\Delta G_{\mathrm{bnd,\ raw}} = -27.6$ kJ·mol$^{-1}$ reported in Table 1.

The standard computational approach to determine the absolute binding free energy of a small ligand like benzene to a buried binding pocket is the DD method, which was also applied here for comparison to the PMF approach with simultaneous alchemical modification. It gives a standard absolute binding free energy of $\Delta G^{o}_{\mathrm{bnd}} = -9.8$ or −8.5 kJ·mol$^{-1}$, depending on which restraining and TI scheme was used (Section 3.3 and Table 1). The values are comparable to and have similar statistical precision as the PMF approach (Table 1). TI curves are shown in Figure S2.

It is unclear why the two binding free energies obtained with the DD method are of slightly lower magnitude (by 16 and 27%, respectively) than the PMF result. The average van der Waals interaction energies between the benzene molecule and the protein as well as the benzene molecule and the solvent (water and ions) in the bound state differ by only 1.5 or 0.9 kJ·mol$^{-1}$ between the two DD simulations and the PMF calculation (Table S2). Due to the correspondingly lower magnitudes, the differences are even smaller for electrostatic interaction energies and interaction energies with the solvent (Table S2). We investigated structural characteristics of helices 4 and 7 in the bound and unbound state of the PMF and DD calculations. No significant differences in terms of backbone rmsd (Figure S7) or secondary structure content (Table S3) were found except for the backbone rmsd of helix 7, which was slightly higher in the DD simulations of the bound state than in the US simulation of the bound state for the PMF calculation (0.32 and 0.38 nm *vs* 0.25 nm, respectively; Figure S7).

In terms of ease of use, the PMF and DD approaches are comparable: On the one hand, for the alchemically modified PMF approach, physical insight is required to define the Λ-function. In case of a conventional PMF calculation exempt from simultaneous alchemical modification, the effort of physical insight would shift to the definition of an appropriate reaction coordinate. For the DD approach, on the other hand, physical insight is required to define one, or multiple appropriate restraints to keep the ligand in the binding pocket for the in-protein decoupling simulation. The computational effort spent in the two approaches is system-size-dependent. The DD approach has the advantage of treating the unbound state in a much smaller computational box, which becomes especially beneficial for large receptors and comparably small ligands, as well as treating the bound state in a smaller computational box, which becomes beneficial for large binding partners with long-range interactions. For the T4L system, in the present context, 28 windows were used for the PMF calculation and convergence was reached after ca. 40−50 ns (Figure 5a). For the DD simulations in the protein and solvent, 21 or 27 $\lambda$ states were simulated for 10 or 14 ns, respectively. In the present situation, the computational effort was hence somewhat lower in the DD simulations. However, when larger or flexible ligand molecules are studied, DD simulations may become more compute-intensive or unfeasible in comparison to a physical-pathway method.

To conclude, we point out the important benefit of introducing alchemical modifications into a physical-pathway method. For the case illustrated here, a buried binding pocket, carrying out PMF calculations would be very difficult in a situation where ligand−protein interactions are unmodified throughout because steric ligand−protein clashes would lead to high forces and entail protein conformational changes to allow ligand dissociation, which may in turn lead to convergence problems or lead to a simulation crash.

However, introducing soft-core interactions along the ligand−protein unbinding pathway within the S-approach (Figure 1) transforms the problem into a series of US simulations with uncomplicated convergence behavior.

*4.1.2. MDM System: Modifying the Strength of Protein−Ligand Interactions within a PPS-Approach.* Besides steric clashes, sticky interactions (*e.g.*, strong salt bridges, hydrogen bonds, or hydrophobic interactions) can render conventional PMF calculations difficult to carry out because barriers have to be overcome that may be difficult to sample in a converged manner and that require tight umbrella window spacing to achieve a sufficient histogram overlap. While the T4L system was used to illustrate the benefits of combining alchemical modifications and physical-pathway methods in the former case (S-approach; Figure 1a), the MDM system, featuring a peptide ligand with two hydrophobic side chains protruding into a deep binding groove with hydrophobic contact sites, is used for investigations concerning the latter case. The problems posed by sticky interactions are best addressed by reducing the strength of these interactions. Since these interactions are favorable, softening, by virtue of removing a singularity, may show some benefits in the case of strong attractive electrostatic interactions at a short distance.



https://doi.org/10.1021/acs.jctc.1c01194
J. Chem. Theory Comput. 2022, 18, 3873−3893

However, here, the dominant interactions are nonpolar, which is why, primarily, the depth of the Lennard−Jones potential between the protein and the concerned hydrophobic side chains was modified (Section 2.1.2). Electrostatic protein−peptide side chain interactions were reduced as well, although the effect may not be so important here. Hence, here the PPS-approach (Figure 1b) was used; *i.e.*, prior to the US simulations, crucial interactions were reduced. This step is akin to the AlchemPMF method of Cruz *et al.*[29] (Section 1). Finally, along the US simulations, the physical magnitude of these interactions was gradually restored; *i.e.*, in contrast to the AlchemPMF method, here the second alchemical modification is done during the US rather than afterward in the unbound state. Again, a Λ-function needed to be chosen to guide the modification of Lennard−Jones and electrostatic interactions during the PMF calculation. Since the maximum stickiness of interactions is expected to occur right within the bound state, the Λ-function should peak at $\lambda = 0$ and gradually decrease toward 0 for the unbound state at $\lambda = 1$. This means that the unbinding process is now accompanied by removing an alchemical modification that has already been established within the bound state. Thus, here two TI calculations are required, namely, a first one that reduces protein−peptide interactions in the bound state and a second one that restores the physical interactions along the unbinding pathway. To challenge the approach, here, a nontrivial reaction coordinate along the *z*-axis was chosen (Figure 3a,b). This pathway passes the surface of the protein and hence may lead to multiple further peptide−protein interactions in addition to the strong interactions in the binding pocket. In comparison, a reaction coordinate along the *y*-axis would have implied mere deattachment of the peptide from the binding site and straight pulling into the solvent regime.

Figure 6 shows the evolution of peptide−protein interaction energies along the unbinding pathway. Of particular note is the difference in the simulated and recalculated peptide−protein van der Waals interaction energy below $z = 0.56$ nm or $\lambda = 0.1$, where the latter is more favorable than the former by 55 kJ·mol$^{-1}$, illustrating the reduction of peptide−protein interactions that was introduced prior to the PMF calculation. A similar albeit somewhat smaller effect is also seen at the intermediate PMF plateau region at $z = 0.98$ nm, where the reduction in Lennard−Jones interactions is still around 30 kJ·mol$^{-1}$. Electrostatic protein−peptide interactions are of almost fourfold magnitude compared to Lennard−Jones interactions, mainly due to the charged groups in the peptide. However, since here only the electrostatic interactions of the F2 and W6 side chains were modified, the difference between the simulated and recalculated electrostatic (and total) interaction energies is rather small in comparison to the magnitude of these energies.

By virtue of conformational restraints on the peptide, the free-energy profile of peptide−protein unbinding converges rather quickly, within 30−40 ns, for both the normal (MDM-0) and the modified PMF (MDM-MOD; Figure 7a). The latter free-energy profile is artificial because it corresponds to an alchemically modified peptide in the bound and intermediate states; *i.e.*, its slopes and plateaus reflect a combination of interactions between the modified peptide and the protein and the nature of the Λ-function that determines the reinstallation of physical peptide−protein interactions. The two free-energy profiles differ from about 0.5 nm on. Here, the PMF for MDM-MOD presents a first plateau region at 10 kJ·mol$^{-1}$, meaning that peptide−protein interactions are sufficiently weak to pull ca. 0.2 nm further along the protein without a concomitant increase in free energy. As evident from Figure 3d, the corresponding λ-range ($\lambda < 0.2$ nm) corresponds to $\Lambda \approx 1$, *i.e.*, minimal interactions between the peptide F2 and W6 side chains with the protein. In contrast, the PMF for MDM-0 rises monotonically until the solvated plateau is reached at 1.9 nm. In particular, it rises extremely sharply until a shoulder at 0.65 nm. The sharp rise is due to removing favorable interactions between the peptide F2 and W6 side chains and the protein. Between 0.9 and 1.15 nm, the PMF for MDM-MOD shows a second plateau at 30 kJ·mol$^{-1}$. This falls mostly into the region of peaking and then decreasing peptide−protein interactions between 1.0 and 1.3 nm, already noted above (Figure 6). The TI profile for reinstallation of interactions has a maximum at 0.98 nm ($\lambda =$

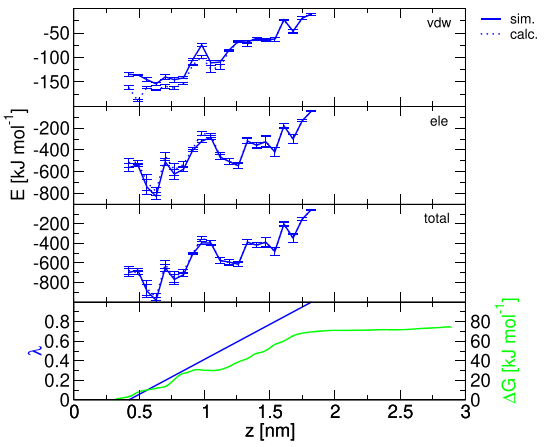



**Figure 6.** Protein−peptide interaction energies in the MDM system along the reaction coordinate *z* of the PMF calculation. The van der Waals ("vdw") and electrostatic ("ele") components as well as the total ("total") interaction energies are shown. The solid lines depict the interaction energies as occurring during the simulation ("sim."), *i.e.*, with the reduced interaction potentials, while the dotted lines depict interaction energies recalculated with the physical force field ("calc."), *i.e.*, with full interactions. Error bars refer to statistical errors from block averaging.[80] In the bottom panel, the variation of the λ-parameter for the simultaneous alchemical modification and the free-energy profile (see also Figure 7a) are shown. Note that the alchemical modification process is already completed at the onset of the unbound plateau in the PMF.

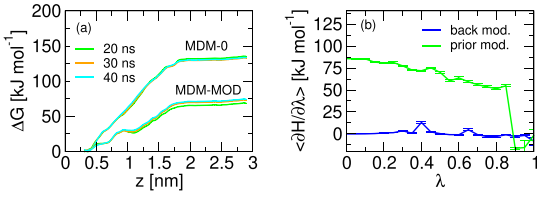



**Figure 7.** PMF and TI curves to calculate the binding free energy for the MDM system. (a) PMF along the peptide−protein distance, anchored to zero in the bound state, for an approach without ("MDM-0") or with ("MDM-MOD") simultaneous alchemical modifications. In the latter case, the depicted curves do not include the free energy of alchemical modifications. Error bars on the PMF values are not shown for clarity. They are at most 0.4 kJ·mol$^{-1}$ per point and were evaluated from bootstrappping.[74] (b) TI curve for introducing ("prior mod.") and reverting ("back mod.") alchemical modifications. Error bars refer to statistical errors from block averaging.[80]



https://doi.org/10.1021/acs.jctc.1c01194
J. Chem. Theory Comput. 2022, 18, 3873−3893

0.4; Figure 7b), meaning that it is unfavorable to establish full peptide F2 and W6 side chain charges and van der Waals interactions here, which is in agreement with the interaction energy profile peaking at negative values of comparably low magnitude here (−400 kJ·mol$^{-1}$; Figure 6). Another minute peak in the TI profile (Figure 7b) occurs at $\lambda$ = 0.65 and coincides with a plateau in peptide−protein interaction energies (−400 kJ·mol$^{-1}$; Figure 6). In total, the integral of the TI curve almost vanishes ($\Delta G_{al,pmf}$ = 0.5 ± 0.6 kJ·mol$^{-1}$; eq 42); *i.e.*, the difference between the plateau values of the PMFs for MDM-0 (132.5 kJ·mol$^{-1}$) and MDM-MOD (71.5 kJ·mol$^{-1}$; Figure 7) is essentially entirely captured by the prior removal of interactions in the bound state ($\Delta G_{al,prior}$ = 59.6 ± 1.3 kJ·mol$^{-1}$; eq 43). Along with PMF error estimates from bootstrapping, thus, $\Delta G_{bnd,raw}$ evaluates to −132.5 ± 0.4 kJ·mol$^{-1}$ for MDM-0 and −131.6 ± 2.3 kJ·mol$^{-1}$ for MDM-MOD. The statistical error is somewhat higher for the MDM-MOD procedure because errors pertaining to three separate free-energy calculations are added. Orthogonal and conformational-restraint corrections are not applied here because the aim of the MDM-MOD and MDM-0 simulations was to merely draw a comparison between the unmodified and the modified (PPS-approach; Figure 1) PMF procedures in terms of a "raw" binding free energy estimate.

Note that one may again integrate the TI curve pertaining to the distance restraint term used in the US, which yields a free energy of 70.2 ± 8.3 kJ·mol$^{-1}$ for MDM-MOD or 132.0 ± 11.6 kJ·mol$^{-1}$ for MDM-0 (Figure S1), in very good agreement with the unbound plateau values reported above. The running integral of the TI curve is virtually identical with the PMF from the WHAM procedure (Figure S1).

Finally, while good agreement was achieved between the MDM-0 and MDM-MOD procedures, in the following, the advantages and disadvantages of the two approaches are discussed. For the present example, the MDM-0 approach is superior to the MDM-MOD approach in terms of compute effort because while the PMFs converge on a similar timescale, the MDM-MOD approach requires an additional prior TI free-energy calculation (PPS-approach; Figure 1b). However, in other systems, where the breaking of strong ligand−protein interactions may require protein conformational rearrangement, compute effort in the US simulations of the modified system may be reduced. The problem would then be shifted to the prior TI simulations where it may be handled more easily because rather than breaking interactions by pulling as in a physical-pathway method, the alchemical TI simulation can break these interactions in a sophisticated, user-defined pathway through an alchemical $\lambda$-space along which the ligand does not leave the bound state. Note that, in certain cases, one may conceive of the requirement of additional restraints to reduce ligand positional sampling in the binding pocket upon the weakening of ligand−protein interactions, similar to the case of DD simulations.

For the present example, in addition, the MDM-0 approach is slightly superior to the MDM-MOD approach in terms of the histogram area overlap. The sampled distance histograms are shown in Figure S3, along with the overlap $A_{overlap}$ between two successive distance probability distributions defined as the integral over the intersecting regions of the two probability distributions. The area overlap is overall of comparable size for systems MDM-0 and MDM-MOD except for the region up to 0.55 nm and the region between 0.8 and 1.0 nm, where MDM-0 shows a better area overlap. In the former region near the bound state, MDM-0 has narrower and higher distance probability distributions because these distributions are shifted toward the bound state due to strong protein−peptide interactions. The latter region corresponds to the pre-plateau region in MDM-MOD where protein−peptide interactions become more unfavorable (Figure 6). On the other hand, in system MDM-0, the breaking of protein−peptide interactions near 0.6 nm goes with reduced overlap, whereas here system MDM-MOD shows improved overlap. Thus, overlap issues are shifted to other areas of the pulling pathway by the PPS-approach. This implies that overlap issues can, by the choice of alchemical modification and $\Lambda$-function, be modulated in a user-defined manner. Plain visual inspection of the distance histograms in Figure S3 suggests that the sampling of distances across the different windows is more uniform in MDM-MOD; *i.e.*, the width of the distance probability distributions varies less in system MDM-MOD compared to MDM-0. This is particularly pronounced near the bound state, where system MDM-0 shows very narrow distributions in the first three windows and a rather wide distribution in the fourth window, with respective full widths at half-maximum of 0.06, 0.07, 0.07, and 0.13 nm, whereas MDM-MOD has distributions of full widths at half-maximum of 0.09, 0.08, 0.10, and 0.09 nm, respectively. From a user point of view, uniform distributions are more convenient because convergence in the different windows is then reached at similar simulation times. In contrast, wider distributions generally require more simulation time to converge than narrow distributions. Defining histogram overlap as the width of the intersecting regions, $W_{overlap}$, it can be seen that the histogram overlap along the PMF is more uniform for system MDM-MOD than MDM-0 (Figure S3). For system MDM-0, the standard deviation of $W_{overlap}$ is 0.04 nm, exceeding the one for MDM-0 (0.03 nm) by 33%. Thus, in terms of distance sampling homogeneity, the MDM-MOD approach is superior to the MDM-0 approach.

The experimental value for the RFMDYWEGL peptide−MDM2 standard binding free energy is −36.5 kJ·mol$^{-1}$.[88] It cannot be compared to the calculated values because we do not include the free energy associated with releasing the restraint on the peptide helical conformation and orientational restraints.

**4.2. Alchemically Modified Nonequilibrium Pulling Simulations: "Soft Jarzynski Method".** The power of combining alchemical transformations and physical-pathway methods was illustrated in Sections 4.1.1 and 4.1.2 in the context of PMF calculations based on US. However, the employed physical-pathway method can also be nonequilibrium pulling in combination with the Jarzynski identity as free-energy estimator (Section 3.4). In the present work, this was done for the T4L system; *i.e.*, multiple pulling simulations driving the benzene ligand out of the buried binding pocket of lysozyme were carried out along with simultaneous softening and desoftening of ligand−protein interactions. Thus, the approach may be succinctly referred to as a "soft Jarzynski method". It corresponds to the S-approach (Figure 1) applied in a nonequilibrium fashion. Without soft-core interactions, the problem posed by the steric clashes becomes even more apparent in a nonequilibrium approach than in US because here it is virtually impossible for the protein to perform slow conformational rearrangements to adapt to the physical space required by the dissociating ligand. This explains the power of the present approach in nonequilibrium physical-pathway methods. The structure of the pocket interface is in a low


https://doi.org/10.1021/acs.jctc.1c01194
J. Chem. Theory Comput. 2022, 18, 3873−3893

backbone rmsd regime (≤0.3 nm) throughout the non-equilibrium pulling simulations when soft-core interactions are installed and deinstalled (Figure S7). In the US simulations of the PMF calculation, it is significantly higher, especially for helix 7 (Figure S7), because here more conformational sampling is done in a given reaction-coordinate distance range. The role of conformational sampling is also apparent when comparing the helix 7 backbone rmsd for slow and fast soft nonequilibrium pulling, where, from about 1 nm distance on, it is consistently lower by about 0.05 nm for the latter (Figure S7). Structural perturbations in helix 7 are much more pronounced if no softness is involved in the pulling (see below and Figure S7).

We find standard binding free energies in good agreement with the PMF method ($\Delta G^{o}_{bnd} = -10.2$ kJ·mol$^{-1}$ for a pulling rate of 0.2 m·s$^{-1}$ compared to −11.6 kJ·mol$^{-1}$ from the PMF; Table 1). Note that the error on the Jarzynski estimate is rather high, which is due to the way it was calculated (standard deviation over random subsamples; Section 4 and Figure S4). It does not quantify the uncertainty in the free energy estimator but rather the spread in the results due to different sampling of phase space. Note also that here a faster pulling rate of 0.4 m·s$^{-1}$ gives a binding free energy of lower magnitude (−4.7 kJ·mol$^{-1}$; Table 1). This is surprising at first sight because one may expect higher pulling speeds to result in higher (more positive) sampled nonequilibrium works and thus a higher magnitude of the binding free energy. However, here, the binding free energy (eq 34) is a composite of a distance restraint contribution $W_{disres}$ and a softness-modification contribution $W_{sft}$ (eq 35). The latter is mainly negative. The faster the pulling is, the more steric clashes occur; *i.e.*, the higher is the magnitude of the softness work because it is increasingly favorable to install soft-core interactions. Hence, the overestimation of the magnitude of $W_{sft}$ in the case of the fast nonequilibrium procedure leads to underestimated magnitudes of the total works $W_{total}$ along the pulling pathway. This can be seen from the probability distributions of $W_{disres}$ and $W_{sft}$ in Figure 8 for the different pulling speeds. The probability distribution of $W_{sft}$ has a global maximum at −14.8 kJ·mol$^{-1}$ for fast pulling compared to one at −10.9 kJ·mol$^{-1}$ for slow pulling. Values $W_{sft} < -18$ kJ·mol$^{-1}$ are consistently sampled with higher probabilities with fast pulling compared to slow pulling (inset of Figure 8). Thus, although higher pulling works $W_{disres}$ are encountered with fast pulling, these appear to be overcompensated by the negative $W_{sft}$, leading to a probability distribution of total works $W_{total}$ with a tail at low work magnitudes of 13−40 kJ·mol$^{-1}$ (Figure 8).

For the slow pulling rate, the Jarzynski estimate for the raw binding free energy $\Delta G_{bnd,raw}$ converges rather fast to a value leading to a $\Delta G^{o}_{bnd}$ in good agreement with the PMF result. Already after 32 pulling simulations, a value of $-\Delta G_{bnd,raw} = 28.2$ kJ·mol$^{-1}$ is reached, which is within $k_BT$ of the value after 70 simulations, 30.1 kJ·mol$^{-1}$. In comparison, for the PMF calculation of system T4L, after 40 ns sampling per window, the in-solvent plateau value is within $k_BT$ of the corresponding value after 60 ns sampling (Figure 5a); *i.e.*, the corresponding compute effort for the nonequilibrium pulling simulations (32 simulations of 10 ns length) is approximately 30% that of the PMF calculation (28 windows of 40 ns length). The low number of nonequilibrium pulling simulations required here may be fortuitous, but it can also be a consequence of a balanced sampling of $W_{sft}$ and $W_{disres}$ contributions leading to

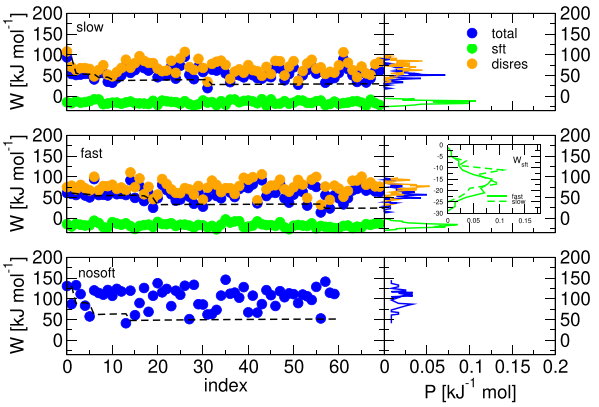



**Figure 8.** Work values sampled in nonequilibrium pulling simulations for the T4L system. Depicted are the total work ("total") and its softness ("sft") and distance restraint ("disres") contributions (eq 35), as a function of simulation index (left) or as a probability distribution (right), for either slow ("slow"; pulling rate 0.2 m·s$^{-1}$) or fast ("fast"; pulling rate 0.4 m·s$^{-1}$) pulling simulations with softness modification and for the pulling without softness modification ("nosoft"). The dashed lines give the running estimate of the free energy *via* the Jarzynski identity (eq 34). The inset compares the probability distribution of the softness work contribution $W_{sft}$ for the slow and fast pulling.

$W_{total}$ values whose distribution does not present a pronounced skewness toward high-magnitude work values (Figure 8). Yet, the distribution of $W_{total}$ does not satisfy Gaussianity; *i.e.*, the application of the cumulant-expansion approach (eq 40) fails (Table S1). This is presumably due to a large variance $\delta W^2$ rendering the resulting $\Delta G$ too low, both when the raw data are used to calculate the mean and variance and when a fitted Gaussian is used instead.

We also attempted a statistical-boosting approach[81] to analyze the nonequilibrium pulling simulations. However, permuting all the different values of $W_{sft}$ and $W_{disres}$ to generate a set of $70^2$ work values $W_{total}$ is only justified if the two composite works are uncorrelated. The 70 sampled pairs of $W_{sft}$ and $W_{disres}$ show Pearson correlation coefficients of −0.32 for the slow (pulling rate 0.2 m·s$^{-1}$) and −0.02 for the fast (pulling rate 0.4 m·s$^{-1}$) pulling, respectively. Kendall rank coefficients are similar (Figure 9a). Clearly, there appears to be a correlation, albeit weak, for the slow pulling, while correlation appears absent for the fast pulling. In Figure S5, the different work contributions and the total work are displayed along an example pulling simulation. At first glance, there appears to be an anticorrelation between $W_{sft}$ and $W_{disres}$ up to about a distance of 1.5 nm, interestingly corresponding to near the turning point in Λ (Figure 2d). Below that distance, the pulling work increases along with a decreasing softness work, passes through a maximum corresponding to a minimum in the softness work, and decreases toward negative values along with an increasing softness work entering the positive regime before both quantities slowly level off in a wiggling manner (Figure S5). Note that $W_{sft}$ has a zero at $\lambda = 0.25$ (distance of 1.09 nm), where the Λ-function peaks (Figure 2d); *i.e.*, below this point, softness rises, leading to predominantly negative $W_{sft}$, while above that point, softness decreases, leading to predominantly positive $W_{sft}$. Due to the different role and importance of protein−ligand softness along the pulling pathway, the extent of correlation between $W_{sft}$ and $W_{disres}$ should intuitively be different in different regimes along the


https://doi.org/10.1021/acs.jctc.1c01194
J. Chem. Theory Comput. 2022, 18, 3873−3893

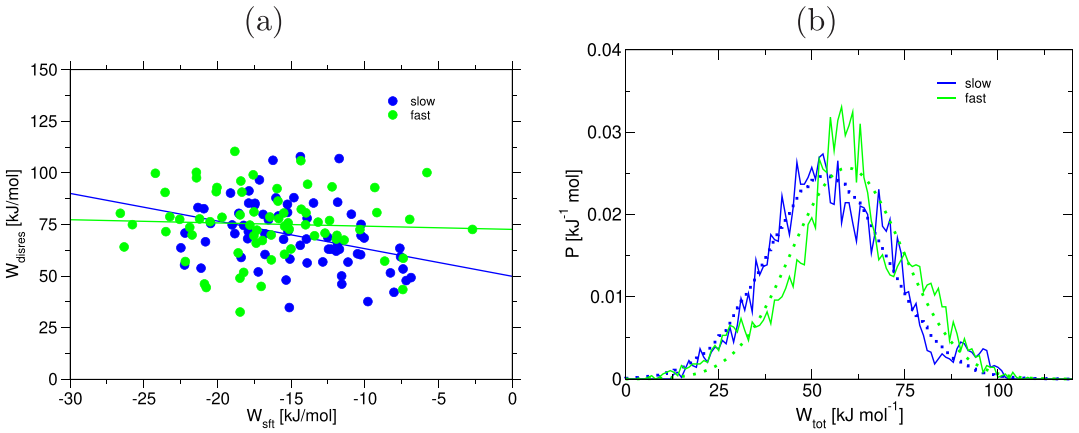



**Figure 9.** (a) Scatterplot showing the sampled pairs of softness ($W_{sft}$) and distance restraint ($W_{disres}$) work contributions (eq 35) for either the slow ("slow"; pulling rate 0.2 m·s$^{-1}$) or fast ("fast"; pulling rate 0.4 m·s$^{-1}$) pulling simulations with softness modification for the T4L system. A total of 70 work pairs are depicted per approach. The Pearson correlation or Kendall rank coefficients are −0.32 or −0.24 for the slow and −0.02 or −0.04 for the fast pulling, respectively. (b) Probability distribution of work values $W_{tot}$ rendered by $70^2$ combinations of softness and distance restraint work contributions (eq 35) obtained for the slow ("slow"; pulling rate 0.2 m·s$^{-1}$) or fast ("fast"; pulling rate 0.4 m·s$^{-1}$) pulling simulations with softness modification, for the T4L system, along with fitted Gaussians (dotted lines).

unbinding path. This was investigated in Figure S6, where the work pairs are shown for the vicinity of the binding pocket (integration up to $\lambda$ = 0.3; distance below 1.19 nm), an intermediate regime (integration between $\lambda$ = 0.3 and 0.6; distances of 1.19−1.79 nm), and the essentially solvated regime (integration from $\lambda$ = 0.6 to 1.0; distance beyond 1.79 nm). For the first regime, uncorrelation is seen (correlation coefficients of 0.03 and 0.17 for slow and fast pulling, respectively). For the intermediate regime, strong anti-correlation effects are observed, with a correlation coefficient about twice as high for slow than for fast pulling. In this regime, the Λ-function is decreasing, meaning that physical ligand−protein interactions are gradually restored, but the softness is still relatively high (Λ-function close to its maximum), which is why $W_{disres}$ is of lower magnitude than in the first regime. The lower $W_{disres}$ is, the more unfavorable it is to reinstall physical interactions, which explains the observed anti-correlation. It is unclear why it appears to be more pronounced for slow rather than for fast pulling. It may be because $W_{disres}$ does not have that low values for fast pulling. Finally, in the solvated regime, we expect vanishing correlation between the pulling work and the softness work. The latter almost vanishes here because of the large ligand−protein separation. Possibly, the small extent of correlation seen here may be related to slow environmental rearrangements creating a "memory effect" in the softness work. This effect appears to be more pronounced for the fast pulling, in agreement with the notion of nonequilibrium phenomena being more pronounced with faster pulling rates. Taken together, the relatively strong correlation between $W_{sft}$ and $W_{disres}$ is restricted to a narrow intermediate $\lambda$-regime, which is why the statistical-boosting approach seemed justified.

Statistical boosting delivers $70^2$ values of $W_{tot}$ whose distribution is shown in Figure 9b. Although a Gaussian can be fitted to these values, the cumulant-expansion approach (eq 40) still fails, possibly for the same reasons as hypothesized above for the nonboosted approach, namely, an overestimation of the variance (Figure S1).

Note, finally, that we also attempted to carry out the (slow) nonequilibrium pulling simulations without any simultaneous alchemical modifications. Doing so in the presence of the helix 4−helix 7 interface restraint, *i.e.*, a challenging situation, leads to 10 simulations out of the 70 started ones crashing due to failures in the bond-constraint algorithm as a consequence of high forces. The resulting $\Delta G_{bnd,raw}$ based on the Jarzynski identity (eq 34) is −51.0 kJ·mol$^{-1}$ (Table 1). The overestimated magnitude is caused by the sampling of comparably high pulling works. Due to the lack of softened interactions and hence the presence of steric clashes, structural perturbations in helix 7 appear to be more pronounced, especially in the region around 1.2−1.5 nm, where a local maximum in the helix 7 backbone rmsd is reached (rmsd of 0.30 nm; Figure S7).

## 5. CONCLUSIONS

In the present work, we combined alchemical modifications and physical-pathway methods for the calculation of absolute binding free energies. Our main motivation was to remove one of the limitations of physical-pathway methods, namely, the difficulty of applying them to occluded binding pockets. Besides, we were interested in how alchemical modifications may be used to deal with sticky protein−ligand interactions that may lead to sampling problems.

We devised two basic approaches: the simultaneous approach (S-approach), where, along the unbinding pathway, an alchemical transformation of ligand−protein interactions is introduced and faded out again to restore the physical interactions, and the prior-plus-simultaneous approach (PPS-approach), where, prior to the physical-pathway calculation, an alchemical transformation concerning ligand−protein interactions in the binding pocket is introduced and afterward faded out again during the physical-pathway calculation to restore the physical interactions. The physical-pathway methods applied in our work were either a PMF calculation using US and WHAM or multiple nonequilibrium pulling simulations in combination with the Jarzynski identity to calculate raw binding free energies. Our approach may be seen as an extension of the work of Cruz *et al.*[29] who also combined alchemical and physical-pathway free-energy calculations in the so-called AlchemPMF method. However, rather than carrying out the alchemical and physical-pathway transformation simultaneously as done in the S-approach, in the AlchemPMF


https://doi.org/10.1021/acs.jctc.1c01194
J. Chem. Theory Comput. 2022, 18, 3873−3893

method, two alchemical transformations are done at the end states of the physical pathway, *i.e.,* one with the aim to weaken or soften problematic interactions in the bound state and the second one with the aim to reinstall the normal interaction potential in the unbound state. The AlchemPMF method bears resemblance to the PPS-approach in the present work, but the difference is that the PPS-approach reinstalls normal interactions concurrent with ligand unbinding. Modifying and reinstalling the normal ligand−protein interaction potential simultaneously with physical ligand unbinding have the advantages of computational efficiency as well as ease of combination with non-equilibrium pulling simulations.

We investigated two different systems to address two separate issues, namely, a lysozyme mutant with a small-molecule ligand in a buried binding pocket (system T4L) where steric obstacles and conformational rearrangement would be encountered in straightforward physical unbinding scenarios and the MDM2 protein with a peptide ligand featuring strong hydrophobic interactions in the binding groove (system MDM). The problems posed by the nature of the crucial interactions were addressed by softening of interactions in the former and reduction of interactions in the latter system.

In principle, one could conceive of all possible combinations of approach (S- or PPS-approach), physical-pathway method (PMF calculation or nonequilibrium pulling), and interaction modification (softening or reduction) to explore the relative merits in the context of the two above systems. Since this would be out of the scope of the current investigation, we limited ourselves to three main combinations: (i) system T4L with PMF calculation and the S-approach with softening of interactions; (ii) system T4L with nonequilibrium pulling and the S-approach with softening of interactions ("soft Jarzynski method"); and (iii) system MDM with PMF calculation and the PPS-approach with reduction of interactions. These choices were guided by physical insight into the problematic processes of the systems, namely, steric clashes of the ligand due to the buried nature of the T4L binding pocket and high energy barriers due to relatively strong peptide−protein interactions in the MDM binding groove.

For points (i) and (ii), we could illustrate the significant benefits introduced by the alchemical modification during the PMF and nonequilibrium pulling simulations. Steric clashes were effectively removed by the S-approach with softening, which allowed successful free-energy evaluation despite the problems an occluded binding pocket would actually pose in standard, unmodified physical-pathway calculations. The PMF result for the standard binding free energy ($-11.6$ kJ·mol$^{-1}$) was in very good agreement with the Jarzynski result ($-10.2$ kJ·mol$^{-1}$) and in good agreement with double decoupling simulations carried out for comparison ($-9.8$ or $-8.5$ kJ·mol$^{-1}$). Nonequilibrium pulling simulations without simultaneous alchemical modifications were seen to fail for an equivalent compute effort. The "soft Jarzynski method" presented here hence appears to be a very promising tool for calculating free energies of processes that are associated with a large dissipated work when carried out in a nonequilibrium fashion, *e.g.,* pulling ligands out of buried binding pockets (present work) or pulling compounds through dense or strongly interacting media. The larger the dissipated work during the process is, the less likely it becomes that trajectories with negative dissipated work are sampled,[60] yet it is those trajectories that are actually giving important contributions to the exponential average in the Jarzynski identity. As we showed, the "soft Jarzynski method" achieves the sampling of the required low nonequilibrium work values and thus gives a free energy result in agreement with an US-based PMF calculation and at less computational cost, namely, for the considered system, approximately 30% of the compute effort. In addition, as a consequence of softened ligand−protein interactions, structural perturbations in the protein during the pulling are reduced in comparison to pulling simulations without modification of soft-core interactions.

Other approaches have been suggested before to improve nonequilibrium simulations in combination with the Jarzynski identity. For example, sampling of a wider work distribution may be enforced by the inclusion of external stochastic noise as a perturbation potential,[89] or the dissipated work may be calculated explicitly and thus serve as a correction.[90] It remains to be seen in future work how the "soft Jarzynski method" presented here compares to these approaches in terms of efficiency and accuracy.

For point (iii), we attained agreement between PMF calculations without or with alchemical modifications, yielding raw binding free energies of $-132.5$ and $-131.6$ kJ·mol$^{-1}$, respectively. Here, the PPS-approach with a reduction of interactions essentially shifted the problem of breaking strong interactions in the binding groove to an alchemical transformation carried out in a separate TI free-energy calculation. The free-energy barrier between the bound and unbound state in the PMF is thus greatly reduced. Furthermore, disentangling physical ligand displacement from the deinstallation of ligand−protein interactions was seen to allow a more uniform sampling of distance histograms in the US of the PMF calculation. This may improve histogram overlap and hence increase computational efficiency in certain systems. For the present system, an improvement of histogram overlap was only observed in small PMF regions; *i.e.,* no overall computational savings were gained in comparison to the unmodified PMF calculation.

Clearly, combining physical-pathway methods with simultaneous alchemical modifications can bring enormous benefits, the most important one being the extension of the applicability of PMF or nonequilibrium pulling simulations to occluded binding pockets. In addition, the method may also be useful to calculate binding free energies for two protein binding partners, *i.e.,* large systems with multiple, possibly sticky interactions and potential steric clashes encountered upon dissociation. A disadvantage of the approach is the requirement to design an appropriate Λ-function. This is less difficult for the PPS-approach, where, in most cases, the physical-pathway method should be accompanied by an initial, maximal plateau in Λ that decreases gradually toward the unbound state. It is more intricate for the S-approach, where the location of the maximum in the Λ-function has to be chosen in accordance with the crucial problematic physical interaction phenomena along the unbinding pathway, and its slopes have to be adapted to the range of problematic interactions along the reaction coordinate. A molecular-level characterization of ligand−environment interactions is thus highly useful. This can be done by analysis of interaction energies or local friction characteristics along the reaction coordinate. Identification of regions along the unbinding pathway that present high interaction energies and/or a high level of friction between the ligand and the environment and hence a corresponding high level of work dissipation during nonequilibrium pulling


https://doi.org/10.1021/acs.jctc.1c01194
J. Chem. Theory Comput. 2022, 18, 3873−3893

can guide the positioning of extrema in the Λ-function. The dissipated work can, *e.g.*, be determined based on the autocorrelation function of the constraint force along the reaction coordinate.[90] Work is ongoing toward an optimum exploitation of and quantitative microscopic insight in the effect of alchemical modifications during physical-pathway simulations.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.1c01194.

Distance-restraint free-energy contribution from TI (Figure S1); TI curves for vanishing interactions (Figure S2); distance histograms and histogram overlap for the PMF calculations of systems MDM-0 and MDM-MOD (Figure S3); average free-energy values from random samples for error calculation (Figure S4); example of a nonequilibrium pulling simulation of the T4L system (Figure S5); scatterplots showing the sampled pairs of softness and distance restraint works (Figure S6); root-mean-square deviation of backbone atoms (Figure S7); free energy values from the cumulant expansion approach (Table S1); average van der Waals and electrostatic interaction energies (Table S2); and accumulated time fraction of helical structure in helices 4 and 7 of system T4L (Table S3) (PDF)

## ■ AUTHOR INFORMATION


### Corresponding Author

**Maria M. Reif** − *Center for Protein Assemblies (CPA), Physics Department, Chair of Theoretical Biophysics (T38), Technical University of Munich, Garching 85748, Germany*; orcid.org/0000-0002-8171-3541; Phone: +49 89 289 12731; Email: maria.reif@tum.de

### Author

**Martin Zacharias** − *Center for Protein Assemblies (CPA), Physics Department, Chair of Theoretical Biophysics (T38), Technical University of Munich, Garching 85748, Germany*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.1c01194


### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

Computer resources for this project were provided by the compute cluster of the T38 group at the Technische Universität München, the Gauss Centre for Supercomputing/Leibniz Supercomputing Centre, and the Regionales Rechenzentrum Erlangen.

## ■ REFERENCES


(1) Ferreira, L. G.; Dos Santos, R. N.; Oliva, G.; Andricopulo, A. D. Molecular docking and structure-based drug design strategies. *Molecules* **2015**, *20*, 13384.

(2) Brooijmans, N.; Kuntz, I. D. Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335.

(3) Berendsen, H. J. C. *Simulating the physical world*; Cambridge University Press: Cambridge, UK, 2007.

(4) de Ruiter, A.; Oostenbrink, C. Advances in the calculation of binding free energies. *Curr. Opin. Struct. Biol.* **2020**, *61*, 207.

(5) Mobley, D. L.; Gilson, M. K. Predicting binding free energies: Frontiers and benchmarks. *Annu. Rev. Biophys.* **2017**, *46*, 531.

(6) Maximova, T.; Moffatt, R.; Ma, B.; Nussinov, R.; Shehu, A. Principles and overview of sampling methods for modeling macromolecular structure and dynamics. *PLoS Comput. Biol.* **2016**, *12*, No. e1004619.

(7) Limongelli, V. Ligand binding free energy and kinetics calculation in 2020. *WIREs Comput. Mol. Sci.* **2020**, *10*, e1455.

(8) Wan, S.; Bhati, A. P.; Zasada, S. J.; Coveney, P. V. Rapid, accurate, precise and reproducible ligand-protein binding free energy prediction. *Interfaces: Focus* **2020**, *10*, 20200007.

(9) Aqvist, J.; Medina, C.; Samuelsson, J.-E. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* **1994**, *7*, 385.

(10) Hansson, T.; Marelius, J.; Åqvist, J. Ligand binding affinity prediction by linear interaction energy methods. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 27.

(11) Genhenden, S.; Ryde, U. Comparison of the efficiency of the LIE and MM/GBSA methods to calculate ligand-binding energies. *J. Chem. Theory Comput.* **2011**, *7*, 3768.

(12) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., III Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889.

(13) Wang, E.; Sun, H.; Wang, J.; Wang, Z.; Liu, H.; Zhang, J. Z. H.; Hou, T. End-point binding free energy calculation with MM/PBSA and MM/GBSA: Strategies and applications in drug design. *Chem. Rev.* **2019**, *119*, 9478.

(14) Onufriev, A. V.; Case, D. A. Generalized Born implicit solvent models for biomolecules. *Annu. Rev. Biophys.* **2019**, *48*, 275.

(15) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187.

(16) Torrie, G. M.; Valleau, J. P. Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. *Chem. Phys. Lett.* **1974**, *28*, 578.

(17) den Otter, W. K. Revisiting the exact relation between potential of mean force and free-energy profile. *J. Chem. Theory Comput.* **2013**, *9*, 3861.

(18) Ferruz, N.; De Fabritiis, G. Binding kinetics in drug discovery. *Mol. Inf.* **2016**, *35*, 216.

(19) Marsh, J. A.; Teichmann, S. A. Structure, dynamics, assembly, and evolution of protein complexes. *Annu. Rev. Biochem.* **2015**, *84*, 551.

(20) Fry, C. Targeting protein-protein interactions for drug discovery. In *Protein-protein interactions*; *Methods Mol. Biol.* Humana press (Springer): New York, USA, Meyerkord, C.; Fu, H., Eds., 2015, 93−106.

(21) Villarreal, O. D.; Yu, L.; Rodriguez, R. A.; Chen, L. Y. Computing the binding affinity of a ligand buried deep inside a protein with the hybrid steered molecular dynamics. *Biochem. Biophys. Res. Commun.* **2017**, *483*, 203.

(22) Reif, M. M.; Fischer, M.; Fredriksson, K.; Hagn, F.; Zacharias, M. The N-Terminal segment of the voltage-dependent anion channel: A possible membrane-bound intermediate in pore unbinding. *J. Mol. Biol.* **2019**, *431*, 223.

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

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

(25) Rocklin, G. J.; Mobley, D. L.; Dill, K. A.; Hünenberger, P. H. Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: An

accurate correction scheme for electrostatic finite-size effects. *J. Chem. Phys.* **2013**, *139*, 184103.
(26) Reif, M. M.; Oostenbrink, C. Net charge changes in the calculation of relative ligand-binding free energies *via* classical atomistic molecular dynamics simulations. *J. Comput. Chem.* **2014**, *35*, 227.
(27) Öhlknecht, C.; Lier, B.; Petrov, D.; Fuchs, J.; Oostenbrink, C. Correcting electrostatic artifacts due to net-charge changes in the calculation of ligand binding free energies. *J. Comput. Chem.* **2020**, *41*, 986.
(28) Wu, J. Z.; Azimi, S.; Khuttan, S.; Deng, N.; Gallicchio, E. Alchemical transfer approach to absolute binding free energy estimation. *J. Chem. Theory Comput.* **2021**, *17*, 3309.
(29) Cruz, J.; Wickstrom, L.; Yang, D.; Gallicchio, E.; Deng, N. Combining alchemical transformation with a physical pathway to accelerate absolute binding free energy calculations of charged ligands to enclosed binding sites. *J. Chem. Theory Comput.* **2020**, *16*, 2803.
(30) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011.
(31) Roux, B. The calculation of the potential of mean force using computer simulation. *Comput. Phys. Commun.* **1995**, *91*, 275.
(32) Zhang, D.; Gullingsrud, J.; McCammon, A. Potentials of mean force for acetylcholine unbinding from the alpha7 nicotinic acetylcholine receptor ligand-binding domain. *J. Am. Chem. Soc.* **2006**, *128*, 3019.
(33) Vuong, Q. V.; Nguyen, T. T.; Li, M. S. A new method for navigating optimal direction for pulling ligand from binding pocket: Application to ranking binding affinity by steered molecular dynamics. *J. Chem. Inf. Model.* **2015**, *55*, 2731.
(34) Ngo, S. T.; Hung, H. M.; Nguyen, M. T. Fast and accurate determination of the relative binding affinities of small compounds to HIV-1 protease using non-equilibrium work. *J. Comput. Chem.* **2016**, *37*, 2734.
(35) Truong, D. T.; Li, M. S. Probing the binding affinity by Jarzynski's nonequilibrium binding free energy and rupture time. *J. Phys. Chem. B* **2018**, *122*, 4693.
(36) Jakubec, D.; Vondrášek, J. Efficient estimation of absolute binding free energy for a homeodomain-DNA complex from nonequilibrium pulling simulations. *J. Chem. Theory Comput.* **2020**, *16*, 2034.
(37) Ho, K.; Truong, D. T.; Li, M. S. How good is Jarzynski's equality for computer-aided drug design? *J. Phys. Chem. B* **2020**, *124*, 5338.
(38) Perthold, J. W.; Oostenbrink, C. GroScore: Accurate scoring of protein-protein binding poses using explicit-solvent free-energy calculations. *J. Chem. Inf. Model.* **2019**, *59*, 5074.
(39) 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. *Chem. Sci.* **2020**, *11*, 1140.
(40) Procacci, P.; Guarnieri, G. SAMPL7 blind predictions using nonequilibrium alchemical approaches. *J. Comput.-Aided Mol. Des.* **2021**, *35*, 37.
(41) Gapsys, V.; Yildirim, A.; Aldeghi, M.; Khalak, Y.; van der Spoel, D.; de Groot, B. L. Accurate absolute free energies for ligand-protein binding based on non-equilibrium approaches. *Commun. Chem.* **2021**, *4*, 61.
(42) Baumann, H. M.; Gapsys, V.; de Groot, B. L.; Mobley, D. L. Challenges encountered applying equilibrium and nonequilibrium free energy calculations. *J. Phys. Chem. B* **2021**, *125*, 4241.
(43) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute binding free energies: A quantitative approach for their calculation. *J. Phys. Chem. B* **2003**, *107*, 9535.
(44) Deng, Y.; Roux, B. Calculation of standard binding free energies: Aromatic molecules in the T4 lysozyme L99A mutant. *J. Chem. Theory Comput.* **2006**, *2*, 1255.
(45) Genheden, S.; Kongsted, J.; Söderhjelm, P.; Ryde, U. Nonpolar solvation free energies of protein-ligand complexes. *J. Chem. Theory Comput.* **2010**, *6*, 3558.
(46) Lim, N. M.; Wang, L.; Abel, R.; Mobley, D. L. Sensitivity in binding free energies due to protein reorganization. *J. Chem. Theory Comput.* **2016**, *12*, 4620.
(47) Xie, B.; Nguyen, T. H.; Minh, D. D. L. Absolute binding free energies between T4 lysozyme and 141 small molecules: Calculations based on multiple rigid receptor configurations. *J. Chem. Theory Comput.* **2017**, *13*, 2930.
(48) Niitsu, A.; Re, S.; Oshima, H.; Kamiya, M.; Sugita, Y. De novo prediction of binders and nonbinders for T4 lysozyme by gREST simulations. *J. Chem. Inf. Model.* **2019**, *59*, 3879.
(49) Feher, V. A.; Schiffer, J. M.; Mermelstein, D. J.; Mih, N.; Pierce, L. C. T.; McCammon, J. A.; Amaro, R. E. Mechanisms for benzene dissociation through the excited state of T4 lysozyme L99A mutant. *Biophys. J.* **2019**, *116*, 205.
(50) Estrada-Ortiz, N.; Neochoritis, C. G.; Dömling, A. How to design a successful p53-MDM2/X interaction inhibitor: A thorough overview based on crystal structure. *ChemMedChem* **2016**, *11*, 757.
(51) Lemos, A.; Leão, M.; Soares, J.; Palmeira, A.; Pinto, M.; Saraiva, L.; Sousa, M. E. Medicinal chemistry strategies to disrupt the p53-MDM2/MDMX interaction. *Med. Res. Rev.* **2016**, *36*, 789.
(52) Grasberger, B. L.; Lu, T.; Schubert, C.; Parks, D. J.; Carver, T. E.; Koblish, H. K.; Cummings, M. D.; LaFrance, L. V.; Milkiewicz, K. L.; Calvo, R. R.; Maguire, D.; Lattanze, J.; Franks, C. F.; Zhao, S.; Ramachandren, K.; Bylebyl, G. R.; Zhang, M.; Manthey, C. L.; Petrella, E. C.; Pantoliano, M. W.; Deckman, I. C.; Spurlino, J. C.; Maroney, A. C.; Tomczuk, B. E.; Molloy, C. J.; Bone, R. F. Discovery and cocrystal structure of benzodiazepinedione HDM2 antagonists that activate p53 in cells. *J. Med. Chem.* **2005**, *48*, 909.
(53) Schmid, N.; Christ, C. D.; Christen, M.; Eichenberger, A. P.; van Gunsteren, W. F. Architecture, implementation and parallelisation of the GROMOS software for biomolecular simulation. *Comput. Phys. Commun.* **2012**, *183*, 890.
(54) Beutler, T. C.; Mark, A. E.; van Schaik, R.; Gerber, P. R.; van Gunsteren, W. F. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* **1994**, *222*, 529.
(55) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration. *J. Chem. Phys.* **1994**, *100*, 9025.
(56) Barker, J. A.; Watts, R. O. Monte Carlo studies of the dielectric properties of water-like models. *Mol. Phys.* **1973**, *26*, 789.
(57) Riniker, S.; Christ, C. D.; Hansen, H. S.; Hünenberger, P. H.; Oostenbrink, C.; Steiner, D.; van Gunsteren, W. F. Calculation of relative free energies for ligand-protein binding, solvation and conformational transitions using the GROMOS biomolecular simulation software. *J. Phys. Chem. B* **2011**, *115*, 13570.
(58) Jarzynski, C. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* **1997**, *78*, 2690.
(59) Jarzynski, C. Equilibrium free-energy differences from non-equilibrium measurements: A master-equation approach. *Phys. Rev. E* **1997**, *56*, 5018.
(60) Gore, J.; Ritort, F.; Bustamante, C. Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 12564.
(61) Xiong, H.; Crespo, A.; Marti, M.; Estrin, D.; Roitberg, A. Free energy calculations with non-equilibrium methods: Applications of the Jarzynski relationship. *Theor. Chem. Acc.* **2006**, *116*, 338.
(62) Hummer, G. Nonequilibrium methods for equilibrium free energy calculations. In *Free energy calculations*; Springer: Berlin, Germany, Chipot, C.; Pohorille, A., Eds., 2007, 171−198.
(63) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction models for water in relation to protein hydration. In *Intermolecular Forces*; Reidel, Dordrecht: The Netherlands, Pullman, B., Ed., 1981, 331−342.


https://doi.org/10.1021/acs.jctc.1c01194
J. Chem. Theory Comput. 2022, 18, 3873−3893

(64) Merski, M.; Fischer, M.; Balius, T. E.; Eidam, O.; Shoichet, B. K. Homologous ligands accommodated by discrete conformations of a buried cavity. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 5039.
(65) Nunes-Alves, A.; Kokh, D. B.; Wade, R. C. Ligand unbinding mechanisms and kinetics for T4 lysozyme mutants from τRAMD simulations. *Curr. Res. Struct. Biol.* **2021**, *3*, 106.
(66) Reif, M. M.; Hünenberger, P. H.; Oostenbrink, C. New interaction parameters for charged amino acid side chains in the GROMOS force field. *J. Chem. Theory Comput.* **2012**, *8*, 3705.
(67) Horta, B. A. C.; Merz, P. T.; Fuchs, P.; Dolenc, J.; Riniker, S.; Hünenberger, P. H. A GROMOS-compatible force field for small organic molecules in the condensed phase: The 2016H66 parameter set. *J. Chem. Theory Comput.* **2016**, *12*, 3825.
(68) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; di Nola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684.
(69) Hockney, R. W. The potential calculation and some applications. *Methods Comput. Phys.* **1970**, *9*, 135.
(70) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **1977**, *23*, 327.
(71) Glättli, A.; Daura, X.; van Gunsteren, W. F. Derivation of an improved simple point charge model for liquid water: SPC/A and SPC/L. *J. Chem. Phys.* **2002**, *116*, 9811.
(72) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular simulation: The GROMOS96 manual and user guide*; Verlag der Fachvereine: Zürich, Switzerland 1996.
(73) Mondal, J.; Ahalawat, N.; Pandit, S.; Kay, L. E.; Vallurupalli, P. Atomic resolution mechanism of ligand binding to a solvent inaccessible cavity in T4 lysozyme. *PLoS Comput. Biol.* **2018**, *14*, No. e1006180.
(74) Grossfield, A. *The weighted histogram analysis method*; http://membrane.urmc.rochester.edu/content/wham
(75) Zwanzig, R. W. High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J. Chem. Phys.* **1954**, *22*, 1420.
(76) Bradford, S. Y. C.; El Khoury, L.; Ge, Y.; Osato, M.; Mobley, D. L.; Fischer, M. Temperature artifacts in protein structures bias ligand-binding predictions. *Chem. Sci.* **2021**, *12*, 11275.
(77) Schiffer, J. M.; Feher, V. A.; Malmstrom, R. D.; Sida, R.; Amaro, R. E. Capturing invsible motions in the transition from ground to rare excited states of T4 lysozyme L99A. *Biophys. J.* **2016**, *111*, 1631.
(78) Doudou, S.; Burton, N. A.; Henchman, R. H. Standard free energy of binding from a one-dimensional potential of mean force. *J. Chem. Theory Comput.* **2009**, *5*, 909.
(79) Duboué-Dijon, E.; Hénin, J. Building intuition for binding free energy calculations: Bound state definition, restraints, and symmetry. *J. Chem. Phys.* **2021**, *154*, 204101.
(80) Allen, M. P.; Tildesley, D. J. *Computer simulation of liquids*; Oxford University Press: New York, USA 1987.
(81) Procacci, P. Precision and computational efficiency of nonequilibrium alchemical methods for computing free energies of solvation. II. Unidirectional estimates. *J. Chem. Phys.* **2019**, *151*, 144115.
(82) Hendrix, D. A.; Jarzynski, C. A "fast growth" method of computing free energy differences. *J. Chem. Phys.* **2001**, *114*, 5974.
(83) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577.
(84) Eichenberger, A. P.; Allison, J. R.; Dolenc, J.; Geerke, D. P.; Horta, B. A. C.; Meier, K.; Oostenbrink, C.; Schmid, N.; Steiner, D.; Wang, D.; van Gunsteren, W. F. The GROMOS++ software for the analysis of biomolecular simulation trajectories. *J. Chem. Theory Comput.* **2011**, *7*, 3379.
(85) Luitz, M. P.; Bomblies, R.; Ramcke, E.; Itzen, A.; Zacharias, M. Adenylylation of Tyr77 stabilizes Rab1b GTPase in an active state: A molecular dynamics simulation analysis. *Sci. Rep.* **2016**, *6*, 19896.
(86) Morton, A.; Matthews, B. W. Specificity of ligand binding in a buried nonpolar cavity of T4 lysozyme: Linkage of dynamics and structural plasticity. *Biochemistry* **1995**, *34*, 8576.
(87) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129*, 124105.
(88) Popowicz, G. M.; Czarna, A.; Rothweiler, U.; Szwagierczak, A.; Krajewski, M.; Weber, L.; Holak, T. A. Molecular basis for the inhibition of p53 by Mdmx. *Cell Cycle* **2007**, *6*, 2386.
(89) Perišić, O.; Lu, H. On the improvement of free-energy calculation from steered molecular dynamics simulations using adaptive stochastic perturbation protocols. *PLoS One* **2014**, *9*, No. e101810.
(90) Wolf, S.; Stock, G. Targeted molecular dynamics calculations of free energy profiles using a nonequilibrium friction correction. *J. Chem. Theory Comput.* **2018**, *14*, 6175.