JCTC Journal of Chemical Theory and Computation

Article

Cite This: J. Chem. Theory Comput. 2017, 13, 5874−5886

pubs.acs.org/JCTC

Binding Free Energies of Host−Guest Systems by Nonequilibrium Alchemical Simulations with Constrained Dynamics: Theoretical Framework

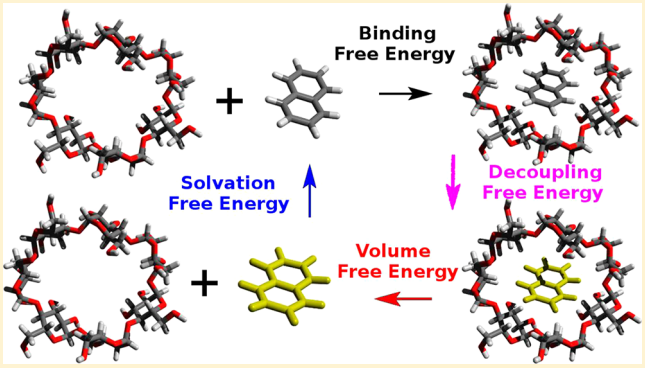
Edoardo Giovannelli,[†] Piero Procacci,[†] Gianni Cardini,[†] Marco Pagliai,[†] Victor Volkov,[‡] and Riccardo Chelli*[,†]

[†]Dipartimento di Chimica, Università di Firenze, Via della Lastruccia 3, I-50019 Sesto Fiorentino, Italy

[‡]Interdisciplinary Biomedical Research Center, School of Science and Technology, Nottingham Trent University, Clifton Lane, Nottingham NG11 8NS, U.K.

Supporting Information

ABSTRACT: The fast-switching decoupling method is a powerful nonequilibrium technique to compute absolute binding free energies of ligand−receptor complexes (Sandberg et al., *J. Chem. Theory Comput.* **2014**, *11*, 423−435). Inspired by the theory of noncovalent binding association of Gilson and co-workers (*Biophys. J.* **1997**, *72*, 1047−1069), we develop two approaches, termed binded-domain and single-point alchemical-path schemes (BiD-AP and SiP-AP), based on the possibility of performing alchemical trajectories during which the ligand is constrained to fixed positions relative to the receptor. The BiD-AP scheme exploits a recent generalization of nonequilibrium work theorems to estimate the free energy difference between the coupled and uncoupled states of the ligand−receptor complex. With respect to the fast-switching decoupling method without constraints, BiD-AP prevents the ligand from leaving the binding site, but still requires an estimate of the positional binding-site volume, which may not be a simple task. On the other side, the SiP-AP scheme allows avoidance of the calculation of the binding-site volume by introducing an additional equilibrium simulation of ligand and receptor in the bound state. In the companion article (DOI: 10.1021/acs.jctc.7b00595), we show that the extra computational effort required by SiP-AP leads to a significant improvement of accuracy in the free energy estimates.



1. INTRODUCTION

The fundamental role of standard absolute binding free energy (ABFE) of ligand−receptor complexes in chemistry, biology, and, especially, in drug discovery has stimulated an intensive research to design efficient computational strategies for fast and accurate free energy estimates[1−6] and for reliable ligand screening.[7,8] In the framework of molecular dynamics (MD) simulations, several approaches exploiting biasing potentials or restrained dynamics have been devised to compute the free energy difference of distinct configurational states of a system, in general, and of bound and unbound states of host−guest assemblies in particular.[9−20] An important class of methodologies revolves around alchemical transformations,[2,21−33] whose efficacy relies on the possibility of splitting the ABFE calculation of a ligand−receptor complex in two parts, one based on decoupling[21] or annihilation[27,34] of the ligand from the solvent in a simulation of the solvated ligand and the other on the decoupling of the ligand from its environment in a simulation of the solvated ligand−receptor complex. Alchemical transformations can in turn be performed by using equilibrium[21,27,34−36] and nonequilibrium[31−33,37] MD simulations. In equilibrium simulations, the intermolecular potential energy between ligand and environment changes reversibly through a series of independent simulations, called replicas, characterized by ligand-environment potential energies different to each other. These potential energies are associated with specific values of a parameter λ, typically ranging in the interval [0,1], where the extremes correspond to the fully coupled and uncoupled states of the ligand. Thus, in the ensemble of replicas, the ligand-environment coupling varies in discrete steps, according to the values of λ assigned to the various replicas. The free energy difference between the coupled and uncoupled state is determined from summing up the free energy differences estimated for the pairs of neighboring λ ensembles through thermodynamic integration,[38] free energy perturbation,[39] or Bennett acceptance ratio.[40,41] The ABFE can finally be computed as the difference between the free energies related to the decoupling processes[27] of the ligand alone in solution and of the ligand in the solvated complex, using a

Received: June 7, 2017
Published: October 9, 2017

DOI: 10.1021/acs.jctc.7b00594
J. Chem. Theory Comput. 2017, 13, 5874−5886

correction[21,42] to account for the reversible work needed to bring the ligand from the volume of the binding site to the volume of the standard state. To avoid the dissociation of the ligand−receptor complex during the simulations carried out at the various λ values, a restraining potential is usually employed.[9,35,36] Such a potential introduces a spurious contribution to the ABFE, which can be removed *a posteriori* with established relationships.[21,35]

In fast-switching alchemical transformations,[43] the free energies relative to decoupling processes are computed according to prescriptions of nonequilibrium work theorems[44] applied to MD simulations.[1] Initial microstates are sampled at equilibrium fixing $\lambda = 0$ (coupled state). Starting from each microstate, a set of nonequilibrium alchemical trajectories is realized by varying λ from 0 to 1 with a fixed time schedule. During such trajectories, the work performed on the system is computed, and the set of work values is employed into nonequilibrium work theorems[37] to find the free energy difference between final ($\lambda = 1$) and initial state ($\lambda = 0$). The absence of any artificial device aimed at restraining the ligand within the binding site of the receptor, makes formally incorrect the formulation of fast-switching alchemical transformations presented in ref 43. In spite of this, it has been shown that the use of restraining potentials can be avoided in the practice without affecting significantly the ABFE estimates. This relies on the fact that simulation times adopted in nonequilibrium alchemical transformations can be, in principle, arbitrary, and hence short enough to virtually prevent ligand−receptor dissociation. Although short simulation times are approachable in several cases without significant loss of accuracy, a check on each alchemical trajectory should always be performed to verify that no dissociation has occurred. Otherwise, the final uncoupled state would not be represented correctly, making the free energy evaluation wrong. In the free energy calculation, only the trajectories for which the final configuration of the complex is still in a bound state can be accepted. This posterior check makes necessary the use of some criterion to evaluate if bound configurations are preserved, thus introducing a certain degree of complexity in the analysis. Another not obvious issue is how to compute the binding-site volume entering the standard state correction in the fast-switching decoupling method.[30] Often, approximate estimates are given for the binding-site volume, based on the mean volume per atom in condensed phases under standard conditions. Typical free energy corrections range around, or less than, 1 kcal mol^{-1}.[31,45]

To tackle the shortcomings discussed above, we supplement the fast-switching decoupling method[30,43] with the possibility of performing alchemical trajectories during which the ligand is constrained to a fixed position relative to the receptor. Here, "position" must be intended as the vector connecting an atom of the receptor, taken as the origin of the receptor-frame, with an atom of the ligand, taken as the origin of the ligand-frame. Two types of approach are presented. The first, called the binded-domain alchemical-path (BiD-AP) scheme, is based on a MD simulation protocol that allows an estimation of the free energy differences between coupled and uncoupled states of the ligand−receptor complex by means of nonequilibrium MD simulations, exploiting the configurational-domain transition method proposed in ref 46. With respect to the fast-switching decoupling method without constraints,[30,43] the present approach prevents the ligand from leaving the binding site, but still requires an estimate of the binding-site volume. In the second alchemical method, called the single-point alchemical-path (SiP-AP) scheme, a reference configuration of the ligand−receptor complex is introduced to split the ligand to receptor/solvent decoupling contribution to the ABFE into two separate energetical terms, one computed from an equilibrium MD simulation of the fully coupled bound state of the complex and the other from fast-switching alchemical simulations of the complex constrained in the reference configuration. Both schemes allow computation of the ABFE without resorting to the calculation of the orientational binding-site volume, related[21] to the change in free energy when the rotationally constrained ligand is allowed to rotate freely. Furthermore, the SiP-AP scheme also avoids the calculation of the positional binding-site volume, which is related to the change in free energy when the constrained gas-phase ligand is allowed to expand to occupy the standard volume (1661 Å^3). These rotational and positional contributions to the ABFE do not simply "disappear" from the calculation, but are accounted for in an implicit way, through a potential of mean force as a function of the ligand position (rotational contribution) and through the integration domain of an integral entering the probability density as a function of the ligand position (positional contribution). The two alchemical schemes are based on a binding descriptor, namely the coordinate devised to establish when the ligand−receptor complex is in place, which corresponds to the position of a reference atom of the ligand relative to a given receptor-frame. In view of applications to inclusion host−guest systems, such as complexes of β-cyclodextrin with aromatic compounds,[47] BiD-AP and SiP-AP are also formulated so as to employ the distance between two generic points of ligand and receptor (atoms or centers of mass) as binding descriptor.

In ref 47, we illustrate important technical and theoretical aspects for a good practice in applying BiD-AP and SiP-AP alchemical schemes through the calculation of ABFEs of a Zn(II)·anion complex and 1:1 complexes of β-cyclodextrin with benzene and naphthalene.

2. THERMODYNAMICS OF THE NONCOVALENT BINDING

BiD-AP and SiP-AP schemes are developed starting from the theory of noncovalent binding association by Gilson and co-workers.[21] In this section, we review the basic relationships for the calculation of the standard ABFE through alchemical transformations, preserving the notation of ref 21 whenever possible. During the discussion, we will outline some differences with respect to outcomes of Gilson and co-workers, especially related to the possible geometries of the ligand. In section 3.1, we report on our approaches to the alchemical transformations based on constrained nonequilibrium MD simulations. In section 4, we describe how the relationships for alchemical transformations are modified from using the ligand−receptor distance as binding descriptor.

The reaction in which we are interested is the association of a ligand L with a receptor R to form a noncovalent complex RL in solution,

$$\mathrm{R} + \mathrm{L} \rightleftarrows \mathrm{RL} \tag{1}$$

At equilibrium, the chemical potentials of L, R, and RL into solution are equalized, namely

$$\mu_{\mathrm{sol,R}} + \mu_{\mathrm{sol,L}} = \mu_{\mathrm{sol,RL}} \tag{2}$$

The chemical potential of a species i at a given concentration C_i can be expressed as

DOI: 10.1021/acs.jctc.7b00594
J. Chem. Theory Comput. 2017, 13, 5874−5886

$$\mu_{\mathrm{sol},i} = \mu^{\circ}_{\mathrm{sol},i} + RT \ln \frac{\gamma_i C_i}{C^{\circ}} \tag{3}$$

where $\mu^{\circ}_{\mathrm{sol},i}$ is the standard chemical potential, γ_i is the activity coefficient, C° is the standard concentration in the same units as C_i (1 M or 1 molecule/1661 Å^3), R is the gas constant, and T is the absolute temperature. As Gilson and co-workers noted, $\mu^{\circ}_{\mathrm{sol},i}$ is the chemical potential in a hypothetical standard state in which each species is at standard concentration in the solvent, but does not interact with other solute molecules. It is worth noting that, in the infinite dilution limit, the activity coefficients of the solute species approach unity.[48,49] Recasting eqs 2 and 3, the relation between the standard free energy of binding and the binding constant K is obtained

$$\begin{aligned}\Delta G^{\circ} &\equiv \mu^{\circ}_{\mathrm{sol,RL}} - \mu^{\circ}_{\mathrm{sol,R}} - \mu^{\circ}_{\mathrm{sol,L}} \\ &= -RT \ln\left(\frac{\gamma_{\mathrm{RL}}}{\gamma_{\mathrm{R}}\gamma_{\mathrm{L}}}\frac{C^{\circ}C_{\mathrm{RL}}}{C_{\mathrm{R}}C_{\mathrm{L}}}\right) \\ &\equiv -RT \ln K\end{aligned} \tag{4}$$

A relationship to link the ABFE (ΔG°), and hence K, to statistical thermodynamic quantities has been derived by Hill in ref 50 and revised by Gilson and co-workers[21] to include explicitly the standard concentration,

$$\mu^{\circ}_{\mathrm{sol,R}} = -RT \ln\left(\frac{1}{V_{N,\mathrm{R}}C^{\circ}}\frac{Q_{N,\mathrm{R}}(V_{N,\mathrm{R}})}{Q_{N,0}(V_{N,0})}\right) + P^{\circ}\bar{V}_{\mathrm{R}} \tag{5}$$

with analogous expressions for the ligand L and the complex RL. In the previous equation, $Q_{N,\mathrm{R}}(V_{N,\mathrm{R}})$ is the canonical partition function for a solution consisting of N solvent molecules and one molecule R at volume $V_{N,\mathrm{R}}$, which is the volume of this solution when it is at equilibrium at the temperature T and standard pressure P°. Analogously, $Q_{N,0}(V_{N,0})$ is the canonical partition function of N solvent molecules alone at the volume $V_{N,0}$, namely the equilibrium volume of the pure-solvent sample at T and P° conditions. Finally, for large values of N, $\bar{V}_{\mathrm{R}}/N_A = V_{N,\mathrm{R}} - V_{N,0}$ is the volume change occurring when one molecule R is added to N molecules of solvent (N_A being the Avogadro's number). It is worth noting that the term $P^{\circ}\bar{V}_{\mathrm{R}}$ into eq 5 is typically very small,[51] because of small values of $\bar{V}_{\mathrm{R}}$.

We now report on a more detailed expression of the standard chemical potentials $\mu^{\circ}_{\mathrm{sol,R}}$ and $\mu^{\circ}_{\mathrm{sol,L}}$, by exploiting the representation of the canonical partition functions in terms of the classical statistical thermodynamics.[52,53] In this framework, the partition function $Q_{N,\mathrm{R}}(V_{N,\mathrm{R}})$ can be written as a phase-space integral separable as the product of an integral over the positional variables, that is, the atomic coordinates, and two integrals over the dynamical variables, that is, the conjugate momenta related to the solute and solvent atoms:

$$\begin{aligned}Q_{N,\mathrm{R}}(V_{N,\mathrm{R}}) = \frac{1}{\sigma_{\mathrm{sol,R}}\sigma_{\mathrm{S}}^{N}} \int e^{-\beta U(\mathbf{r}'_{\mathrm{R}},\mathbf{r}_{\mathrm{S}})}\, d\mathbf{r}'_{\mathrm{R}}\, d\mathbf{r}_{\mathrm{S}} \\ \times \int \exp\left(-\beta \sum_{i=1}^{M_{\mathrm{R}}} \frac{p_i^2}{2m_{\mathrm{R},i}}\right) d\mathbf{p}_{\mathrm{R}} \int \exp\left(-\beta \sum_{i=1}^{N_{\mathrm{S}}N} \frac{p_i^2}{2m_{\mathrm{S},i}}\right) d\mathbf{p}_{\mathrm{S}}\end{aligned} \tag{6}$$

where β is the inverse temperature, p_i^2 is the squared magnitude of the momentum of the generic atom i, M_{R} is the number of atoms of R, $\mathbf{r}'_{\mathrm{R}}$ and $\mathbf{p}_{\mathrm{R}}$ denote the atomic coordinates and conjugate momenta of R, respectively, while $\mathbf{r}_{\mathrm{S}}$ and $\mathbf{p}_{\mathrm{S}}$ are the analogous variables for the $N_{\mathrm{S}}N$ solvent atoms (here, N_{S} is the number of atoms for one solvent molecule). Also, $m_{\mathrm{R},i}$ and $m_{\mathrm{S},i}$ indicate the mass of atom i belonging to receptor and solvent, respectively. We note that, at variance with the integral over the conjugate momenta, the integral over $\mathbf{r}'_{\mathrm{R}}$ and $\mathbf{r}_{\mathrm{S}}$ cannot be split, because the coordinates of solute and solvent are inextricably connected through mixed terms into $U(\mathbf{r}'_{\mathrm{R}},\mathbf{r}_{\mathrm{S}})$. In eq 6, $\sigma_{\mathrm{sol,R}}$ and σ_{S} are the symmetry numbers of R into solution and of a solvent molecule into a pure solvent sample. Specifying that the symmetry number of R is related to the solution environment is mandatory, because similar factors will be introduced for the gas phase and the complex RL. It is worth considering that the factor arising from the quantum-mechanical correction is not included in the expression of $Q_{N,\mathrm{R}}(V_{N,\mathrm{R}})$, because it cancels out with other analogous contributions to the ABFE. We now introduce a molecular axis system to separate the lab-frame coordinates $\mathbf{r}'_{\mathrm{R}}$ into internal and external. This molecular axis system is built taking as reference three atoms of R. Atom 1 becomes the origin of the molecular coordinates, denoted as $\mathbf{R}_{\mathrm{R}}$. The vector joining atom 1 with atom 2 defines the x-axis. The direction of the y-axis is given by the direction of the vector joining atoms 2 and 3, minus the x-component of this vector. The z-axis is constructed as the cross-product of vectors along the x and y-axes. The six coordinates thus fixed, namely $\mathbf{R}_{\mathrm{R}}$ plus the Eulerian angles $\xi_{\mathrm{R},1}$, $\xi_{\mathrm{R},2}$, and $\xi_{\mathrm{R},3}$ that specify the orientation of the molecular frame relative to the lab-frame, correspond to the external coordinates of R. The set of $3M_{\mathrm{R}} - 6$ internal coordinates of R will be indicated with $\mathbf{r}_{\mathrm{R}}$. Noting that the integrals over $\mathbf{r}_{\mathrm{R}}$ and $\mathbf{r}_{\mathrm{S}}$ do not depend upon the position and orientation of R, the integrals over $\mathbf{R}_{\mathrm{R}}$, $\xi_{\mathrm{R},1}$, $\xi_{\mathrm{R},2}$, and $\xi_{\mathrm{R},3}$ can be done at once. Considering that R is typically a polyatomic nonlinear molecule, the integrals yield $8\pi^2 V_{N,\mathrm{R}}$. Moreover, considering that the integral over the momenta components of an atom of mass m yields a factor $(2\pi mRT)^{3/2}$, the partition function of eq 6 can be written as

$$\begin{aligned}Q_{N,\mathrm{R}}(V_{N,\mathrm{R}}) = \frac{8\pi^2 V_{N,\mathrm{R}} Z_{N,\mathrm{R}}}{\sigma_{\mathrm{sol,R}}\sigma_{\mathrm{S}}^{N}} \prod_{i=1}^{M_{\mathrm{R}}} (2\pi m_{\mathrm{R},i}RT)^{3/2} \\ \prod_{i=1}^{N_{\mathrm{S}}N} (2\pi m_{\mathrm{S},i}RT)^{3/2}\end{aligned} \tag{7}$$

where

$$Z_{N,\mathrm{R}} = \int e^{-\beta U(\mathbf{r}_{\mathrm{R}},\mathbf{r}_{\mathrm{S}})}\, d\mathbf{r}_{\mathrm{R}}\, d\mathbf{r}_{\mathrm{S}} \tag{8}$$

is the configuration integral for a system consisting of one R molecule into N solvent molecules. In a similar way, we may express the partition function of N solvent molecules as

$$\begin{aligned}Q_{N,0}(V_{N,0}) &= \frac{1}{\sigma_{\mathrm{S}}^{N}} \int e^{-\beta U(\mathbf{r}_{\mathrm{S}})}\, d\mathbf{r}_{\mathrm{S}} \int \exp\left(-\beta \sum_{i=1}^{N_{\mathrm{S}}N} \frac{p_i^2}{2m_{\mathrm{S},i}}\right) d\mathbf{p}_{\mathrm{S}} \\ &= \frac{Z_{N,0}}{\sigma_{\mathrm{S}}^{N}} \prod_{i=1}^{N_{\mathrm{S}}N} (2\pi m_{\mathrm{S},i}RT)^{3/2}\end{aligned} \tag{9}$$

where $Z_{N,0}$ is the configuration integral for the solvent sample

$$Z_{N,0} = \int e^{-\beta U(\mathbf{r}_{\mathrm{S}})}\, d\mathbf{r}_{\mathrm{S}} \tag{10}$$

Upon substitution of eqs 7 and 9 into eq 5, we obtain

DOI: 10.1021/acs.jctc.7b00594
J. Chem. Theory Comput. 2017, 13, 5874−5886

$$\mu^{\circ}_{\mathrm{sol,R}} = -RT\ln\left(\frac{8\pi^2}{C^{\circ}\sigma_{\mathrm{sol,R}}}\prod_{i=1}^{M_{\mathrm{R}}}(2\pi m_{\mathrm{R},i}RT)^{3/2}\frac{Z_{N,\mathrm{R}}}{Z_{N,0}}\right) + P^{\circ}\overline{V}_{\mathrm{R}} \tag{11}$$

Similar arguments lead to a relationship for $\mu^{\circ}_{\mathrm{sol,L}}$. However, considering that the ligand can be also linear in shape and even a single atom, integration over the orientational degrees of freedom can give $8\pi^2$, 4π, and 1, respectively (from now on, this geometry factor will be denoted as $\mathcal{V}_{\xi_{\mathrm{L}}}$). Therefore, the expression for $\mu^{\circ}_{\mathrm{sol,L}}$ is

$$\mu^{\circ}_{\mathrm{sol,L}} = -RT\ln\left(\frac{\mathcal{V}_{\xi_{\mathrm{L}}}}{C^{\circ}\sigma_{\mathrm{sol,L}}}\prod_{i=1}^{M_{\mathrm{L}}}(2\pi m_{\mathrm{L},i}RT)^{3/2}\frac{Z_{N,\mathrm{L}}}{Z_{N,0}}\right) + P^{\circ}\overline{V}_{\mathrm{L}} \tag{12}$$

where the product is extended to the M_{L} atoms of the ligand and $m_{\mathrm{L},i}$ is the mass of atom i of the ligand. The calculation of the standard chemical potential of the complex $\mu^{\circ}_{\mathrm{sol,RL}}$ requires a specific treatment of the external and internal coordinates of RL. The former are assumed to be the external coordinates of R, while the external coordinates of L, indicated as $\zeta_{\mathrm{L}} \equiv (\mathbf{R}_{\mathrm{L}}, \xi_{\mathrm{L},1}, \xi_{\mathrm{L},2}, \xi_{\mathrm{L},3})$, are taken to be defined relative to R, so that they become internal coordinates of the complex. The arguments adopted to determine $\mu^{\circ}_{\mathrm{sol,R}}$ and $\mu^{\circ}_{\mathrm{sol,L}}$ may also be used here with the difference that the configuration integral of the complex must be restricted to the configurations for which R and L are complexed.[53] This can be realized introducing an indicator function $I(\zeta_{\mathrm{L}})$ that holds 1 for bound configurations and 0 otherwise. We then obtain the following expression,

$$\begin{aligned}\mu^{\circ}_{\mathrm{sol,RL}} = {} & -RT\ln\left(\frac{8\pi^2}{C^{\circ}\sigma_{\mathrm{cp,L}}\sigma_{\mathrm{cp,R}}}\frac{Z_{N,\mathrm{RL}}}{Z_{N,0}}\right) \\ & -RT\ln\left(\prod_{i=1}^{M_{\mathrm{L}}}(2\pi m_{\mathrm{L},i}RT)^{3/2}\prod_{i=1}^{M_{\mathrm{R}}}(2\pi m_{\mathrm{R},i}RT)^{3/2}\right) \\ & + P^{\circ}\overline{V}_{\mathrm{RL}}\end{aligned} \tag{13}$$

In the previous equation, $Z_{N,\mathrm{RL}}$ is the configuration integral of RL into solution

$$Z_{N,\mathrm{RL}} = \int I(\zeta_{\mathrm{L}})J_{\zeta_{\mathrm{L}}}\, e^{-\beta U(\zeta_{\mathrm{L}},\mathbf{r}_{\mathrm{L}},\mathbf{r}_{\mathrm{R}},\mathbf{r}_{\mathrm{S}})}\, d\zeta_{\mathrm{L}}\, d\mathbf{r}_{\mathrm{L}}\, d\mathbf{r}_{\mathrm{R}}\, d\mathbf{r}_{\mathrm{S}} \tag{14}$$

where $J_{\zeta_{\mathrm{L}}}$ is the absolute value of the Jacobian determinant for the rotation and translation of L relative to R, and $\mathbf{r}_{\mathrm{L}}$ denotes the internal coordinates of L. We remark that, for purposes of generality, we keep the full dependence of $J_{\zeta_{\mathrm{L}}}$ on the translational and rotational (external) coordinates of L. Instead, Gilson and co-workers[21] take a Jacobian determinant only dependent on the rotation of L, implicitly assuming that the position of L, namely $\mathbf{R}_{\mathrm{L}}$, is relative to a Cartesian reference R-frame. Furthermore, in eq 13, $\sigma_{\mathrm{cp,L}}$ and $\sigma_{\mathrm{cp,R}}$ are the symmetry numbers associated with L and R when the complex is in place.[54] A critical discussion on the symmetry numbers is reported in section I of the Supporting Information. Recasting eqs 11, 12, and 13 into eq 4, the expression for the standard ABFE is recovered

$$\begin{aligned}\Delta G^{\circ} = {} & -RT\ln\left(\frac{C^{\circ}}{\mathcal{V}_{\xi_{\mathrm{L}}}}\frac{\sigma_{\mathrm{sol,L}}\sigma_{\mathrm{sol,R}}}{\sigma_{\mathrm{cp,L}}\sigma_{\mathrm{cp,R}}}\frac{Z_{N,\mathrm{RL}}Z_{N,0}}{Z_{N,\mathrm{R}}Z_{N,\mathrm{L}}}\right) \\ & + P^{\circ}(\overline{V}_{\mathrm{RL}} - \overline{V}_{\mathrm{R}} - \overline{V}_{\mathrm{L}})\end{aligned} \tag{15}$$

3. THE DOUBLE-DECOUPLING METHOD

The double-decoupling method is a route to the estimate of ΔG° and is based on the calculation of the free energy differences associated with two independent processes entering the thermodynamic cycle represented in Figure 1. One process,

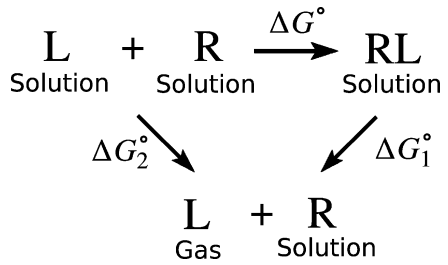


Figure 1. Thermodynamic cycle describing the double-decoupling method.

related to the free energy change ΔG°_1, is the decoupling of L from the solvated RL complex (right process in Figure 1). The other process, related to the free energy change ΔG°_2, is the decoupling of L from the solvent (left process in Figure 1). While the former process is physically meaningless, the latter corresponds to the desolvation free energy of L. In the former case, decoupling is accomplished by turning off the interactions of L with solvent and receptor R in a solution of RL, while in the latter case decoupling is realized by turning off the interactions of L with the solvent in a solution of L. It is important to remark that, in both situations, we do not deal with a total annihilation of L, because its intramolecular interactions are left in place, and hence L is virtually "transformed" into an ideal-gas molecule.

Before discussing the decoupling processes and in particular the details of our approach, it is mandatory to relate ΔG° to the quantities ΔG°_1 and ΔG°_2. According to Gilson and co-workers,[21] ΔG°_1 and ΔG°_2 can be written as

$$\Delta G^{\circ}_1 = \mu^{\circ}_{\mathrm{sol,R}} + \mu^{\circ}_{\mathrm{gas,L}} - \mu^{\circ}_{\mathrm{sol,RL}} \tag{16}$$

$$\Delta G^{\circ}_2 = \mu^{\circ}_{\mathrm{gas,L}} - \mu^{\circ}_{\mathrm{sol,L}} \tag{17}$$

where $\mu^{\circ}_{\mathrm{gas,L}}$ is the standard chemical potential of L in the ideal-gas phase and the other standard chemical potentials are defined in eqs 11, 12, and 13. Considering eq 4 together with eqs 16 and 17, we can immediately show that

$$\Delta G^{\circ} = \Delta G^{\circ}_2 - \Delta G^{\circ}_1 \tag{18}$$

Such a relation is also inferred straightforwardly by the thermodynamic cycle reported in Figure 1. In the next sections, we will show how ΔG°_1 and ΔG°_2 can be expressed in terms of configuration integrals, ultimately allowing for a description through potentials of mean force.

3.1. Decoupling the Ligand from Solvent and Receptor: ΔG°_1 Calculation. The standard chemical potential of L in the ideal-gas phase, $\mu^{\circ}_{\mathrm{gas,L}}$, is related to the natural logarithm of the molecular partition function as

$$\mu^{\circ}_{\mathrm{gas,L}} = -RT\ln Q_{0,\mathrm{L}}(V^{\circ}) \tag{19}$$

where it is explicitly reported that the partition function must be evaluated in the phase space limited to the standard volume $V^{\circ} = 1/C^{\circ}$. Following the arguments leading to eq 12, we get

DOI: 10.1021/acs.jctc.7b00594
J. Chem. Theory Comput. 2017, 13, 5874−5886

$$\mu^{\circ}_{\mathrm{gas,L}} = -RT \ln\left(\frac{\mathcal{V}_{\xi_L} Z_{0,\mathrm{L}}}{C^{\circ}\sigma_{\mathrm{gas,L}}} \prod_{i=1}^{M_L} (2\pi m_{\mathrm{L},i} RT)^{3/2}\right) \tag{20}$$

In the previous equation, $\mathcal{V}_{\xi_L}$ is from the integral over the orientation of L ($\mathcal{V}_{\xi_L} = 8\pi^2$, 4π, 1 for nonlinear, linear, and single-atom ligands, respectively), $\sigma_{\mathrm{gas,L}}$ is the symmetry number of L in the ideal-gas phase, and $Z_{0,\mathrm{L}}$ is the configuration integral in the internal coordinates:

$$Z_{0,\mathrm{L}} = \int \mathrm{e}^{-\beta U(\mathbf{r}_L)}\, \mathrm{d}\mathbf{r}_L \tag{21}$$

The external coordinates of L are integrated in eq 20, giving the contribution $\mathcal{V}_{\xi_L}/(C^{\circ}\sigma_{\mathrm{gas,L}})$. Substituting eqs 11, 13, and 20 into eq 16, we obtain

$$\Delta G^{\circ}_{\mathrm{I}} = -RT \ln\left(\frac{\mathcal{V}_{\xi_L}}{C^{\circ}} \frac{\sigma_{\mathrm{cp,L}}\sigma_{\mathrm{cp,R}}}{\sigma_{\mathrm{gas,L}}\sigma_{\mathrm{sol,R}}} \frac{Z_{N,\mathrm{R}} Z_{0,\mathrm{L}}}{Z_{N,\mathrm{RL}}}\right) + P^{\circ}(\bar{V}_{\mathrm{R}} - \bar{V}_{\mathrm{RL}}) \tag{22}$$

In the double-decoupling method, an artificial energy function $U(\lambda, \zeta_L, \mathbf{r}_L, \mathbf{r}_R, \mathbf{r}_S)$ dependent on a control parameter $\lambda \in [0,1]$ is introduced, whose functional form is rather arbitrary. The only requirements are that for $\lambda = 0$ and $\lambda = 1$ the function must correspond to the energy functions of the coupled and uncoupled states of the ligand in the complex, respectively:

$$U(0, \zeta_L, \mathbf{r}_L, \mathbf{r}_R, \mathbf{r}_S) = U(\zeta_L, \mathbf{r}_L, \mathbf{r}_R, \mathbf{r}_S) \tag{23}$$

$$U(1, \zeta_L, \mathbf{r}_L, \mathbf{r}_R, \mathbf{r}_S) = U(\mathbf{r}_R, \mathbf{r}_S) + U(\mathbf{r}_L) \tag{24}$$

Exploiting the artificial energy function, a free energy function dependent parametrically on λ can be built as

$$g(\lambda) = -RT \ln\left(\int I(\zeta_L) J_{\zeta_L}\, \mathrm{e}^{-\beta U(\lambda,\zeta_L,\mathbf{r}_L,\mathbf{r}_R,\mathbf{r}_S)}\, \mathrm{d}\zeta_L\, \mathrm{d}\mathbf{r}_L\, \mathrm{d}\mathbf{r}_R\, \mathrm{d}\mathbf{r}_S\right) \tag{25}$$

According to $g(\lambda)$ and to the requirements of eqs 23 and 24, the free energy difference between the final and initial states is

$$\begin{aligned} g(1) - g(0) &= -RT \ln\left(\frac{\int I(\zeta_L) J_{\zeta_L}\, \mathrm{e}^{-\beta U(\mathbf{r}_R,\mathbf{r}_S)}\, \mathrm{e}^{-\beta U(\mathbf{r}_L)}\, \mathrm{d}\zeta_L \mathrm{d}\mathbf{r}_L\, \mathrm{d}\mathbf{r}_R\, \mathrm{d}\mathbf{r}_S}{\int I(\zeta_L) J_{\zeta_L}\, \mathrm{e}^{-\beta U(\zeta_L,\mathbf{r}_L,\mathbf{r}_R,\mathbf{r}_S)}\, \mathrm{d}\zeta_L\, \mathrm{d}\mathbf{r}_L\, \mathrm{d}\mathbf{r}_R\, \mathrm{d}\mathbf{r}_S}\right) \\ &= -RT \ln\left(\frac{V_{\zeta_L} Z_{N,\mathrm{R}} Z_{0,\mathrm{L}}}{Z_{N,\mathrm{RL}}}\right) \end{aligned} \tag{26}$$

where the definitions of the configuration integrals, eqs 8, 14, and 21, have been used in deriving the second line of eq 26, after integrating the numerator over ζ_L. This operation is allowed because the exponential functions in the numerator are independent of ζ_L. Such an integration gives the binding-site volume $\int I(\zeta_L) J_{\zeta_L} \mathrm{d}\zeta_L = V_{\zeta_L}$, a quantity to be estimated numerically. Here, it is worth noting that, owing to the inextricable connection between the translational and rotational coordinates of L ($\mathbf{R}_L$, $\xi_{L,1}$, $\xi_{L,2}$, and $\xi_{L,3}$) within the indicator function $I(\zeta_L)$, the integral $\int I(\zeta_L) J_{\zeta_L} \mathrm{d}\zeta_L$ cannot be split straightforwardly into a product of a translational volume V_I and a rotational volume ξ_I, as instead reported in ref 21 (cf. eq 26 above with eq 25 of ref 21). Factorization of this integral in the product $V_I\xi_I$ would lead inevitably to an overestimate of the overall binding-site volume. Thus, translational and rotational degrees of freedom of L relative to R (external coordinates of L) within the binding site contribute to the ABFE, and eq 26 quantifies such a contribution.

Substituting eq 26 into eq 22, leads to

$$\Delta G_{\mathrm{I}}{}^{\circ} = g(1) - g(0) - RT \ln\left(\frac{\mathcal{V}_{\xi_L}}{C^{\circ} V_{\zeta_L}} \frac{\sigma_{\mathrm{cp,L}}\sigma_{\mathrm{cp,R}}}{\sigma_{\mathrm{gas,L}}\sigma_{\mathrm{sol,R}}}\right) + P^{\circ}(\bar{V}_{\mathrm{R}} - \bar{V}_{\mathrm{RL}}) \tag{27}$$

This relationship is the main outcome of Gilson and co-workers that we report here in a slightly modified form for the sake of comparison (cf. with eq 28 of ref 21):

$$\Delta G^{\circ}_{\mathrm{I}} = g(1) - g(0) - RT \ln\left(\frac{8\pi^2}{C^{\circ} V_{\mathrm{I}} \xi_{\mathrm{I}}} \frac{\sigma_{\mathrm{RL}}}{\sigma_{\mathrm{L}}\sigma_{\mathrm{R}}}\right) + P^{\circ}(\bar{V}_{\mathrm{R}} - \bar{V}_{\mathrm{RL}}) \tag{28}$$

Three significant differences between eqs 27 and 28 can be remarked. The first, discussed above, relies on the association of the integral over ζ_L with a single term, V_{ζ_L}, instead of the product $V_I\,\xi_I$ coming from translational and rotational degrees of freedom of L. Moreover, eq 27 provides a more detailed definition of symmetry numbers, outlining the fact that ligand and receptor may change symmetry upon going from one phase to another. Furthermore, we have also considered the possibility of dealing with linear molecules or single atoms as ligands. This is disclosed by the introduction of the parameter $\mathcal{V}_{\xi_L}$ instead of the factor $8\pi^2$ appearing in eq 28, the latter being valid only for nonlinear ligands. The quantity $g(1) - g(0)$ can be evaluated via equilibrium MD simulations exploiting the method of thermodynamic integration.[21] However, to gain an estimate of $\Delta G^{\circ}_{\mathrm{I}}$ one must also determine V_{ζ_L}, which may not be a straightforward task. Also, it is worth noting that eq 27 strictly holds for simulations performed in the canonical (NVT) ensemble, since only the artificial potential energy function appears in the exponential function of $g(\lambda)$ (see eq 25). When MD simulations are performed by adopting equations of motion which preserve NPT conditions, we are actually employing a free energy function supplemented with a $P^{\circ}V$ term in the exponential function. As proved in section II of the Supporting Information, the use of a NPT-like free energy function allows us to access directly to $\Delta G^{\circ}_{\mathrm{I}}$ without any correction for the partial molar volumes $\bar{V}_{\mathrm{R}}$ and $\bar{V}_{\mathrm{RL}}$. In the following, in order to adhere to the Gilson and co-workers' treatment,[21] we preserve the assumptions for the canonical ensemble, keeping in mind that the pressure-times-volume corrections must not be considered when simulating in the NPT conditions.

In this study, we propose to modify the paradigm for the ligand−receptor binding, adopting a criterion based only on the position of L relative to R, previously denoted as $\mathbf{R}_L$. This corresponds to turn from a binding function expressed in terms of position and orientation of L, that is, $I(\zeta_L) \equiv I(\mathbf{R}_L, \xi_{L,1}, \xi_{L,2}, \xi_{L,3})$, to a binding function expressed in terms of the position of L alone, that is, $I(\mathbf{R}_L)$. This assumption is consistent with the common idea that binding occurs basically when the ligand and receptor come into contact, regardless of the mutual orientation defined by the variables $\xi_{L,1}$, $\xi_{L,2}$, and $\xi_{L,3}$. Furthermore, limiting the indicator function to the only positional coordinates of L could be advantageous in the practice, since it would lead to simpler constraining protocols for L. Of course, for a generic

DOI: 10.1021/acs.jctc.7b00594
J. Chem. Theory Comput. 2017, 13, 5874−5886

position $\mathbf{R}_L$ satisfying the binding condition $I(\mathbf{R}_L) = 1$, most orientations of L relative to R will have a negligible probability of being observed in the coupled state, because of strong atomic overlaps between R and L. As a consequence, these configurations will contribute negligibly to the denominator of eq 26. This scheme allows us to rewrite the free energy function of eq 25 as

$$g(\lambda) = -RT \ln\left(\int I(\mathbf{R}_L) J_{\mathbf{R}_L} J_{\xi_L} \, e^{-\beta U(\lambda, \mathbf{R}_L, \xi_L, \mathbf{r}_L, \mathbf{r}_R, \mathbf{r}_S)} \, d\mathbf{R}_L \, d\xi_L \, d\mathbf{r}_L \, d\mathbf{r}_R \, d\mathbf{r}_S \right) \tag{29}$$

where ξ_L is a shorthand for $(\xi_{L,1}, \xi_{L,2}, \xi_{L,3})$, J_{ξ_L} and $J_{\mathbf{R}_L}$ are the absolute values of the Jacobian determinants for the (external) rotational and translational coordinates of L, respectively, and $d\xi_L \equiv d\xi_{L,1} d\xi_{L,2} d\xi_{L,3}$. As noted below eq 14, the Jacobian determinant $J_{\mathbf{R}_L}$ is in general different from 1, being 1 only when $\mathbf{R}_L$ is expressed in a Cartesian reference system. The free energy difference $g(1) - g(0)$ of eq 26 is then restated as

$$\begin{aligned} g(1) - g(0) &= -RT \ln\left(\frac{\int I(\mathbf{R}_L) J_{\mathbf{R}_L} J_{\xi_L} \, e^{-\beta U(\mathbf{r}_R, \mathbf{r}_S)} e^{-\beta U(\mathbf{r}_L)} \, d\mathbf{R}_L \, d\xi_L \, d\mathbf{r}_L \, d\mathbf{r}_R \, d\mathbf{r}_S}{\int I(\mathbf{R}_L) J_{\mathbf{R}_L} J_{\xi_L} \, e^{-\beta U(\mathbf{R}_L, \xi_L, \mathbf{r}_L, \mathbf{r}_R, \mathbf{r}_S)} \, d\mathbf{R}_L \, d\xi_L \, d\mathbf{r}_L \, d\mathbf{r}_R \, d\mathbf{r}_S} \right) \\ &= -RT \ln\left(V_I \mathcal{V}_{\xi_L} \frac{Z_{N,R} Z_{0,L}}{Z_{N,RL}} \right) \end{aligned} \tag{30}$$

In the third line of eq 30, we have carried out the integration over $\mathbf{R}_L$ and ξ_L in the numerator, obtaining $V_I = \int I(\mathbf{R}_L) J_{\mathbf{R}_L} \, d\mathbf{R}_L$ and $\mathcal{V}_{\xi_L} = \int J_{\xi_L} \, d\xi_L$, the latter being $8\pi^2$, 4π, or 1, according to the structure of L. Contrarily to eq 26, separation of these integrals can be done into eq 30 because the adopted binding criterion does not involve the rotational coordinates ξ_L.

Since the bound states of the complex are identified on the basis of $\mathbf{R}_L$, it is convenient to introduce a potential of mean force as a function of λ, which includes the internal coordinates of R and L, the coordinates of the solvent and the orientational coordinates of L. This potential results to be a function of both λ and $\mathbf{R}_L$:

$$e^{-\beta\phi(\lambda, \mathbf{R}_L)} = \int J_{\xi_L} \, e^{-\beta U(\lambda, \mathbf{R}_L, \xi_L, \mathbf{r}_L, \mathbf{r}_R, \mathbf{r}_S)} d\xi_L d\mathbf{r}_L \, d\mathbf{r}_R \, d\mathbf{r}_S \tag{31}$$

According to the above definition of potential of mean force, the free energy function $g(\lambda)$ (eq 29) takes the following simplified form

$$g(\lambda) = -RT \ln\left(\int I(\mathbf{R}_L) J_{\mathbf{R}_L} \, e^{-\beta\phi(\lambda, \mathbf{R}_L)} \, d\mathbf{R}_L \right) \tag{32}$$

Using the definition 32 of $g(\lambda)$ into eq 30, we obtain

$$g(1) - g(0) = -RT \ln\left(\frac{\int I(\mathbf{R}_L) J_{\mathbf{R}_L} \, e^{-\beta\phi(1, \mathbf{R}_L)} \, d\mathbf{R}_L}{\int I(\mathbf{R}_L) J_{\mathbf{R}_L} \, e^{-\beta\phi(0, \mathbf{R}_L)} \, d\mathbf{R}_L} \right) \tag{33}$$

It is worthwhile to note that, in the previous equation, $\phi(1, \mathbf{R}_L)$ does not depend on $\mathbf{R}_L$.[55] Nonetheless, in order to preserve consistency of notation, from now on the symbol $\mathbf{R}_L$ will be explicitly indicated into $\phi(\lambda, \mathbf{R}_L)$, regardless of λ.

3.1.1. BiD-AP Scheme in Nonequilibrium Alchemical Transformations. In the following two sections, we report on two alternative schemes to compute ΔG_I°. The first, termed BiD-AP (binded-domain alchemical-path) scheme, is based on the direct estimate of $g(1) - g(0)$ (eq 33). In this aspect the methodology is analogous to that of Gilson and co-workers.[21] In particular, $g(1) - g(0)$ is computed from nonequilibrium MD simulations, instead of using thermodynamic integration via equilibrium MD simulations. In nonequilibrium alchemical transformations, according to the fact that the end states must be related to the complex RL (see eq 33), the initial microstates have to represent bound RL configurations sampled at equilibrium.[43] Moreover, in order to attain a bound RL configuration in the final microstate, we must prevent the ligand from leaving the binding site during the sampling. When dealing with a strongly bound complex, the correct sampling weight of the initial microstates can be guaranteed implicitly by an equilibrium MD simulation, without enforcing any constraint to keep the ligand in the bound state. In such a case, a precise definition of bound-complex configurations is unimportant so long as binding is tight and all the statistically important bound configurations are sampled during the simulation.[21] For weak complexes, preserving bound configurations during a standard equilibrium MD simulation can instead be difficult. This requires that bound RL configurations are sampled by enforcing some walled potential matching $I(\mathbf{R}_L)$. This equilibrium sampling provides an amount of isothermally and isobarically sampled microstates, say N_{traj}, to be taken as initial phase-space points for the nonequilibrium alchemical trajectories. Equation 33 establishes that L must be in the same binding site in both the initial and final states. This can be accomplished by creating a bijective mapping between these states, with the aim of preventing the ligand from leaving the binding site. Recently, some of us developed a nonequilibrium approach able to guarantee such a mapping,[46] allowing the estimate of free energy differences between two configurational domains by means of steered MD simulations combined with nonequilibrium work theorems.[44,56−58] The method is based on the creation of a phase-space mapping applied during the nonequilibrium trajectories, whether to the control parameter employed to switch the system from the initial to the final state or to some phase-space variable (not directly correlated to the control parameter) taken to define the two configurational domains. The latter is just the situation that we may apply to the alchemical transformations. Evolving in time the λ control parameter from 0 (coupled ligand) to 1 (uncoupled ligand) according to some established time schedule, the coordinate $\mathbf{R}_L$ of the ligand relative to the receptor is mapped to bring the system from a coupled to an uncoupled configuration within the binding site. This is accomplished by fixing the $\mathbf{R}_L$ coordinate to the initial value (obtained from the equilibrium sampling) during the alchemical transformation, thus preventing the ligand from leaving the binding site. A constraint to $\mathbf{R}_L$ can be applied whether using some constraining method, such as RATTLE[59] or SHAKE,[60] or more simply by enforcing stiff (harmonic) potentials to the three components of $\mathbf{R}_L$. Using this simulation scheme, we thus produce N_{traj} alchemical trajectories that allow computation of the free energy difference $g(1) - g(0)$ by using the Jarzynski equality:[44]

$$g(1) - g(0) = -RT \ln\langle e^{-\beta W} \rangle \tag{34}$$

where the average is performed over the N_{traj} work values W associated with the alchemical trajectories. For a generic trajectory, the work is computed with the standard formula[1]

DOI: 10.1021/acs.jctc.7b00594
J. Chem. Theory Comput. 2017, 13, 5874−5886

$$W = \int_0^\tau \frac{\partial U(\lambda, \mathbf{R}_L, \xi_L, \mathbf{r}_L, \mathbf{r}_R, \mathbf{r}_S)}{\partial t}\, dt \quad (35)$$

where $\mathbf{R}_L$ is fixed to the value of the initial microstate and τ is the duration time of the alchemical trajectory. We outline that the explicit dependence on time lies only on the λ parameter, while the other variables are uncontrolled degrees of freedom. Furthermore, it is worth remarking that the validity of eq 34 for computing $g(1) - g(0)$ of eq 33 stems from having imposed a mapping which leaves the RL complex within the same binding site for the whole ensemble of alchemical paths.

Once the quantity $g(1) - g(0)$ is estimated, the contribution ΔG_1° to the ABFE can be computed through the following relationship (use the third line of eq 30 into eq 22)

$$\Delta G^\circ{}_1 = g(1) - g(0) - RT \ln\left(\frac{1}{V_I C^\circ}\frac{\sigma_{cp,L}\sigma_{cp,R}}{\sigma_{gas,L}\sigma_{sol,R}}\right) + P^\circ(\bar{V}_R - \bar{V}_{RL}) \quad (36)$$

It is important to note that, as in the thermodynamic integration method, the calculation of the binding-site volume V_I needs to be carried out. An illustrative example on how to compute V_I for a Zn(II)·anion complex is given in ref 47. A schematic illustration of the BiD-AP scheme is reported and shortly described in Figure 2.

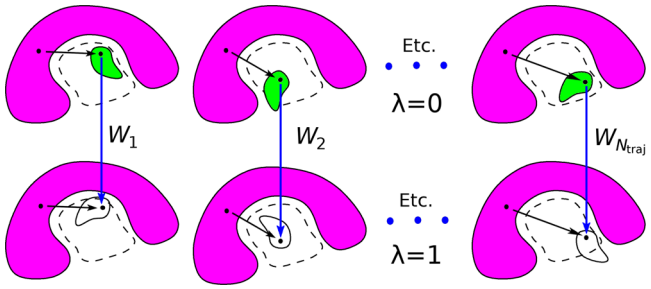


Figure 2. Schematic illustration of the BiD-AP scheme. R is displayed in magenta, while L in the coupled and uncoupled states is in green and white, respectively. The black circles are the origins of the R and L-frames. The volume $V_I = \int I(\mathbf{R}_L) J_{\mathbf{R}_L} d\mathbf{R}_L$ entering eq 36 is computed from an equilibrium simulation of the complex in the binding site defined by the dashed lines. The initial microstates of the alchemical trajectories are represented by the top configurations. They are sampled from an equilibrium simulation of the complex in the binding site, with L coupled to R and solvent ($\lambda = 0$). Such a simulation is the one also adopted for computing V_I. The position of the L-frame relative to the R-frame, $\mathbf{R}_L$ (black arrows), is fixed during each alchemical trajectory. The final microstates of the alchemical trajectories are represented by the bottom configurations (with L decoupled from R and solvent, that is, $\lambda = 1$). The work values W_1, W_2, ..., $W_{N_{traj}}$ performed on the system during the alchemical trajectories are calculated using eq 35 and employed into eq 34 to recover $g(1) - g(0)$, to be finally used into eq 36.

3.1.2. SiP-AP Scheme in Nonequilibrium Alchemical Transformations. To avoid the calculation of V_I, which implies knowledge of a way of evaluating the function $I(\mathbf{R}_L)$, we propose a different way to compute the ratio of integrals appearing in the second line of eq 30. This second approach, termed SiP-AP (single-point alchemical-path) scheme, has some similarity with other alchemical methods based on equilibrium MD simulations.[35,36,61] Noting that $e^{-\beta\phi(1,\mathbf{R}_L)}$ does not depend on $\mathbf{R}_L$[55] and that $\int I(\mathbf{R}_L) J_{\mathbf{R}_L} d\mathbf{R}_L = V_I$, we can rewrite eq 33 as

$$g(1) - g(0) = -RT \ln\left(\frac{V_I\, e^{-\beta\phi(1,\mathbf{R}_L)}}{\int I(\mathbf{R}_L) J_{\mathbf{R}_L} e^{-\beta\phi(0,\mathbf{R}_L)} d\mathbf{R}_L}\right) \quad (37)$$

In the previous equation, the quantity $\mathcal{V}_{\xi_L}$ does not appear in the numerator because the integral over the orientational coordinates of L is included into $e^{-\beta\phi(1,\mathbf{R}_L)}$ (see eq 31). Substituting eq 37 into eq 36 yields

$$\Delta G_1{}^\circ = -RT \ln\left(\frac{\sigma_{cp,L}\sigma_{cp,R}}{C^\circ\sigma_{gas,L}\sigma_{sol,R}}\frac{e^{-\beta\phi(1,\mathbf{R}_L)}}{\int I(\mathbf{R}_L) J_{\mathbf{R}_L} e^{-\beta\phi(0,\mathbf{R}_L)} d\mathbf{R}_L}\right) + P^\circ(\bar{V}_R - \bar{V}_{RL}) \quad (38)$$

With respect to the BiD-AP scheme represented by eq 36, explicit knowledge of the positional binding-site volume V_I is not necessary in eq 38. On the other side, here we need to compute the integral over $\mathbf{R}_L$, which implies to determine the difference between the potentials of mean force for the coupled and uncoupled systems as a function of $\mathbf{R}_L$, *i.e.* $\phi(0, \mathbf{R}_L) - \phi(1, \mathbf{R}_L)$. Indeed, this may not be a simple task. To tackle this problem, we resort to a reference configuration of the complex RL featured by an established position of L, say $\mathbf{R}_L'$. The definition of this configurational state allows us to write

$$\frac{e^{-\beta\phi(1,\mathbf{R}_L)}}{\int I(\mathbf{R}_L) J_{\mathbf{R}_L} e^{-\beta\phi(0,\mathbf{R}_L)} d\mathbf{R}_L} = \frac{e^{\beta[\phi(0,\mathbf{R}_L') - \phi(1,\mathbf{R}_L')]}}{\int I(\mathbf{R}_L) J_{\mathbf{R}_L} e^{\beta[\phi(0,\mathbf{R}_L') - \phi(0,\mathbf{R}_L)]} d\mathbf{R}_L} \quad (39)$$

where, considering that $\phi(1,\mathbf{R}_L)$ is independent of $\mathbf{R}_L$, the equality $\phi(1,\mathbf{R}_L) = \phi(1,\mathbf{R}_L')$ has been used. Numerator and denominator of the right-hand side of eq 39 can be computed separately. The denominator can be computed from an equilibrium MD simulation of the RL complex (for tight binding), or using some method to sample preferentially bound configurations of the complex, such as the umbrella sampling method[62] (for weak binding). In any case, regardless of the employed simulation method, configurations featured by $\mathbf{R}_L = \mathbf{R}_L'$ must be sampled during the equilibrium MD simulation, as the function $\phi(0,\mathbf{R}_L)$ must be defined at the configuration $\mathbf{R}_L'$. Therefore, even if the position $\mathbf{R}_L'$ of the reference configuration can in principle be chosen arbitrarily, it is statistically convenient that $I(\mathbf{R}_L') = 1$, or better that $\mathbf{R}_L'$ falls in a binding-site region with small value of the potential of mean force (high probability region).

We point out that the denominator of the second line of eq 39 corresponds to the probability density of finding the ligand at the position $\mathbf{R}_L'$ once the complex RL is formed. This can be recognized writing the denominator as follows

$$\rho(\mathbf{R}_L') \equiv \frac{e^{-\beta\phi(0,\mathbf{R}_L')}}{\int I(\mathbf{R}_L) J_{\mathbf{R}_L} e^{-\beta\phi(0,\mathbf{R}_L)} d\mathbf{R}_L} = \frac{\Delta p(\mathbf{R}_L')}{J_{\mathbf{R}_L'}\Delta\mathbf{R}_L} \quad (40)$$

where $\Delta p(\mathbf{R}_L')$ is the infinitesimal probability that L is found in the volume element $J_{\mathbf{R}_L'}\Delta\mathbf{R}_L$ centered into $\mathbf{R}_L'$ during an equilibrium sampling of the complex in the bound state. Note that the Jacobian determinant $J_{\mathbf{R}_L'}$ is computed at the position

DOI: 10.1021/acs.jctc.7b00594
J. Chem. Theory Comput. 2017, 13, 5874−5886

$\mathbf{R}'_L$. Let us suppose that the bound state of the complex is sampled through an equilibrium simulation, or, more generally, through a simulation adopting some biasing potential, for example, using umbrella sampling.[62] In such a situation, we can define a position $\mathbf{R}'_L$ of L and a resolution $\Delta\mathbf{R}_L$ for establishing when the system takes that position. Denoting the number of times the system visits the configuration $\mathbf{R}'_L$ as $\Delta\mathcal{N}_{\mathbf{R}'_L}$ and the total number of bound configurations sampled during the MD simulation as $\mathcal{N}_{tot}$, the probability of interest is simply computed as

$$\Delta p(\mathbf{R}'_L) = \frac{\Delta\mathcal{N}_{\mathbf{R}'_L}}{\mathcal{N}_{tot}} \tag{41}$$

As stated above, $\Delta p(\mathbf{R}'_L)$ must be computed from an equilibrium MD simulation of the bound complex. This requirement leads to the sampling problems already discussed for the BiD-AP scheme, specifically when dealing with a weakly bound complex. As suggested in section 3.1.1, we may resort to hard-walled or restraining potentials to enforce the sampling of bound configurations. Additionally, soft potentials combined to reweighting procedures[62] may be employed to restrain the ligand in the binding site.[63−65] In such cases, an energy term would appear in eq 24 that is explicitly dependent on $\mathbf{R}_L$, meaning that the potential of mean force at $\lambda = 1$ is no longer independent from $\mathbf{R}_L$. This implies that the equality $\phi(1,\mathbf{R}_L) = \phi(1, \mathbf{R}'_L)$, required to write down eq 39, no longer holds. The difference between the two potentials of mean force can however be estimated analytically, as described in refs 35 and 36.

The numerator of the second line of eq 39 is estimated through an alchemical transformation. Analogously to the BiD-AP scheme, a number of initial microstates are sampled at equilibrium by fixing the position of L to $\mathbf{R}'_L$. Starting from these microstates, nonequilibrium trajectories are performed with an established time schedule for λ, from $\lambda = 0$ to $\lambda = 1$. The works computed from these trajectories via eq 35 are thus employed in the Jarzynski equality[44] (eq 34) to evaluate the free energy difference between the initial and final states, which corresponds to $\phi(1, \mathbf{R}'_L) - \phi(0, \mathbf{R}'_L)$.

In summary, considering the introduction of a reference configuration (eq 39) and the definition of probability density (eq 40), ΔG_1° can be rewritten as

$$\Delta G_1^\circ = \phi(1, \mathbf{R}'_L) - \phi(0, \mathbf{R}'_L) - RT \ln\left(\rho(\mathbf{R}'_L)\frac{\sigma_{cp,L}\sigma_{cp,R}}{C^\circ\sigma_{gas,L}\sigma_{sol,R}}\right) + P^\circ(\overline{V}_R - \overline{V}_{RL}) \tag{42}$$

where $\rho(\mathbf{R}'_L)$ and the difference $\phi(1,\mathbf{R}'_L) - \phi(0,\mathbf{R}'_L)$ are computed as described above. A schematic illustration of the SiP-AP scheme is shown and shortly described in Figure 3. We point out that, when NPT simulations are performed in the place of NVT simulations, eq 42 still holds, with the only difference being that no corrections dependent upon the partial molar volumes $\overline{V}_R$ and $\overline{V}_{RL}$ have to be considered.

Moreover, it is important to note that the SiP-AP scheme can be applied with both equilibrium (e.g., thermodynamic integration[35]) and nonequilibrium alchemical simulations, while the BiD-AP methodology is intrinsically a nonequilibrium simulation technique. As a matter of fact, applying the SiP-AP scheme in an equilibrium simulation framework, which simply corresponds to enforcing a constraint to the translations of the ligand, is straightforward, as shown, for example, by Deng and Roux in ref 36. In this regard, eq 42 may be viewed as a reformulation[29,35] of the Deng and Roux approach in the limit of strong restraints, where only the ligand position vector is held fixed at $\mathbf{R}'_L$ during the alchemical decoupling.

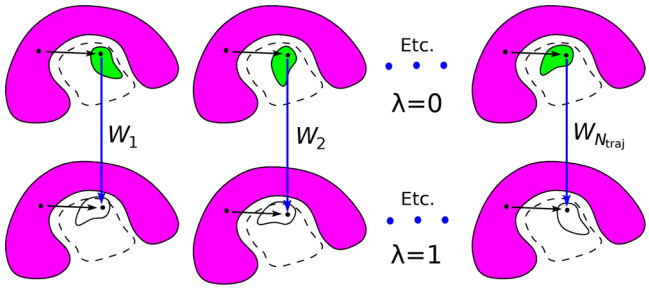


Figure 3. Schematic illustration of the SiP-AP scheme. R is displayed in magenta, while L in the coupled and uncoupled states is in green and white, respectively. The black circles are the origins of the R and L-frames. The quantity $\rho(\mathbf{R}'_L)$ entering eq 42 is computed from an equilibrium simulation of the complex in the binding site defined by the dashed lines (no constraints are applied to L). The initial microstates of the alchemical trajectories are represented by the top configurations. They are sampled from an equilibrium simulation of the complex in which L is fixed at the position $\mathbf{R}'_L$ (black arrows) and coupled to R and solvent ($\lambda = 0$). The position of L, $\mathbf{R}'_L$, is fixed during each alchemical trajectory and is the same for all trajectories. The final microstates of the alchemical trajectories are represented by the bottom configurations (with L decoupled from R and solvent, that is, $\lambda = 1$). The work values W_1, W_2, ..., $W_{N_{traj}}$ performed on the system during the alchemical trajectories are calculated using eq 35 and employed into eq 34 to recover $\phi(1, \mathbf{R}'_L) - \phi(0,\mathbf{R}'_L)$, to be finally used into eq 42.

As a final remark, we notice that the BiD-AP and SiP-AP methods can be extended to alchemical simulations involving also rotational constraints. Extension of the theory to account for alchemical processes where the ligand is subject to both translational and rotational constraints is formulated in section III of the Supporting Information.

3.2. Decoupling the Ligand from the Solvent: ΔG_2° Calculation. The contribution ΔG_2° to ΔG° corresponds to the free energy difference between the state in which L is decoupled from the solvent (L in the gas phase and pure solvent in the condensed phase) and the state in which L is coupled to the solvent (solution of L in the solvent). From the physical standpoint, ΔG_2° therefore represents the desolvation free energy of L. It is obtained by substituting eqs 12 and 20 into eq 17,

$$\Delta G_2^\circ = -RT \ln\left(\frac{\sigma_{sol,L}}{\sigma_{gas,L}}\frac{Z_{N,0}Z_{0,L}}{Z_{N,L}}\right) - P^\circ\overline{V}_L \tag{43}$$

where $Z_{N,L}$, $Z_{N,0}$, and $Z_{0,L}$ are the usual configuration integrals. At variance with ΔG_1°, the contribution ΔG_2° does not depend upon the choice of the standard concentration. In this case, the artificial energy function $U(\lambda, \zeta_L, \mathbf{r}_L, \mathbf{r}_S)$ does not depend upon the internal coordinates of R, as we simply deal with L in the solvent. The requirements on U are that, for $\lambda = 0$ and $\lambda = 1$, the artificial energy function must correspond to the energy functions of the coupled and uncoupled states of L in the solvent, respectively:

$$U(0, \zeta_L, \mathbf{r}_L, \mathbf{r}_S) = U(\zeta_L, \mathbf{r}_L, \mathbf{r}_S) \tag{44}$$

DOI: 10.1021/acs.jctc.7b00594
J. Chem. Theory Comput. 2017, 13, 5874−5886

$$U(1, \zeta_L, \mathbf{r}_L, \mathbf{r}_S) = U(\mathbf{r}_S) + U(\mathbf{r}_L) \quad (45)$$

where the external coordinates of L are now relative to the lab-frame. A free energy function dependent parametrically on λ can be built exploiting the artificial energy function as

$$g(\lambda) = -RT \ln\left(\int J_{\zeta_L} e^{-\beta U(\lambda, \zeta_L, \mathbf{r}_L, \mathbf{r}_S)} d\zeta_L \, d\mathbf{r}_L \, d\mathbf{r}_S\right) \quad (46)$$

According to $g(\lambda)$ and to the requirements of eqs 44 and 45, the free energy difference between the final and initial states is

$$g(1) - g(0) = -RT \ln\left(\frac{\int J_{\zeta_L} e^{-\beta U(\mathbf{r}_S)} e^{-\beta U(\mathbf{r}_L)} d\zeta_L \, d\mathbf{r}_L \, d\mathbf{r}_S}{\int J_{\zeta_L} e^{-\beta U(\zeta_L, \mathbf{r}_L, \mathbf{r}_S)} d\zeta_L \, d\mathbf{r}_L \, d\mathbf{r}_S}\right) = -RT \ln\left(\frac{Z_{N,0} Z_{0,L}}{Z_{N,L}}\right) \quad (47)$$

Note that the integrals over the internal coordinates of the solute, $\mathbf{r}_L$, and over the coordinates of the solvent, $\mathbf{r}_S$, do not depend upon the position or orientation of the solute, ζ_L, and hence the integrals over ζ_L may be carried out at once yielding $V\mathcal{V}_{\xi_L}$, where V is the volume of the container (simulation box) and arises from the integral over the position $\mathbf{R}_L$, while $\mathcal{V}_{\xi_L}$ arises from the integral over the orientation $(\xi_{L,1}, \xi_{L,2}, \xi_{L,3})$. As this volume term appears in both numerator and denominator of eq 47, it cancels out.

We may now define the free energy function $g(\lambda)$ in terms of the potential of mean force as a function of position and orientation of the ligand:

$$g(\lambda) = -RT \ln\left(\int J_{\zeta_L} e^{-\beta \Phi(\lambda, \zeta_L)} d\zeta_L\right) \quad (48)$$

where

$$e^{-\beta \Phi(\lambda, \zeta_L)} = \int e^{-\beta U(\lambda, \zeta_L, \mathbf{r}_L, \mathbf{r}_S)} d\mathbf{r}_L \, d\mathbf{r}_S \quad (49)$$

As observed above, the integrals over $\mathbf{r}_L$ and $\mathbf{r}_S$ into eq 49 do not depend upon ζ_L. For this reason, the potential of mean force $\Phi(\lambda, \zeta_L)$ is independent of ζ_L and hence it will be denoted as $\Phi(\lambda)$. This allows to write eq 48 as

$$g(\lambda) = \Phi(\lambda) - RT \ln(V\mathcal{V}_{\xi_L}) \quad (50)$$

Using eq 50 into eq 47 for expressing $g(0)$ and $g(1)$ and substituting the resulting equation into eq 43 yields

$$\Delta G^\circ_2 = \Phi(1) - \Phi(0) - RT \ln\left(\frac{\sigma_{sol,L}}{\sigma_{gas,L}}\right) - P^\circ \overline{V}_L \quad (51)$$

We notice that the knowledge of $\sigma_{gas,L}$ is not mandatory, because it drops out when eq 51 is recombined with eq 36 (if using BiD-AP) or eq 42 (if using SiP-AP) to recover ΔG° via eq 18.

Operatively, ΔG°_2 can be computed using nonequilibrium MD simulations in the usual way. First, a set of initial microstates is produced through an equilibrium MD simulation of one L molecule into N solvent molecules (without any constraint). Starting from these microstates, nonequilibrium trajectories are performed with an established time schedule for λ, starting from $\lambda = 0$ (coupled ligand) and ending to $\lambda = 1$ (uncoupled ligand). The works computed from these alchemical trajectories by means of a relationship analogous to eq 35[66] are then employed in the Jarzynski equality[44] (eq 34) to evaluate the free energy difference $\Phi(1) - \Phi(0)$, to be finally used into eq 51.

4. USING THE LIGAND–RECEPTOR DISTANCE AS BINDING DESCRIPTOR IN THE DOUBLE-DECOUPLING METHOD

The alchemical schemes presented in sections 3.1.1 and 3.1.2 are based on a binding descriptor relying on the position of a reference atom of L relative to an atom of R, specifically the $\mathbf{R}_L$ vector. A special important case of such an approach is to use a binding descriptor based on the magnitude of $\mathbf{R}_L$. Here, the ligand–receptor distance is assumed to be the fundamental quantity defining a bound state, which is, in general, a reliable assumption when configurations belonging to a limited orientational subspace of $\mathbf{R}_L$ contribute to the bound state. In addition, the ligand–receptor distance is suitable for describing complexes forming isotropic or nearly isotropic bound configurations of the complex, such as a ligand within a cage-like binding site. This binding descriptor has been employed to determine the ABFEs of complexes of β-cyclodextrin with aromatic compounds.[47] In this section, we discuss how the basic relationships of the method, namely eq 36 for the BiD-AP scheme and eq 42 (together with the companion eqs 40 and 41) for the SiP-AP scheme, are modified upon using $|\mathbf{R}_L|$ as binding descriptor.

To simplify the notation, we define the distance between the origins of the L and R-frames as r, namely $r \equiv |\mathbf{R}_L|$. Without loss of generality, the origin of the L-frame, as well as that of the R-frame, can be an atom, the centroid of a subset of atoms or the center of mass. The distance r is the parameter taken to establish when the complex is or is not in place, according to the value of the indicator function $I(r)$, which can be 1 or 0. In principle, to apply this criterion, we need to define two threshold distances, say r_1 and r_2, such that $I(r) = 1$ if $r_1 < r < r_2$ and $I(r) = 0$ otherwise. However, as emerged from the previous discussion, the indicator function enters the double-decoupling method in no explicit way. This suggests that one may not need to define r_1 and r_2, provided that a "way" can be devised to sample most of the bound configurations during an equilibrium MD simulation. As already discussed in sections 3.1.1 and 3.1.2, for complexes with large binding constants, this "way" can be guaranteed from the equilibrium simulation itself, because the complex, owing to its stability, never dissociates during the simulation. Problems may instead occur when dealing with weakly bound complexes. These situations can be treated only introducing some external information on shape and size of the binding site, through a geometrical definition of $I(r)$, via hard or soft potential. Of course, in these restraining strategies, significant errors can be introduced, arising from an ill definition of the binding free energy basin. For this reason, the weaker the binding is, the greater is the error. In the limit case of an almost flat free energy binding basin, one has to resort to some arbitrary criterion to define $I(r)$, calling into play physical features of the complex, which do not include the mere energetical stability.

Considering that the coordinate r corresponds to the distance between the origins of the R- and L-frames, it is a natural choice to use spherical polar coordinates for representing $\mathbf{R}_L$, that is, $\mathbf{R}_L \equiv (r, \theta, \varphi)$, where θ is the angle between $\mathbf{R}_L$ and the z-axis of the R-frame and φ is the angle formed by the projection of $\mathbf{R}_L$ on the xy-plane of the R-frame and the x-axis of the same frame. Then, we make explicit the

DOI: 10.1021/acs.jctc.7b00594
J. Chem. Theory Comput. 2017, 13, 5874–5886

coordinates r, θ and φ into eq 29, expressing the indicator function $I(\mathbf{R}_L)$ in terms of the r coordinate:

$$g(\lambda) = -RT \ln\left(\int I(r) r^2 \sin\theta \, J_{\xi_L} \, e^{-\beta U(\lambda, \mathbf{R}_L, \xi_L, \mathbf{r}_L, \mathbf{r}_R, \mathbf{r}_S)} \, d\mathbf{R}_L \, d\xi_L \, d\mathbf{r}_L \, d\mathbf{r}_R d\mathbf{r}_S\right) \quad (52)$$

where $r^2 \sin\theta$ is the Jacobian determinant $J_{\mathbf{R}_L}$ and, for the sake of compactness, $\mathbf{R}_L \equiv (r, \theta, \varphi)$ and $d\mathbf{R}_L \equiv dr \, d\theta \, d\varphi$. The other symbols in eq 52 preserve their original meaning. Thus, the free energy difference $g(1) - g(0)$ of eq 30 becomes

$$\begin{aligned} & g(1) - g(0) \\ &= -RT \ln\left(\frac{\int I(r) r^2 \sin\theta \, J_{\xi_L} \, e^{-\beta U(\mathbf{r}_R, \mathbf{r}_S)} \, e^{-\beta U(\mathbf{r}_L)} \, d\mathbf{R}_L \, d\xi_L \, d\mathbf{r}_L \, d\mathbf{r}_R \, d\mathbf{r}_S}{\int I(r) r^2 \sin\theta \, J_{\xi_L} \, e^{-\beta U(\mathbf{R}_L, \xi_L, \mathbf{r}_L, \mathbf{r}_R, \mathbf{r}_S)} \, d\mathbf{R}_L \, d\xi_L \, d\mathbf{r}_L \, d\mathbf{r}_R \, d\mathbf{r}_S}\right) \\ &= -RT \ln\left(4\pi V_I \mathcal{V}_{\xi_L} \frac{Z_{N,R} Z_{0,L}}{Z_{N,RL}}\right) \end{aligned} \quad (53)$$

In the third line of the previous equation, the factor 4π arises from integration over θ and φ, the factor $\mathcal{V}_{\xi_L}$ (equal to $8\pi^2$, 4π, or 1 according to the structure of L) arises from integration over the orientational coordinates of L (i.e., ξ_L) and $V_I = \int I(r) r^2 \, dr$. The third line of eq 53 allows us to write ΔG_1° of eq 22 as (viz. eq 36)

$$\Delta G^\circ_1 = g(1) - g(0) - RT \ln\left(\frac{1}{4\pi V_I C^\circ} \frac{\sigma_{cp,L}\sigma_{cp,R}}{\sigma_{gas,L}\sigma_{sol,R}}\right) + P^\circ(\bar{V}_R - \bar{V}_{RL}) \quad (54)$$

This relationship allows an estimation of ΔG_1° through the BiD-AP scheme, as explained in section 3.1.1.

To adopt the SiP-AP scheme, we have to recognize that the unnormalized average probability of finding the ligand in a generic point at a distance r from the origin of the R-frame (for a given λ) corresponds, up to a multiplication factor, to the radial distribution function, which, in turn, equals the exponential of the potential of mean force, $e^{-\beta\phi(\lambda,r)}$:

$$e^{-\beta\phi(\lambda,r)} = \frac{1}{4\pi} \int \sin\theta \, J_{\xi_L} \, e^{-\beta U(\lambda, r, \theta, \phi, \xi_L, \mathbf{r}_L, \mathbf{r}_R, \mathbf{r}_S)} \, d\theta \, d\phi \, d\xi_L \, d\mathbf{r}_L \, d\mathbf{r}_R \, d\mathbf{r}_S \quad (55)$$

The quantity $4\pi r^2 e^{-\beta\phi(\lambda,r)} dr$ is therefore proportional to the probability of finding L into a spherical shell of radius r and thickness dr centered at the origin of the R-frame. According to the above definition of potential of mean force, the free energy function $g(\lambda)$ (eq 52) becomes

$$g(\lambda) = -RT \ln\left(\int I(r) 4\pi r^2 \, e^{-\beta\phi(\lambda,r)} \, dr\right) \quad (56)$$

Used in the second line of eq 53, the previous equation gives the free energy difference $g(1) - g(0)$

$$g(1) - g(0) = -RT \ln\left(\frac{4\pi V_I \, e^{-\beta\phi(1,r)}}{\int I(r) 4\pi r^2 \, e^{-\beta\phi(0,r)} \, dr}\right) \quad (57)$$

In this equation, integration over r is carried out because $\phi(1,r)$ does not depend on r (analogously to $\phi(1, \mathbf{R}_L)$ in eq 37). Using eq 57 in eq 54 yields

$$\Delta G_1^\circ = -RT \ln\left(\frac{\sigma_{cp,L}\sigma_{cp,R}}{C^\circ \sigma_{gas,L}\sigma_{sol,R}} \frac{e^{-\beta\phi(1,r)}}{\int I(r) 4\pi r^2 \, e^{-\beta\phi(0,r)} \, dr}\right) + P^\circ(\bar{V}_R - \bar{V}_{RL}) \quad (58)$$

As done in section 3.1.2, we introduce a reference configuration corresponding to $r = r'$, with r' being an arbitrary established value of r (in analogy with $\mathbf{R}'_L$ of section 3.1.2). This allows to write

$$\frac{e^{-\beta\phi(1,r)}}{\int I(r) 4\pi r^2 \, e^{-\beta\phi(0,r)} \, dr} = \frac{e^{\beta[\phi(0,r') - \phi(1,r')]} \, e^{-\beta\phi(0,r')}}{\int I(r) 4\pi r^2 \, e^{-\beta\phi(0,r)} \, dr} \quad (59)$$

where, the equality $\phi(1,r) = \phi(1,r')$ has been used.

According to the SiP-AP scheme, the free energy difference $\phi(1,r') - \phi(0,r')$ in the numerator of eq 59 is estimated by means of alchemical transformations. A number of initial microstates of the coupled system ($\lambda = 0$) are sampled at the fixed $r = r'$. Starting from these microstates, nonequilibrium trajectories are performed with an established time schedule for λ, from $\lambda = 0$ to $\lambda = 1$. The works computed from these trajectories via eq 35 are thus employed in the Jarzynski equality[44] (eq 34).

The remaining part of eq 59 is computed upon considering that it corresponds to the probability density of finding L in a generic point at the distance r' from the origin of the R-frame, once the complex is in a bound configuration, that is, $I(r) = 1$:

$$\rho(r') = \frac{e^{-\beta\phi(0,r')}}{\int I(r) 4\pi r^2 \, e^{-\beta\phi(0,r)} \, dr} = \frac{\Delta p(r')}{4\pi r'^2 \Delta r} \quad (60)$$

where $\Delta p(r')$ is the infinitesimal probability that L is found in a spherical shell of radius r' and volume $4\pi r'^2 \Delta r$ (the center being the reference R point) during an equilibrium MD simulation with the complex restrained in the bound state. This simulation can be carried out as explained in section 3.1.2 (see discussion of eq 41).

In summary, considering the introduction of a reference configuration (eq 59) and the definition of probability density (eq 60), ΔG_1° of eq 58 can be rewritten as

$$\Delta G_1^\circ = \phi(1, r') - \phi(0, r') - RT \ln\left(\rho(r') \frac{\sigma_{cp,L}\sigma_{cp,R}}{C^\circ \sigma_{gas,L}\sigma_{sol,R}}\right) + P^\circ(\bar{V}_R - \bar{V}_{RL}) \quad (61)$$

As discussed previously, the quantity $\phi(1,r') - \phi(0,r')$ is computed evaluating the difference of the potential of mean force of coupled and uncoupled states via nonequilibrium alchemical transformations by constraining the ligand−receptor distance r to the value of r'. The quantity $\rho(r')$ is computed from eq 60.

5. CONCLUDING REMARKS

The fast-switching decoupling method is a powerful technique to compute absolute binding free energies of ligand−receptor (RL) complexes. Compared to equilibrium alchemical approaches such as those based on the free energy perturbation method,[67,68] it has been shown[31] that nonequilibrium alchemical simulations may provide comparable or even better performances both in terms of precision and computer time. It is also worth noting that, in such techniques, the sampling for performing alchemical transformations is limited to the initial

DOI: 10.1021/acs.jctc.7b00594
J. Chem. Theory Comput. 2017, 13, 5874−5886

state, in which the RL system is fully coupled. In equilibrium approaches, correct sampling must instead be guaranteed also for each intermediate nonphysical state in which the alchemical path is divided. While sampling of the fully coupled system may somehow be verified on the basis of some preliminary (experimental) information of the system itself, in equilibrium methods the reliability in sampling the intermediate states can be verified with greater difficulties, because no information can be gained from any experimental technique. In the current implementations, fast-switching decoupling is applied without constraining the RL complex in the bound state.[30] Even if this has been revealed computationally effective,[31,43] a sound theoretical ground requires that the bound state of the complex is preserved during decoupling of the ligand from receptor and solvent. Here, we have addressed this issue supplementing the method with the possibility of performing alchemical trajectories with the ligand constrained to a fixed position relative to the receptor. Binded-domain alchemical-path (BiD-AP) and single-point alchemical-path (SiP-AP) schemes allow us to compute the decoupling free energy contribution to the absolute binding free energy without resorting to the explicit calculation of the orientational binding-site volume.[21] With respect to fast-switching decoupling without constraints,[30] BiD-AP prevents the ligand from leaving the binding site, but still requires an estimate of the positional binding-site volume. SiP-AP is an evolution of BiD-AP, in which a reference configuration of the RL complex is introduced to split the decoupling free energy of the ligand from solvent and receptor into two separate terms, one computed from an equilibrium MD simulation of the fully coupled bound state of the complex and the other from nonequilibrium alchemical transformations of the complex constrained in the reference configuration. The improvement with respect to the BiD-AP scheme is that the SiP-AP scheme allows to avoid the calculation of the positional binding-site volume. The price that we have to pay to obtain such an advantage is to carry out an additional equilibrium MD simulation to compute the probability density.

BiD-AP and SiP-AP techniques are based on a binding descriptor corresponding to the position of a reference atom of the ligand with respect to a given atom of the receptor. As shown, the two schemes can also be devised to employ the simple distance between the two atoms as binding descriptor. The drawback of such an approach is that one has to assume that complete orientational sampling of the ligand is attained during the equilibrium MD simulations of the bound RL complex, whether in the MD simulation performed to get the initial microstates of the alchemical trajectories or, in the case of the SiP-AP scheme, in the one aimed at computing $\rho(\mathbf{R}'_L)$. However, in most cases, specific ligand−receptor interactions make the ligand orientationally hindered within the binding site, namely the ligand makes librations around a well-defined free energy minimum. This supports the assumption that an exhaustive orientational sampling is reached during an equilibrium MD simulation of the complex. The fact that alternative orientational poses are possible for a complex, with comparable binding affinity, could introduce important errors in the methodology. Nonetheless, this type of problem is common to all the double-decoupling based methods. In such cases, one must introduce an a priori knowledge of possible poses of the ligand or to resort to some advanced sampling technique based, for example, on replica exchange or serial generalized ensemble schemes.[16,69−73] Techniques based on the freezing of atoms far from the binding site could also be implemented to speed up the nonequilibrium trajectories.[74−77]

A further important problem, intrinsically related to fast-switching, lies in the fact that only decoupling alchemical trajectories are employed or, using a term borrowed from nonequilibrium work theories,[57,58] that the fast-switching process is monodirectional. This leads to the use of the Jarzynski equality[44] as free energy estimator, which is known to provide biased free energy estimates so long as the decoupling simulation trajectories are fast. A possible way to reduce the bias and hence to improve the accuracy, is to increase the number of decoupling trajectories using, for example, non-equilibrium alchemical simulations powered with path-breaking based algorithms.[78,79]

Finally, it is worth noting that the introduction of constraints in the fast-switching decoupling method, as done with both BiD-AP and SiP-AP, makes the methodology suitable for being extended to bidirectional processes, namely to combine in a unified scheme[40,41] both decoupling and coupling trajectories.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.7b00594.

Symmetry numbers; extension of eq 27 of the main text to the NPT ensemble; decoupling the ligand from solvent and receptor: ΔG°_{1} calculation using translational and rotational constraints (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: riccardo.chelli@unifi.it.

ORCID

Piero Procacci: 0000-0003-2667-3847
Gianni Cardini: 0000-0002-7292-3555
Marco Pagliai: 0000-0003-0240-161X
Riccardo Chelli: 0000-0003-3492-5709

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Chipot, C., Pohorille, A., Eds. *Free Energy Calculations: Theory and Applications in Chemistry and Biology*; Springer: Berlin, 2007; Vol. *86*.

(2) Deng, Y.; Roux, B. Computations of standard binding free energies with molecular dynamics simulations. *J. Phys. Chem. B* **2009**, *113*, 2234−2246.

(3) De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of molecular dynamics and related methods in drug discovery. *J. Med. Chem.* **2016**, *59*, 4035−4061.

(4) Klimovich, P. V.; Shirts, M. R.; Mobley, D. L. Guidelines for the analysis of free energy calculations. *J. Comput.-Aided Mol. Des.* **2015**, *29*, 397−411.

(5) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150−160.

(6) Procacci, P. Alchemical determination of drug-receptor binding free energy: where we stand and where we could move to. *J. Mol. Graphics Modell.* **2017**, *71*, 233−241.

(7) Lavecchia, A.; Giovanni, C. D. Virtual screening strategies in drug discovery: a critical review. *Curr. Med. Chem.* **2013**, *20*, 2839−2860.

DOI: 10.1021/acs.jctc.7b00594
J. Chem. Theory Comput. 2017, 13, 5874−5886

(8) Shoichet, B. Virtual screening of chemical libraries. *Nature* **2004**, *432*, 862−865.
(9) Woo, H.-J.; Roux, B. Calculation of absolute protein-ligand binding free energy from computer simulation. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6825−6830.
(10) Colizzi, F.; Perozzo, R.; Scapozza, L.; Recanatini, M.; Cavalli, A. Single-molecule pulling simulations can discern active from inactive enzyme inhibitors. *J. Am. Chem. Soc.* **2010**, *132*, 7361−7371.
(11) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.
(12) Gallicchio, E.; Levy, R. M.; Lapelosa, M. Binding energy distribution analysis method (BEDAM) for estimation of protein-ligand binding affinities. *J. Chem. Theory Comput.* **2010**, *6*, 2961−2977.
(13) Fasnacht, M.; Swendsen, R. H.; Rosenberg, J. M. Adaptive integration method for Monte Carlo simulations. *Phys. Rev. E* **2004**, *69*, 056704.
(14) Chelli, R.; Procacci, P. A potential of mean force estimator based on nonequilibrium work exponential averages. *Phys. Chem. Chem. Phys.* **2009**, *11*, 1152−1158.
(15) Nicolini, P.; Frezzato, D.; Gellini, C.; Bizzarri, M.; Chelli, R. Toward quantitative estimates of binding affinities for protein-ligand systems involving large inhibitor compounds: a steered molecular dynamics simulation route. *J. Comput. Chem.* **2013**, *34*, 1561−1576.
(16) Procacci, P.; Bizzarri, M.; Marsili, S. Energy-driven undocking (EDU-HREM) in solute tempering replica exchange simulations. *J. Chem. Theory Comput.* **2014**, *10*, 439−450.
(17) Sotomayor, M.; Schulten, K. Single-molecule experiments in vitro and in silico. *Science* **2007**, *316*, 1144−1148.
(18) Vashisth, H.; Abrams, C. F. Ligand escape pathways and (un)binding free energy calculations for the hexameric insulin-phenol complex. *Biophys. J.* **2008**, *95*, 4193−4204.
(19) Ytreberg, F. M. Absolute FKBP binding affinities obtained via nonequilibrium unbinding simulations. *J. Chem. Phys.* **2009**, *130*, 164906.
(20) Shen, J.; Li, W.; Liu, G.; Tang, Y.; Jiang, H. Computational insights into the mechanism of ligand unbinding and selectivity of estrogen receptors. *J. Phys. Chem. B* **2009**, *113*, 10436−10444.
(21) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J.* **1997**, *72*, 1047−1069.
(22) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc.* **2015**, *137*, 2695−2703.
(23) Hansen, N.; van Gunsteren, W. F. Practical aspects of free-energy calculations: a review. *J. Chem. Theory Comput.* **2014**, *10*, 2632−2647.
(24) Gallicchio, E.; Levy, R. M. Advances in all atom sampling methods for modeling protein-ligand binding affinities. *Curr. Opin. Struct. Biol.* **2011**, *21*, 161−166.
(25) Jorgensen, W. L.; Thomas, L. L. Perspective on free-energy perturbation calculations for chemical equilibria. *J. Chem. Theory Comput.* **2008**, *4*, 869−876.
(26) Shirts, M. R.; Mobley, D. L.; Chodera, J. D. Alchemical free energy calculations: ready for prime time? *Annu. Rep. Comput. Chem.* **2007**, *3*, 41−59.
(27) Jorgensen, W. L.; Ravimohan, C. Monte Carlo simulation of differences in free energies of hydration. *J. Chem. Phys.* **1985**, *83*, 3050−3054.
(28) Gumbart, J. C.; Roux, B.; Chipot, C. Standard binding free energies from computer simulations: what is the best strategy? *J. Chem. Theory Comput.* **2013**, *9*, 794−802.
(29) Procacci, P.; Chelli, R. Statistical mechanics of ligand-receptor noncovalent association, revisited: binding site and standard state volumes in modern alchemical theories. *J. Chem. Theory Comput.* **2017**, *13*, 1924−1933.
(30) Procacci, P. I. Dissociation free energies of drug-receptor systems via non-equilibrium alchemical simulations: a theoretical framework. *Phys. Chem. Chem. Phys.* **2016**, *18*, 14991−15004.
(31) Nerattini, F.; Chelli, R.; Procacci, P., II. Dissociation free energies in drug-receptor systems via nonequilibrium alchemical simulations: application to the FK506-related immunophilin ligands. *Phys. Chem. Chem. Phys.* **2016**, *18*, 15005−15018.
(32) Chelli, R. Local sampling in steered Monte Carlo simulations decreases dissipation and enhances free energy estimates via nonequilibrium work theorems. *J. Chem. Theory Comput.* **2012**, *8*, 4040−4052.
(33) Cossins, B. P.; Foucher, S.; Edge, C. M.; Essex, J. W. Assessment of nonequilibrium free energy methods. *J. Phys. Chem. B* **2009**, *113*, 5508−5519.
(34) Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; Tirado-Rives, J. Efficient computation of absolute free energies of binding by computer simulations. Application to methane dimer in water. *J. Chem. Phys.* **1988**, *89*, 3742−3746.
(35) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute binding free energies: a quantitative approach for their calculation. *J. Phys. Chem. B* **2003**, *107*, 9535−9551.
(36) Deng, Y.; Roux, B. Calculation of standard binding free energies: aromatic molecules in the T4 lysozyme L99A mutant. *J. Chem. Theory Comput.* **2006**, *2*, 1255−1273.
(37) Procacci, P.; Cardelli, C. Fast switching alchemical transformations in molecular dynamics simulations. *J. Chem. Theory Comput.* **2014**, *10*, 2813−2823.
(38) Kirkwood, J. G. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **1935**, *3*, 300−313.
(39) Zwanzig, R. W. Hightemperature equation of state by a perturbation method. I. Nonpolar gases. *J. Chem. Phys.* **1954**, *22*, 1420−1426.
(40) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245−268.
(41) Shirts, M. R.; Bair, E.; Hooker, G.; Pande, V. S. Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods. *Phys. Rev. Lett.* **2003**, *91*, 140601.
(42) General, I. J. A note on the standard state's binding free energy. *J. Chem. Theory Comput.* **2010**, *6*, 2520−2524.
(43) Sandberg, R. B.; Banchelli, M.; Guardiani, C.; Menichetti, S.; Caminati, G.; Procacci, P. Efficient nonequilibrium method for binding free energy calculations in molecular dynamics simulations. *J. Chem. Theory Comput.* **2015**, *11*, 423−435.
(44) Jarzynski, C. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* **1997**, *78*, 2690−2693.
(45) Jayachandran, G.; Shirts, M. R.; Park, S.; Pande, V. S. Parallelized-over-parts computation of absolute binding free energy with docking and molecular dynamics. *J. Chem. Phys.* **2006**, *125*, 084901.
(46) Giovannelli, E.; Cardini, G.; Volkov, V.; Chelli, R. Nonequilibrium work theorems applied to transitions between configurational domains. *J. Stat. Mech.: Theory Exp.* **2016**, *2016*, 123204.
(47) Giovannelli, E.; Cioni, M.; Procacci, P.; Cardini, G.; Pagliai, M.; Volkov, V.; Chelli, R. Binding free energies of host-guest systems by nonequilibrium alchemical simulations with constrained dynamics: illustrative calculations and numerical validation. *J. Chem. Theory Comput.* **2017**, xx DOI: 10.1021/acs.jctc.7b00595.
(48) Glasstone, S. *Thermodynamics for chemists*; Van Nostrand: New York, 1947; p 351.
(49) Lewis, G.; Randall, M. *Thermodynamics*, 2nd ed.; Pitzer, K. S., Brewer, L., Eds.; McGraw-Hill, 1961; pp 244−249.
(50) Hill, T. L. *Cooperativity Theory in Biochemistry*, 1st ed.; Springer Series in Molecular Biology; Springer-Verlag: New York, 1985; p 8.
(51) Ben-Naim, A. *Statistical Thermodynamics for Chemists and Biochemists*; Plenum Press: New York, 1992; p 677.
(52) McQuarrie, D. A. *Statistical mechanics*; Harper & Row: New York, 1973.

DOI: 10.1021/acs.jctc.7b00594
J. Chem. Theory Comput. 2017, 13, 5874−5886

(53) Chandler, D.; Pratt, L. R. Statistical mechanics of chemical equilibria and intramolecular structures of nonrigid molecules in condensed phases. *J. Chem. Phys.* **1976**, *65*, 2925−2940.
(54) We notice that in ref 21 the symmetry number of the complex is indicated as σ_{AB}.
(55) The potential of mean force of L (eq 31) is independent of $\mathbf{R}_L$ when L is decoupled from the rest of the system ($\lambda = 1$); combine eqs. 24 and 31.
(56) Jarzynski, C. Equilibrium free-energy differences from non-equilibrium measurements: a master-equation approach. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1997**, *56*, 5018−5035.
(57) Crooks, G. E. Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems. *J. Stat. Phys.* **1998**, *90*, 1481−1487.
(58) Crooks, G. E. Path-ensemble averages in systems driven far from equilibrium. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **2000**, *61*, 2361−2366.
(59) Andersen, H. C. RATTLE: a velocity version of the SHAKE algorithm for molecular dynamics calculations. *J. Comput. Phys.* **1983**, *52*, 24−34.
(60) 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*, 327−341.
(61) General, I. J.; Meirovitch, H. Absolute free energy of binding and entropy of the FKBP12-FK506 complex: effects of the force field. *J. Chem. Theory Comput.* **2013**, *9*, 4609−4619.
(62) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.
(63) Hermans, J.; Shankar, S. The free energy of xenon binding to myoglobin from molecular dynamics simulation. *Isr. J. Chem.* **1986**, *27*, 225−227.
(64) Roux, B.; Nina, M.; Pomes, R.; Smith, J. C. Thermodynamic stability of water molecules in the bacteriorhodopsin proton channel: a molecular dynamics free energy perturbation study. *Biophys. J.* **1996**, *71*, 670−681.
(65) Zhang, L.; Hermans, J. Hydrophilicity of cavities in proteins. *Proteins: Struct., Funct., Genet.* **1996**, *24*, 433−438.
(66) With respect to eq 35, the energy takes the form $U(\lambda, \mathbf{r}_L, \mathbf{r}_S)$, that is it does not include the coordinates of R and is independent of the external coordinates of L.
(67) Fujitani, H.; Tanida, Y.; Matsuura, A. Massively parallel computation of absolute binding free energy with well-equilibrated states. *Phys. Rev. E* **2009**, *79*, 021914.
(68) Fujitani, H.; Tanida, Y.; Ito, M.; Jayachandran, G.; Snow, C. D.; Shirts, M. R.; Sorin, E. J.; Pande, V. S. Direct calculation of the binding free energies of FKBP ligands. *J. Chem. Phys.* **2005**, *123*, 084108.
(69) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(70) Swendsen, R. H.; Wang, J.-S. Replica Monte Carlo simulation of spin-glasses. *Phys. Rev. Lett.* **1986**, *57*, 2607−2609.
(71) Marinari, E.; Parisi, G. Simulated tempering: a new Monte Carlo scheme. *Europhys. Lett.* **1992**, *19*, 451−458.
(72) Chelli, R. Optimal weights in serial generalized-ensemble simulations. *J. Chem. Theory Comput.* **2010**, *6*, 1935−1950.
(73) Chelli, R.; Signorini, G. F. Serial generalized ensemble simulations of biomolecules with self-consistent determination of weights. *J. Chem. Theory Comput.* **2012**, *8*, 830−842.
(74) Nicolini, P.; Chelli, R. Improving fast-switching free energy estimates by dynamical freezing. *Phys. Rev. E* **2009**, *80*, 041124.
(75) Nicolini, P.; Frezzato, D.; Chelli, R. Exploiting configurational freezing in nonequilibrium Monte Carlo simulations. *J. Chem. Theory Comput.* **2011**, *7*, 582−593.
(76) Chelli, R. Local sampling in steered Monte Carlo simulations decreases dissipation and enhances free energy estimates via nonequilibrium work theorems. *J. Chem. Theory Comput.* **2012**, *8*, 4040−4052.
(77) Giovannelli, E.; Cardini, G.; Chelli, R. Elastic barrier dynamical freezing in free energy calculations: a way to speed up nonequilibrium molecular dynamics simulations by orders of magnitude. *J. Chem. Theory Comput.* **2016**, *12*, 1029−1039.
(78) Chelli, R.; Gellini, C.; Pietraperzia, G.; Giovannelli, E.; Cardini, G. Path-breaking schemes for nonequilibrium free energy calculations. *J. Chem. Phys.* **2013**, *138*, 214109.
(79) Giovannelli, E.; Gellini, C.; Pietraperzia, G.; Cardini, G.; Chelli, R. Combining path-breaking with bidirectional nonequilibrium simulations to improve efficiency in free energy calculations. *J. Chem. Phys.* **2014**, *140*, 064104.

DOI: 10.1021/acs.jctc.7b00594
J. Chem. Theory Comput. 2017, 13, 5874−5886