Alchemical Transformations for Concerted Hydration Free Energy Estimation with Explicit Solvation

Sheenam Khuttan*,[1, a)] Solmaz Azimi*,[1, b)] Joe Z. Wu,[1, c)] and Emilio Gallicchio[1, d)]
Department of Chemistry, Brooklyn College of the City University of New York, New York, NY.
**These authors contributed equally to this work.*

(Dated: 9 November 2020)

We present a family of alchemical perturbation potentials that enable the calculation of hydration free energies of small to medium-sized molecules in a single concerted alchemical coupling step instead of the commonly used sequence of two distinct coupling steps for Lennard-Jones and electrostatic interactions. The perturbation potentials we employ are non-linear functions of the solute-solvent interaction energy designed to focus sampling near entropic bottlenecks along the alchemical pathway. We present a general framework to optimize the parameters of alchemical perturbation potentials of this kind. The optimization procedure is based on the λ-function formalism and the maximum-likelihood parameter estimation procedure we developed earlier to avoid the occurrence of multi-modal distributions of the coupling energy along the alchemical path. A novel soft-core function applied to the overall solute-solvent interaction energy rather than individual interatomic pair potentials critical for this result is also presented. Because it does not require modifications of core force and energy routines, the soft-core formulation can be easily deployed in molecular dynamics simulation codes. We illustrate the method by applying it to the estimation of the hydration free energy in water droplets of compounds of varying size and complexity. In each case, we show that convergence of the hydration free energy is achieved rapidly. This work paves the way for the ongoing development of more streamlined algorithms to estimate free energies of molecular binding with explicit solvation.

I. INTRODUCTION

The hydration free energy of a compound is defined as the reversible work for transferring one molecule of the compound from the gas phase to the water phase.[1] The hydration free energy is an important characteristic of a substance. For instance, it is one of the determinant factors of the water solubility of drug formulations[2] and of the binding affinity of an inhibitor towards a protein receptor.[3] Hydration free energies are most commonly derived from Henry's constant measurements.[4]

Free energies of hydration can also be estimated computationally.[5] Most commonly, this is accomplished by molecular simulations of alchemical transformations in which solute-solvent interactions are progressively turned on. The nature of the alchemical transformation is critical to obtaining reliable results. A single direct alchemical process in which the coupling between the solute and the solvent is increased in a simple linear fashion has been found to be problematic for anything other than the smallest solutes (monoatomic atoms and ions, water, methane, and similar).[6] One issue is the singularity of the derivative of the alchemical potential with respect to the parameter λ near the decoupled state.[7] Simple approaches of this kind can also be found to converge slowly due to bottlenecks along the alchemical path that are caused by poorly sampled conformational equilibria.[8]

These issues have been the subject of intense studies. This collective effort has resulted in a set of recommended best practices for alchemical calculations commonly employed by the computational chemistry community that are presently implemented in popular molecular simulation software packages.[5,9,10] For example, it is generally recommended to split the coupling of the solute and the solvent into two phases, the first in which the volume-exclusion core repulsion and dispersion interactions are turned on, followed by a second phase in which electrostatic interactions are turned on. Separation-shifted soft-core pair potentials[11] are commonly introduced to avoid end-point singularities, especially when volume-exclusion core repulsion terms are introduced. In addition, it is often necessary for large solutes to introduce one small group of atoms at a time or to resort to bond-growing processes in which solute atoms are pushed out from a central point rather than directly created in the solvent.[12]

While successful, these strategies add layers of complexity to hydration free energy calculations.[13] Splitting the coupling process into two calculation phases requires specifying two alchemical schedules that are often very different from each other. This practice also relies on simulating a nonphysical intermediate state consisting of the uncharged solute in water which may have very different conformational propensities than the actual solute. The unphysical uncharged solute may, for example, undergo hydrophobic collapse and revert slowly to the physical extended hydrated state.

Software implementations of soft-core potential functions are cumbersome and difficult to maintain, and they require the optimization of specific parameters for each interaction

[a)]Ph.D. Program in Biochemistry, The Graduate Center of the City University of New York, New York, NY
[b)]Ph.D. Program in Biochemistry, The Graduate Center of the City University of New York, New York, NY; These two authors contributed equally to the work
[c)]Ph.D. Program in Chemistry, The Graduate Center of the City University of New York, New York, NY
[d)]Ph.D. Program in Biochemistry, The Graduate Center of the City University of New York, New York, NY; Ph.D. Program in Chemistry, The Graduate Center of the City University of New York, New York, NY; Corresponding author: egallicchio@brooklyn.cuny.edu

potential type.[13] Conventional separation-shifted scaled soft-core pair potentials[11] also lead to non-linear and non-algebraic forms of the alchemical potential that require a cumbersome energy rescoring process for the estimation of free energy changes with thermodynamic reweighting.[14,15]

Thus, it would be beneficial to develop more streamlined alchemical protocols that (i) compute the hydration free energy in one continuous transformation process, (ii) do not require modifications of the standard form of pair potential functions, and (iii) can be parameterized using a linear progression of the charging parameter λ, while (iv) preserving or improving the rate of equilibration and convergence of free energy estimates relative to mainstream protocols. Here, we present a method with potentially all of these characteristics, while being applicable to a range of alchemical transformations, including those used in structure-based drug design[16], which is our primary focus.[17,18]

The approach presented here is based on the work we carried out recently to optimize binding free energy calculations in the context of implicit solvation.[19,20] The previous work yielded a family of optimized alchemical potential energy functions that modify all interactions in one concerted step and that do not require customized soft-core pair potentials. The effort also resulted in the development of a general framework, based on generalized non-Boltzmann sampling theory[21] and maximum likelihood estimation,[22] for the optimization of the parameters of the alchemical potentials to accelerate conformational mixing and convergence.

In the present work, we demonstrate that the same approach applies to the estimation of the hydration free energies of small to medium-sized molecules using an explicit representation of the solvent. As in our previous work,[20] the approach is based on the observation that the slow convergence of free energies is due to rare conformational transitions caused by entropic bottlenecks akin to first-order phase transitions,[23] and that these can be circumvented by an appropriate choice of the alchemical perturbation energy function.

This work is organized as follows. We first review the theoretical and computational protocol developed in the previous work. Then we apply it to a model system of hydration in which four diverse solutes of a wide range of sizes and hydrophobic characters (ethanol, alanine dipeptide, 1-naphthol, and 3,4-diphenyltoluene) are transferred from the gas phase to a water droplet. We show that the optimized protocols yield results that are consistent with more conventional approaches when they both reach convergence. However, for larger solutes, only the optimized protocol can achieve rapid and reliable convergence. These promising outcomes pave the way for a novel generation of more streamlined methodologies for alchemical transformations in condensed phases.

II. THEORY AND METHODS

A. Alchemical Transformations for the Estimation of Hydration Free Energies

Alchemical transformations are based on an alchemical potential energy function that interpolates from the potential energy function U_0 of the starting state to that of the final state U_1 as the progress parameter λ goes, conventionally, from 0 to 1. The most straightforward approach is a linear interpolating function of the form:

$$U_\lambda(x) = U_0(x) + \lambda\,[U_1(x) - U_0(x)] = U_0(x) + \lambda u(x) \qquad (1)$$

where x represents the degrees of freedom of the system and we have introduced the perturbation function

$$u(x) = U_1(x) - U_0(x) \qquad (2)$$

which in the case of hydration corresponds to the solute-solvent interaction energy.

While straightforward, the linear alchemical energy function (1) leads to instabilities and poor convergence, especially when volume-exclusion terms are introduced.[13] To address this issue we consider the family of alchemical potential energy functions of the form:

$$U_\lambda(x) = U_0(x) + W_\lambda[u_{\rm sc}(x)], \qquad (3)$$

where $u_{\rm sc}(x)$, defined below, is a soft core-modified solute-solvent interaction energy, and $W_\lambda(u_{\rm sc})$ is an alchemical perturbation function defined such that $W_0(u_{\rm sc}) = 0$ and $W_1(u_{\rm sc}) = u_{\rm sc}$ at $\lambda = 0$ and $\lambda = 1$, respectively. We will consider in this work the linear function

$$W_\lambda(u_{\rm sc}) = \lambda u_{\rm sc} \qquad (4)$$

and the generalized softplus function

$$W_\lambda(u_{\rm sc}) = \frac{\lambda_2 - \lambda_1}{\alpha} \ln\left[1 + e^{-\alpha(u_{\rm sc} - u_0)}\right] + \lambda_2 u_{\rm sc} + w_0, \qquad (5)$$

where the parameters λ_2, λ_1, α, u_0, and w_0 are functions of λ. The derivative of the softplus function is the generalized logistic function (Fermi's function):

$$\frac{\partial W_\lambda(u_{\rm sc})}{\partial u_{\rm sc}} = \frac{\lambda_2 - \lambda_1}{1 + e^{-\alpha(u_{\rm sc} - u_0)}} + \lambda_1\,. \qquad (6)$$

The softplus perturbation function's parameters are optimized using the procedure described in reference 20 and briefly described later in this Section.

Note that the perturbation functions $W_\lambda(u)$ that we are considering are algebraic functions of only λ and the solute-solvent interaction energy u, which itself does not depend on λ. This characteristic makes them unlike alchemical energy functions with soft-core pair potentials whose solute-solvent interaction energy function also depends on λ implicitly (and non-algebraically) through the λ-dependent scaled distance function.[24] The lack of λ-dependence in the solute-solvent interaction energy is a critical assumption of the alchemical theory employed here. This feature also simplifies the process of

the calculation of the free energy change by thermodynamic reweighting,[15,25] since it does not require calling the molecular mechanics engine to recalculate the perturbation energy of sampled configurations at each λ-state.

In this work, we employ the following definition of the soft-core solute-solvent interaction energy:[20]

$$u_{\rm sc}(u) = \begin{cases} u & u \le 0 \\ u_{\rm max} f_{\rm sc}(u/u_{\rm max}) & u > 0 \end{cases} \tag{7}$$

and

$$f_{\rm sc}(y) = \frac{z^a - 1}{z^a + 1}, \tag{8}$$

where $u_{\rm max} > 0$ is the maximum allowed value of the soft-core solute-solvent interaction energy, $z = 1 + 2y/a + 2(y/a)^2$ and a is an adjustable dimensionless exponent (here $u_{\rm max} = 50$ kcal/mol and $a = 1/16$). With these definitions, $u_{\rm sc}(u)$ is a $C(2)$-smooth one-to-one map from the original solute-solvent interaction energy to the soft-core interaction energy. For favorable solute-solvent interaction energies ($u < 0$), the original and soft-core solute-solvent interaction energies are the same. However, for unfavorable solute-solvent interaction energies ($u > 0$), such as when two atoms clash as they approach each other, the soft-core interaction energy $u_{\rm sc}$ grows less rapidly than u, and it eventually reaches a maximum plateau, unlike the solute-solvent interaction energy that grows indefinitely.

We stress that in this soft-core methodology, there are no soft-core modifications of pair potentials. The soft-core function is applied to the overall solute-solvent interaction energy u evaluated with the standard form of the Coulomb and Lennard-Jones interatomic potentials without soft-core modifications.

To obtain the hydration free energy, a set of samples of the soft-core solute-solvent interaction energies, $u_{\rm sc}(i)$, are collected during molecular dynamics simulations performed at a sequence of λ values between 0 and 1. The free energy profile as a function of λ, $\Delta G(\lambda)$, is obtained by multi-state reweighting[15] using the UWHAM method.[25] The hydration free energy, ΔG_h^* in the Ben-Naim standard state[26] is by definition the value of free energy profile at $\lambda = 1$. In this notation, $\Delta G_h^* = \Delta G(1)$.

B. Analytical Theory of Alchemical Transformations

The hydration process is analyzed and optimized using the theory we recently developed for alchemical potential energy functions of the form in Eq. (3).[22] Briefly, following the potential distribution theorem,[27] here we consider $p_0(u_{\rm sc})$, the probability density of the soft-core solute-solvent interaction energy $u_{\rm sc}$ in the ensemble in which solute and solvent are uncoupled ($\lambda = 0$). All other quantities of the alchemical transformation can be obtained from $p_0(u_{\rm sc})$.[28,29] In particular, given $p_0(u_{\rm sc})$, the probability density for the binding energy $u_{\rm sc}$ for the state with perturbation potential $W_\lambda(u_{\rm sc})$, is

$$p_\lambda(u_{\rm sc}) = \frac{1}{K(\lambda)} p_0(u_{\rm sc}) \exp\left[-\beta W_\lambda(u_{\rm sc})\right], \tag{9}$$

where $\beta = 1/k_B T$,

$$K(\lambda) = \int_{-\infty}^{+\infty} p_0(u_{\rm sc}) \exp\left[-\beta W_\lambda(u_{\rm sc})\right] du_{\rm sc} = \langle \exp\left[-\beta W_\lambda(u_{\rm sc})\right] \rangle_{\lambda=0} \tag{10}$$

is the excess component of the equilibrium constant for binding and

$$\Delta G(\lambda) = -\frac{1}{\beta} \ln K(\lambda) \tag{11}$$

is the corresponding free energy profile. Note that for a linear perturbation potential, $W_\lambda(u_{\rm sc}) = \lambda u_{\rm sc}$, Eqs. (10) and (11) state that the free energy profile is related to the double-sided Laplace transform of $p_\lambda(u_{\rm sc})$.

An analytical description of $p_0(u_{\rm sc})$, and thus of all the quantities derived from it, is available. Briefly, the theory is based on the assumption that the statistics of the solute-solvent interaction energy u in the decoupled state is the convolution of two processes: one that is described by the sum of many "soft" background solute-solvent interactions, which follows central limit statistics, and another process that describes "hard" atomic collisions, which follows max statistics. (See reference 22 for the full derivation.) The probability density $p_0(u)$ is further expressed as the superposition of probability densities of a small number of modes

$$p_0(u) = \sum_i c_i p_{0i}(u), \tag{12}$$

where c_i are adjustable statistical weights that sum to 1 and $p_{0i}(u)$ is the probability density that corresponds to mode i. The probability density for mode i, $p_{0i}(u)$, is described analytically as:

$$\begin{aligned} p_{0i}(u) &= p_{bi}\, g(u; \bar{u}_{bi}, \sigma_{bi}) \\ &+ (1 - p_{bi}) \int_0^{+\infty} p_{WCA}(u'; n_{li}, \varepsilon_i, \tilde{u}_i)\, g(u - u'; \bar{u}_{bi}, \sigma_{bi})\, du' \end{aligned} \tag{13}$$

where $g(u; \bar{u}, \sigma)$ is the normalized Gaussian density function of mean $\bar{u}$, the standard deviation is σ, and

$$p_{WCA}(u; n_l, \varepsilon, \tilde{u}) = n_l \left[1 - \frac{(1+x_C)^{1/2}}{(1+x)^{1/2}}\right]^{n_l - 1} \frac{H(u)}{4\varepsilon_{LJ}} \frac{(1+x_C)^{1/2}}{x(1+x)^{3/2}}, \tag{14}$$

where $x = \sqrt{u/\varepsilon + \tilde{u}/\varepsilon + 1}$ and $x_C = \sqrt{\tilde{u}/\varepsilon + 1}$. (See reference 22 and appendix A of reference 20 for the complete derivation.) The model for each mode i depends on a number of adjustable parameters that correspond to the following physical quantities[22]:

- c_i: statistical weight of mode i
- p_{bi}: probability that no atomic clashes occur while in mode i
- $\bar{u}_{bi}$: the average background interaction energy of mode i
- σ_{bi}: the standard deviation of background interaction energy of mode i

- n_{li}: the effective number of statistical uncorrelated atoms of the solute in mode i
- ε_i: the effective ε parameter of a hypothetical Lennard-Jones interaction energy potential that describes the solute-solvent interaction energy in mode i
- $\tilde{u}_i$: the solute-solvent interaction energy value above which the collisional energy is not zero in mode i

Together with the weights c_i, the parameters above are varied using a maximum likelihood criterion to fit the distributions of the soft-core solute-solvent interaction energy obtained from numerical simulations.[22] The distribution of the soft-core solute-solvent interaction energy $p_0(u_{sc})$ is obtained from $p_0(u)$ using the standard formula for the change of random variable.[20]

In this work, we use the analytical theory above and the results of trial alchemical simulations to optimize the parameters of the softplus alchemical perturbation potential in Eq. (5).[20] The procedure consists of running trial alchemical calculations using the the linear alchemical potential $W_\lambda(u_{sc}) = \lambda u_{sc}$, to obtain the set of samples of the soft-core binding energies as a function of λ. The samples obtained from the trial simulation are then used to derive optimized parameters for the analytical model for $p_0(u_{sc})$ [Eqs. (12) to (14)] using a maximum likelihood approach.[30] We developed an application based on Tensorflow[31] to conduct the maximum likelihood optimization (`https://github.com/egallicc/femodel-tf-optimizer`).

The analytical representation of $p_0(u_{sc})$ is then analytically differentiated with respect to u_{sc} to obtain the λ-function defined as[20]

$$\lambda_0(u_{sc}) \equiv \frac{1}{\beta}\frac{\partial \ln p_0(u_{sc})}{\partial u_{sc}}. \tag{15}$$

Minima and maxima of $p_\lambda(u_{sc})$ occur when the λ-function and $\partial W_\lambda(u_{sc})/\partial u_{sc}$ intersect[20]

$$\lambda_0(u_{sc}) = \frac{\partial W_\lambda(u_{sc})}{\partial u_{sc}}. \tag{16}$$

Hence, the λ-function can be used as a guide to design alchemical potentials that avoid the occurrence of multi-modal distributions that are difficult to converge. The linear alchemical potential $W_\lambda(u_{sc}) = \lambda u_{sc}$ leads to $\partial W_\lambda(u_{sc})/\partial u_{sc} = \lambda$, which is a constant, and in many cases intersects the λ-function at multiple points (see Figure 4). To avoid multi-modal distributions, here we use the softplus alchemical potential and we vary the parameters λ_2, λ_1, α, u_0, and w_0 as a function of λ, such that the derivative of the softplus function in Eq. (6) intersects $\lambda_0(u_{sc})$ at a single point at each λ or, when this is not easily achievable, such that it at least intersects it at nearby points. As discussed in reference 20, this procedure removes or reduces the severity of entropic sampling bottlenecks during the alchemical coupling process and enhances conformational sampling efficiency in order to achieve convergence of the binding free energy estimates.

C. Computational Details

In this work, we employ a hydration model in which solutes are transferred from vacuum to near the center of mass of a water droplet that is about 27 Å in diameter and that is composed of 357 TIP3P water molecules (Figure 1). The droplet is confined in a spherical region defined by a flat-bottom harmonic restraining potential centered at the origin and acting on each TIP3P water oxygen atom. The flat-bottom potential tolerance was set to 24 Å which results in a confinement region of approximately twice the droplet volume, and a force constant of 5 kcal/(mol Å^2) beyond this tolerance. This simplified hydration model is selected here to illustrate and assess the proposed methodology using our existing Single Decoupling Method (SDM) plugin in OpenMM,[20,32] which was originally designed for binding free energy calculations with implicit solvation. The droplet model adopted here also avoids potential issues not relevant to this work that concern the correctness of implementations of alchemical processes with periodic boundary conditions and long-range electrostatics.[33,34]

All solutes were prepared with Maestro (Schrödinger, LLC) and GAFF/AM1-BCC force field parameters were assigned using the Antechamber program.[35] Single Decoupling alchemical calculations were prepared using the SDM workflow (`github.com/egallicc/openmm_sdm_worflow.git`) using 16 λ states. The MD calculations employed the OpenMM[32] MD engine and the SDM integrator plugins (`github.com/rajatkrpal/openmm_sdm_plugin.git`) using the OpenCL platform. The ASyncRE software,[36] customized for OpenMM and SDM (`github.com/egallicc/-async_re-openmm.git`), was used for the Hamiltonian Replica Exchange in λ space with a uniform λ schedule between 0 and 1. The λ-dependent parameters that were used with the softplus potential are listed in Tables III to VI. Molecular dynamics runs were conducted for a minimum of 5 ns per replica with a 1 fs time-step at 300 K, such that λ values were exchanged approximately every 5 ps. A Langevin thermostat at 300 K with a relaxation time constant of 20 ps was used. Binding energy samples and trajectory frames were recorded every 5 ps. The calculations were performed on the XSEDE Comet GPU HPC cluster at the San Diego Supercomputing Center.

To accelerate conformational sampling, in this work we employ the Hamiltonian Replica Exchange algorithm[37–40] in alchemical space.[28,41,42] For the calculations reported here, we have employed the asynchronous implementation of Replica Exchange (ASyncRE)[36] with the Gibbs Independence Sampling algorithm for state reassignments.[20]

III. RESULTS

A. Probability Distributions of the Solute-Solvent Interaction Energy with the Linear Alchemical Potential

As discussed in Theory and Methods, the free energy profile is determined by the probability distributions of the perturbation energy along the alchemical path. The probability

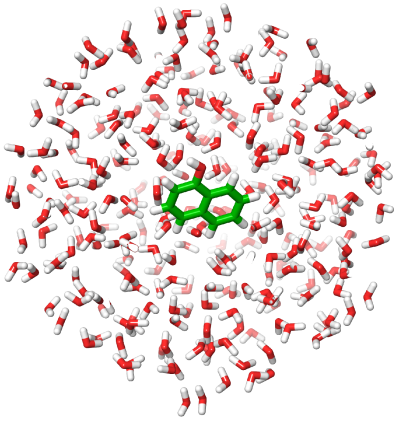

FIG. 1. Illustration of 1-naphthol (green carbon atoms) inserted at the center of a droplet of water. Water molecules that are obscuring the solute are not shown for clarity.

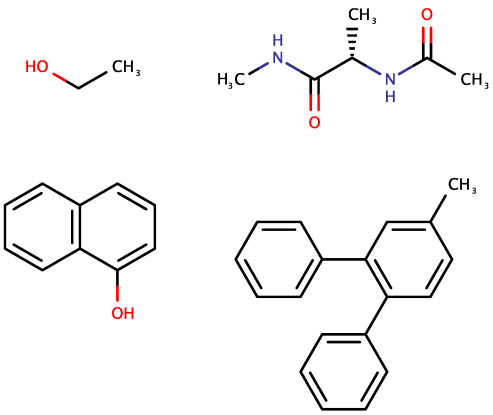

FIG. 2. Chemical structures of the solutes considered in this work. Clockwise starting from the upper left: ethanol, alanine dipeptide, 1-naphthol, and 3,4-diphenyltoluene. This series is ordered here and elsewhere by increasing bulkiness and hydrophobicity.

distributions $p_\lambda(u_{sc})$ for selected values of λ that were obtained from the concerted alchemical hydration with the linear alchemical potential are shown in Figure 3 (dots). At small λ values, the distributions are clustered around large and unfavorable values of the solute-solvent interaction energy, reflecting the occurrence of frequent atomic clashes when the solute and solvent are weakly coupled. Conversely, at values of λ near 1, the solute and solvent are more strongly coupled, and the solute-solvent interaction energies tend to be favorable. As λ increases, however, the distributions do not uniformly move towards lower values of u_{sc}. Instead, they become bimodal at intermediate values of λ with a trough near $u_{sc} = 0$ kcal/mol separating the weakly decoupled and strongly coupled states. The conversion from weakly coupled to strongly coupled behavior occurs by the transfer of population between these two limiting states, rather than by the formation of states with intermediate values of the interaction energy. This behavior is the opposite of what is expected for linear response behavior.[43] Rather, it is the hallmark of the occurrence of a strong phase transition.[20,23]

The transition from the weakly coupled to the strongly coupled states occurs gradually in the case of the concerted alchemical linear process for ethanol, the smallest solute considered. For this solute, the weakly coupled and strongly coupled modes partially overlap at $\lambda = 0.4$ (Figure 3, ethanol, green dots). However, the transition occurs sharply for all of the other solutes. In the case of 1-naphthol, for example, the system transitions from being weakly coupled to strongly coupled in the small span of λ from 0.333 to 0.4 (Figure 3, 1-naphthol, yellow and green dots). Similar biphasic behavior occurs for alanine dipeptide and especially for 3,4-diphenyltoluene. Evidently, there is a critical λ value for each solute when the weakly coupled and strongly coupled modes are equally probable. However, because interconversions from one state to the other are extremely rare, it would be very difficult to pinpoint the equilibrium value accurately (see Figure 6). Indeed, the equilibration analysis described later (Table I and Figure 7) shows that, with the exception of ethanol and perhaps alanine dipeptide, the sequence of distributions in Figure 3 obtained with the linear alchemical potential is unlikely to be converged.

The small number of transitions between uncoupled (dehydrated) and coupled (hydrated) states is due to the small probability of visiting the so-called no man's land of interaction energies of low probability between the two states when applying the concerted alchemical transformation. For 3,4-diphenyltoluene, for example, the probability for observing a configuration with zero solute-solvent interaction energy is immeasurably small (Figure 3: 3,4-diphenyltoluene). For the less extreme examples of alanine dipeptide and 1-naphthol, it is possible to observe distributions with mixtures of weakly coupled and strongly coupled states with very small, yet observable density in the no man's land region. Even for these solutes, however, hydration and dehydration events are rarely observed in the simulations with the linear alchemical potential (Table I and Figure 6).

Due to the lack of a sufficient number of transitions, it is not feasible to accurately estimate the relative equilibrium populations of hydrated and dehydrated states of the bulky solutes at any of the λ-states. In turn, because the hydration free energy depends on the relative populations of coupled and uncoupled states as a function of λ,[22] it is expected (and largely

confirmed, see below) that with the linear alchemical potential the hydration free energy estimates for the solutes, with the exception of ethanol, are likely to be substantially biased by finite sampling.

Collectively, this data shows that conventional linear alchemical interpolation schemes are not generally suitable for the implementation of a reliable concerted alchemical protocols in which Lennard-Jones and electrostatic interactions are removed simultaneously.

B. Probability Distributions of the Solute-Solvent Interaction Energy with the Softplus Alchemical Potential

The data shown in Figure 3 confirms that the observed probability distributions of the soft-core solute-solvent interaction energies, which were obtained from the concerted alchemical calculations (dots in Figure 3), are accurately reproduced by the analytical model for $p_0(u_{sc})$[20,22] that was parameterized to each dataset (continuous lines in Figure 3). The parameters of the model that were optimized by the maximum likelihood method are listed in Table II. The model reproduces the distributions' positions and their variations as a function of λ, including the transition points from dehydrated to hydrated states.

The parameterized analytical functions for $p_0(u_{sc})$ are differentiated analytically with respect to u_{sc} to obtain the corresponding λ-functions $\lambda_0(u_{sc})$ [Eq. (15)][20] shown in Figure 4 (black curves). These functions are used to predict, graphically, the location of the maxima and minima of the probability distributions of the soft-core solute-solvent interaction energy as λ is varied. The procedure consists of finding the intersections between the λ-function and, in the case of a linear perturbation potential, the horizontal line drawn at the level of the desired value of λ [see Eq. (16)].

Indeed, the predictions from the λ-functions in Figure 4 are quantitatively consistent with the observations in Figure 3. In each case, the distributions near $\lambda = 0$ are expected to have one mode at large values of u_{sc}. As λ increases, (that is, as the horizontal line shifts up), a back-bending region[23] of the λ-function is encountered, in which three intersections typically occur. The first and the last correspond to maxima of the distributions and the middle intersection to a minimum. More complex patterns can arise when there are multiple back-bending regions. In the case of ethanol, the distributions are predicted to be unimodal at favorable values of the solute-solvent interaction energy above approximately $\lambda = 0.4$ (Figure 4, ethanol). For all the other solutes, the back-bending of the λ-function is predicted to be so extreme in order to prevent, in principle, unimodal distributions with the linear potential function even for values of λ close to 1. While the λ-functions can be used to identify the locations of maxima and minima of the distributions, they alone do not predict the relative populations of competing modes. Thus, for example, 1-naphthol at $\lambda = 1$ is overwhelmingly more likely to form favorable interactions with the solvent even though a second, immeasurably small, mode is predicted to exist at unfavorable interaction energies.

The softplus perturbation potential can be tuned to reduce the perturbation energy gap across the hydration/dehydration transition and avoid hard-to-converge bimodal distributions.[20] As shown in Figure 4 for ethanol, the parameters of the softplus potential can be tuned to avoid multiple intersections with the λ-function, thereby avoiding the occurrence of bimodal distributions (see below). When this is not feasible, as for alanine dipeptide, 1-naphthol, and 3,4-diphenyltoluene in Figure 4, the softplus potential can still be tuned to reduce the energy gap between competing modes and to favor one mode over another.

The alchemical simulations with the softplus potential, with the optimized parameters (Table II) that were designed based on the predicted λ-functions (Figure 4), result in the elimination or at least the reduction of solute-solvent interaction energy gaps along the alchemical pathway (Figure 5). Unlike the concerted simulations with the linear alchemical potential (Figure 3), simulations with the softplus alchemical potential yield a gradual shift of the probability distributions of the solute-solvent interaction energy as the solute is coupled to the solvent (Figure 5). The distributions with the softplus potential are also generally unimodal, which reflect either lack of multiple stable conformational macrostates or the gradual shift from one macrostate to another as λ is varied. The softplus potential effect is not as significant for ethanol, which does not exhibit a sharp hydration transition with the linear alchemical potential (Figure 3). However, in the case of 1-naphthol, the softplus potential has a very substantial effect. In this case, an interaction energy gap of more than 30 kcal/mol (Figure 3) is virtually eliminated (Figure 5), thereby facilitating transitions between the weakly coupled and strongly coupled states on opposite sides of the gap (see below). As illustrated below, the hydration calculations for alanine dipeptide and 3,4-diphenyltoluene also significantly benefit from using the softplus alchemical potential, even though the no man's land with respect to the solute-solvent interaction energy is not eliminated (see Figure 5) in these cases.

C. Analysis of Replica Exchange Efficiency

Hamiltonian replica exchange efficiency has been monitored here in terms of the extent of diffusion of replicas in the solute-solvent interaction energy space. In particular, we monitored the rate of transitions between coupled states with favorable solute-solvent interaction energies and uncoupled states with large and unfavorable interaction energy. The time trajectories of the soft-core solute-solvent interaction energies that were sampled by the replicas are shown in Figure 6 for each alchemical hydration simulation. The number of transitions from uncoupled to coupled states (hydration) and vice versa (dehydration) are presented in Table I.

The data indicates that ethanol undergoes frequent hydration/dehydration transitions with either the linear and softplus perturbation potentials. However, the bulkier solutes (alanine dipeptide, 1-naphthol, and 3,4-diphenyltoluene) undergo very few hydration/dehydration transitions with the linear alchem-

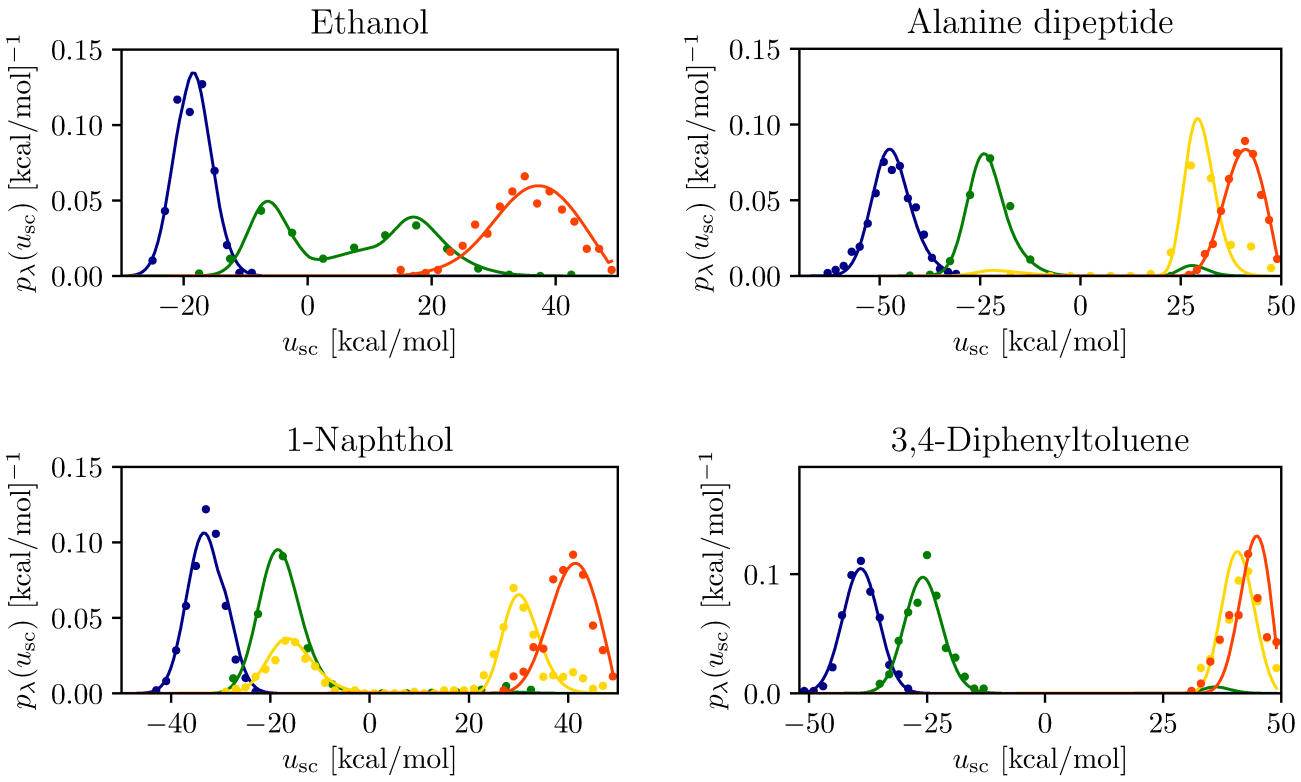

FIG. 3. The predicted (continuous lines) and observed (dots) probability densities of the soft-core solute-solvent interaction energy, $p_\lambda(u_{sc})$, for the concerted alchemical hydration calculations with the linear alchemical potential $W_\lambda(u_{sc}) = \lambda u_{sc}$ for (clockwise starting from the upper left) ethanol, alanine dipeptide, 1-naphthol, and 3,4-diphenyltoluene. Ethanol: $\lambda = 0$ (orange/red), $\lambda = 0.267$ (green), and $\lambda = 1$ (blue). Alanine dipeptide: $\lambda = 0$ (orange/red), $\lambda = 0.333$ (gold), $\lambda = 0.400$ (green), and $\lambda = 1$ (blue). 1-naphthol: $\lambda = 0$ (orange/red), $\lambda = 0.333$ (gold), $\lambda = 0.4$ (green), and $\lambda = 1$ (blue). 3,4-diphenyltoluene: $\lambda = 0$ (orange/red), $\lambda = 0.2$ (gold), $\lambda = 0.467$ (green), and $\lambda = 1$ (blue). The predicted distributions are obtained using the analytical model for $p_0(u_{sc})$, Eqs. (12) to (14)], with the parameters listed in Table II.

ical potential. This behavior is due to the wide interaction energy gap between the distributions of the solute-solvent interaction energies for the the coupled and uncoupled states (Figure 3). The bulky nature of these solutes makes it very improbable for a solute molecule to find a suitable configuration in the solvent in order to make favorable interactions with the solvent when nearly decoupled from it. Conversely, these solutes are unlikely to overcome the energetic penalty necessary to become decoupled from the solvent upon making some favorable interactions. The coupled and decoupled states of the bulky solutes are essentially separated by a first-order phase transition that is entropically frustrated in the hydration direction and energetically frustrated in the dehydration direction.[23]

On the other hand, the optimized softplus perturbation potential is very effective at circumventing the phase transition. As shown in Figure 6 and Table I, many more hydration/dehydration transitions are observed with the softplus potential than with the linear potential. A higher number of transitions make it possible to more rapidly equilibrate and to converge the hydration free energy estimates for the bulkier

TABLE I. Hydration free energy estimates, equilibration times and the number of hydration and dehydration transitions for the solutes considered in this work.

Protocol	ΔG_h° [a]	t_{eq} [b]	n_{hydr}	n_{dehydr}
	Ethanol[c]			
Linear	-2.23 ± 0.09	0.00	96	96
Softplus	-2.12 ± 0.11	0.63	322	292
	Alanine dipeptide[d]			
Linear	-8.06 ± 0.14	0.13	0	2
Softplus	-8.25 ± 0.21	0.00	45	43
	1-Naphthol[e]			
Linear	-4.53 ± 0.18	3.50	6	7
Softplus	-3.66 ± 0.14	0.00	34	32
	3,4-Diphenyltoluene[f]			
Linear	-0.33 ± 0.18	1.38	0	2
Softplus	3.51 ± 0.26	3.38	15	14

[a] kcal/mol
[b] ns
[c] $u_{lower} = -10$, $u_{upper} = 25$ kcal/mol
[d] $u_{lower} = -35$, $u_{upper} = 25$ kcal/mol
[e] $u_{lower} = -20$, $u_{upper} = 25$ kcal/mol
[f] $u_{lower} = -30$, $u_{upper} = 30$ kcal/mol

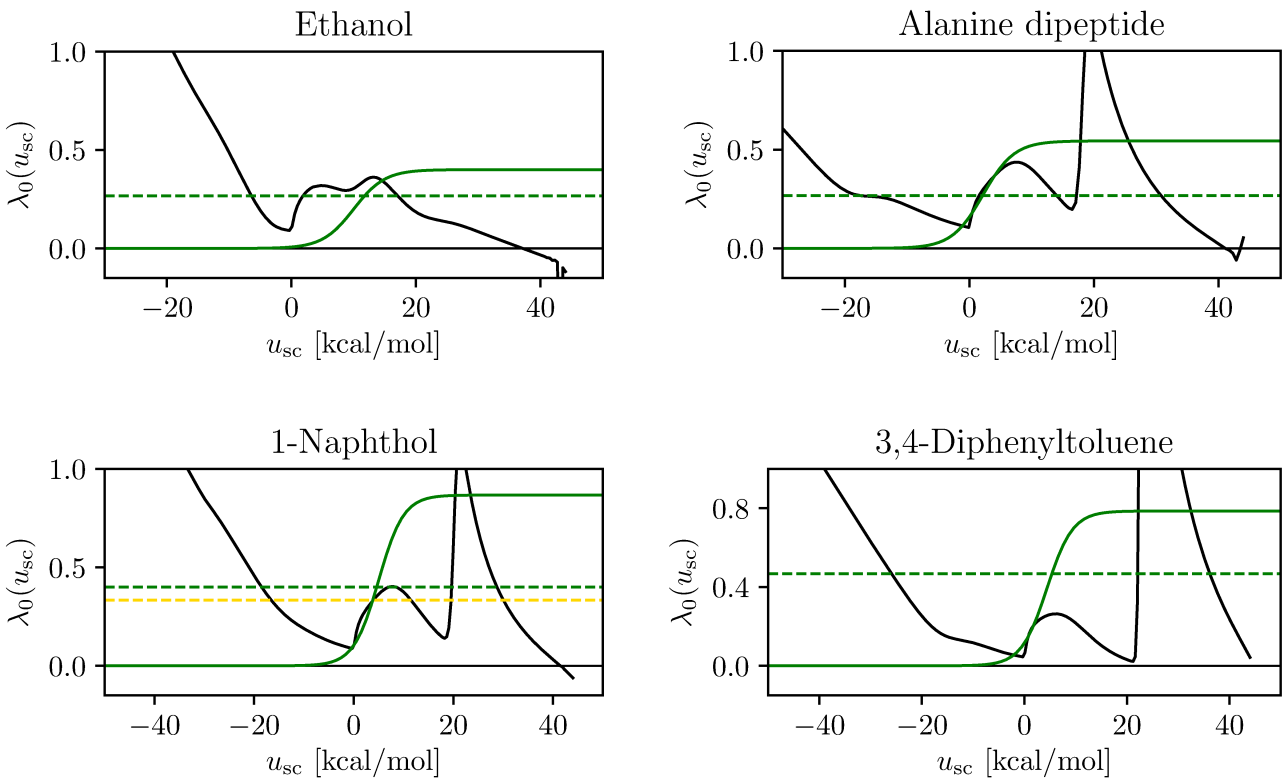

FIG. 4. The predicted λ-functions (black continuous lines) for the concerted hydration of the indicated solutes were obtained from the analytical model for $p_0(u_{sc})$ [Eqs. (12) to (14)], and Eq. (15) with the parameters listed in Table II. The dashed horizontal lines correspond to the values of λ where the probability distributions of the soft-core solute solvent energy with the linear alchemical potential are observed to be bimodal. Ethanol: $\lambda = 0.267$; alanine dipeptide: $\lambda = 0.0$ and $\lambda = 0.544$, ; 1-naphthol: $\lambda = 0.333$ and $\lambda = 0.4$, 3,4-diphenytoluene: $\lambda = 0.467$. The intersections of the horizontal lines with the λ-function correspond to the maxima and minima of the corresponding distributions with the linear alchemical potential (Figure 3).[20] The solid sigmoid curve is $\partial W_\lambda(u_{sc})/\partial u_{sc}$, the logistic function that results from the differentiation of the softplus function at the following values of λ. Ethanol: $\lambda = 0.267$; alanine dipeptide: $\lambda = 0.267$; 1-naphthol: $\lambda = 0.4$; 3,4-diphenyltoluene: $\lambda = 0.4$. See Tables III–VI.

solutes without additional computational expense (see below).

D. Equilibration of the Hydration Free Energy Estimates

The computed hydration free energy estimates are plotted in Figure 7 as a function of the amount of data discarded from the start of the simulation (the equilibration time). These plots, referred to as reverse cumulative equilibration profiles,[19,44,45] are used to determine the time after which the time series of data generated by the simulation becomes stationary and not biased by the starting conformation of the system. It is not obvious to pinpoint such a time because, while the accuracy of the binding free energy presumably improves as more non-equilibrated samples are discarded, the precision of the estimate (indicated by the error bars in Figure 7) decreases as more initial samples are discarded. Here we take the approach of choosing the hydration free energy estimate as the value that corresponds to the shortest equilibration time, which gives a value statistically indistinguishable from those at all subsequent equilibration times.[19]

Based on this criterion, we conclude that the hydration free energy estimate for ethanol equilibrates almost immediately after the start of the simulation with either the linear or the softplus alchemical potentials. The hydration free energy estimates for ethanol with the linear and softplus potentials are in statistical agreement (Table I). Given the very different nature of the alchemical paths in the two simulations, we conclude that equilibration and convergence of the hydration free energies have been achieved in this case. This positive outcome is consistent with the high rate of hydration and dehydration transitions observed (Figure 6 and Table I) for ethanol with either the linear and softplus potentials. The result for ethanol confirms the validity of the concerted alchemical protocol and the correctness of the simulation algorithms in producing a canonical ensemble of conformations with either potential.

A similar analysis for the bulkier solutes (alanine dipeptide, 1-naphthol, and 3,4-diphenyltoluene) shows that equilibration of the concerted free energy protocol is achieved rapidly when using the optimized softplus potential (Figure

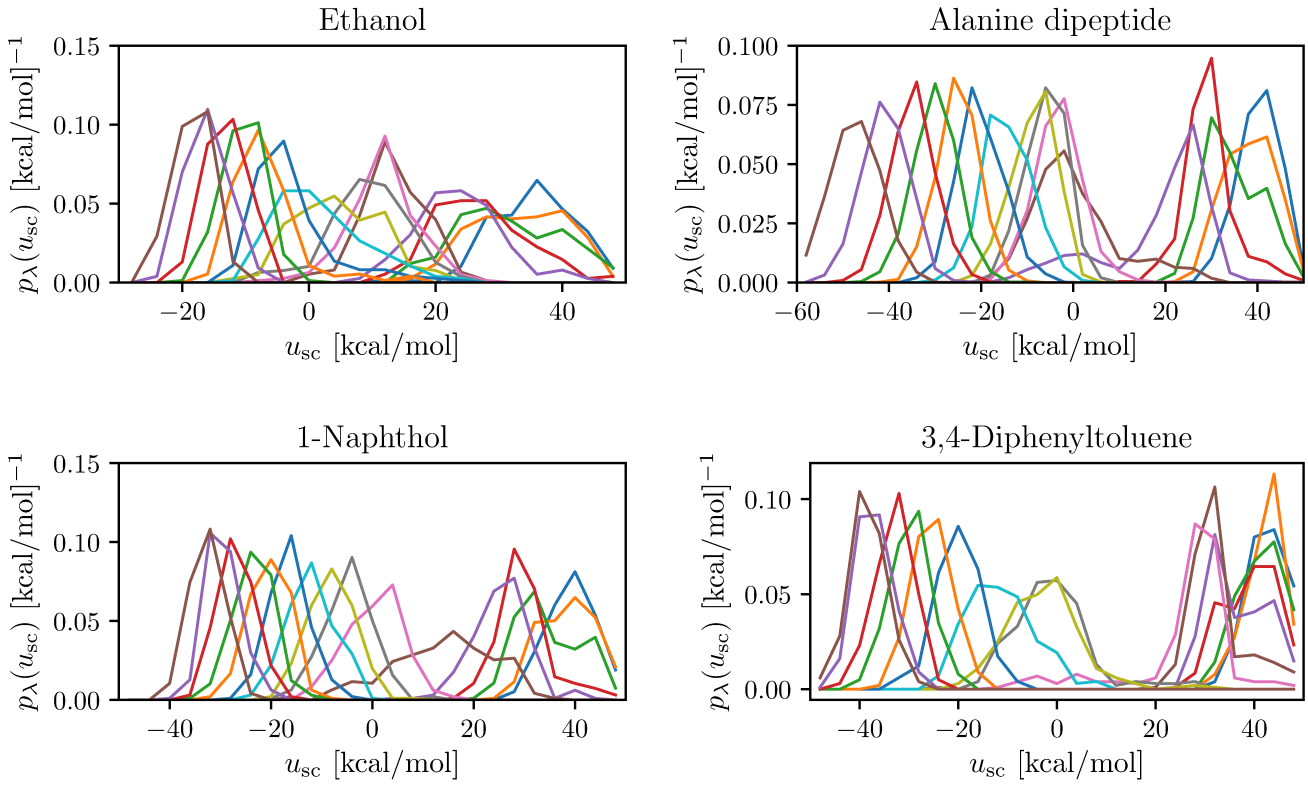

FIG. 5. The probability densities of the soft-core solute-solvent interaction energy obtained from the concerted alchemical hydration simulations performed with the softplus alchemical potential [Eq. (5)] and the parameters in Table II for the hydration of the solutes indicated. The probability densities, drawn with alternating colors, shift from right to left as the λ alchemical parameter progressively varies from 0 (vacuum state) to 1 (hydrated state). The probability densities shown here have been obtained from the second half of the simulation trajectories.

7: second column). Based on the relatively large numbers of hydration/dehydration transitions observed for these solutes, the hydration free energy estimates are considered to be well converged. 3,4-diphenyltoluene, the largest and most hydrophobic solute, is confirmed as a stringent test case for one-step concerted alchemical protocols.[13] Despite the relatively few hydration/dehydration transitions observed even after optimization of the softplus potential, the approach employed here appears to yield a converged and reliable hydration free energy estimate for this compound. After taking into account long-range van der Waals interactions, which are expected to contribute as much as 3 kcal/mol of additional solute-solvent interaction energy, the hydration free energy obtained here for 3,4-diphenyltoluene in a droplet is consistent with earlier calculations with periodic boundary conditions.[13] To our knowledge, the measured hydration free energy for this compound has never been reported.

In contrast to the softplus potential, the concerted hydration alchemical protocol with the linear potential fails to reach reliable convergence for the bulkier solutes. This is indicated by the very few hydration/dehydration transitions (Table I) and the cumulative equilibration profiles (Figure 7: first column, alanine dipeptide, 1-naphthol, and 3,4-diphenyltoluene). For alanine dipeptide, the rapid equilibration of the hydration free energy estimate with the linear potential is likely circumstantial as it hinges on only two dehydration transitions early in the replica exchange trajectories (Figure 6: alanine dipeptide, left panel). With the linear perturbation potential, the hydration free energy of 1-naphthol appears to equilibrate after 3 ns at a value of -4.53 ± 0.18 kcal/mol, which is statistically inconsistent with the less biased estimate of -3.66 ± 0.14 obtained with the softplus potential (Table I and Figure 7: 1-naphthol). The concerted alchemical hydration of 3,4-diphenyltoluene is a good example of the potential convergence pitfalls of free energy calculations. With the linear potential (Figure 7: 3,4-diphenyltoluene), the hydration free energy appears to equilibrate after approximately 1.3 ns at a very stable plateau value, which may be mistaken as a converged estimate. In actuality, the value of the plateau is not converged as confirmed by the lack of hydration/dehydration transition events after 1.3 ns of simulation per replica (Figure 6: 3,4-diphenyltoluene, left panel).

The data collected for these solutes strongly supports the finding that commonly used linear alchemical perturbation potentials are unsuitable for obtaining converged free energy estimates with a concerted alchemical process. In contrast,

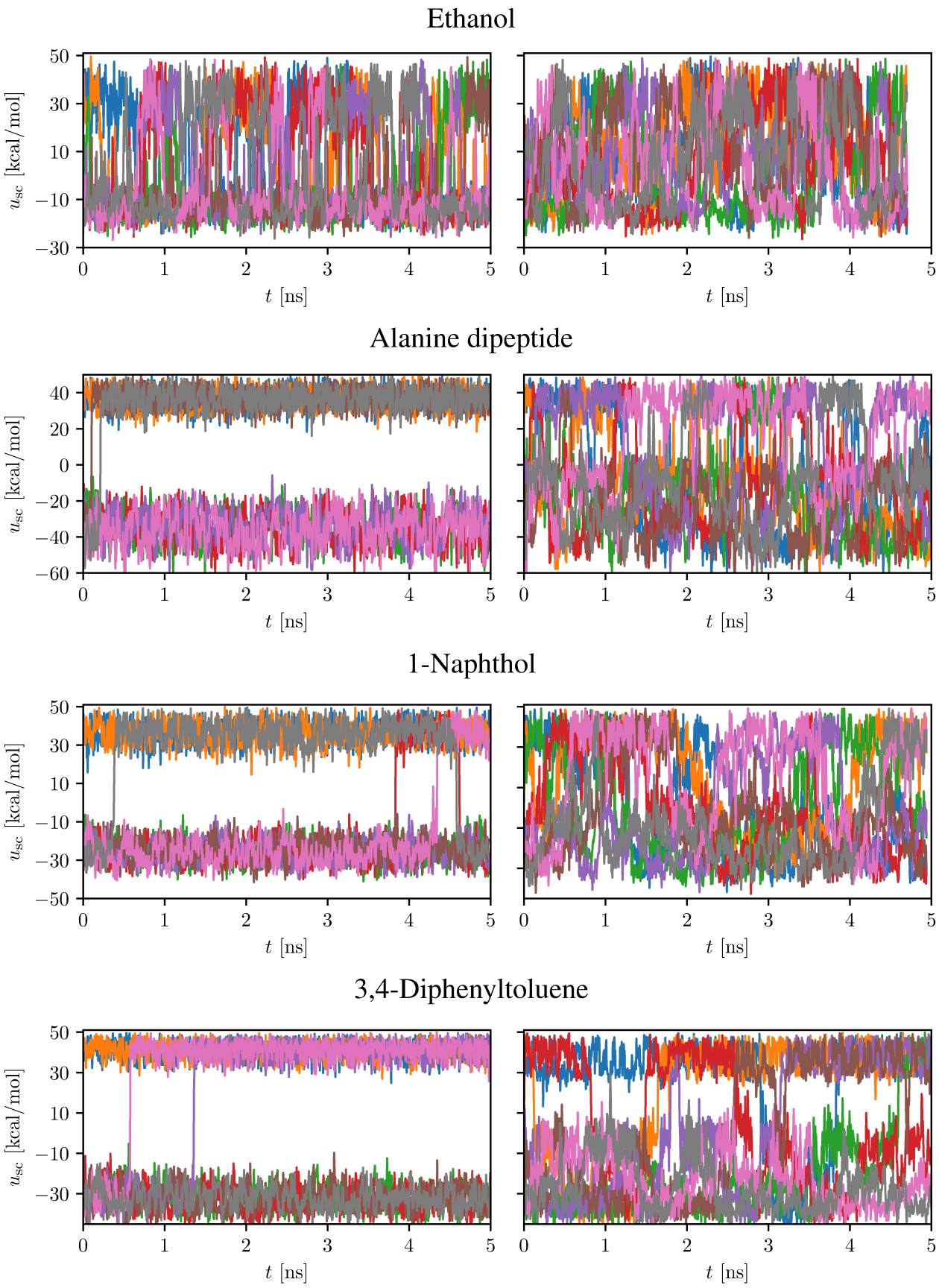

FIG. 6. Time trajectories of the soft-core solute-solvent interaction energy for selected replicas of the replica exchange simulations of the alchemical hydration of ethanol, alanine dipeptide, 1-naphthol, and 3,4-diphenyltoluene with the linear (left) and softplus (right) alchemical perturbation potentials. Ethanol undergoes frequent hydration and dehydration transitions with either alchemical perturbation function. In contrast, significant numbers of hydration and dehydration transitions are observed only with the softplus perturbation potential for the bulkier solutes (alanine dipeptide, 1-naphthol, and 3,4-diphenyltoluene).

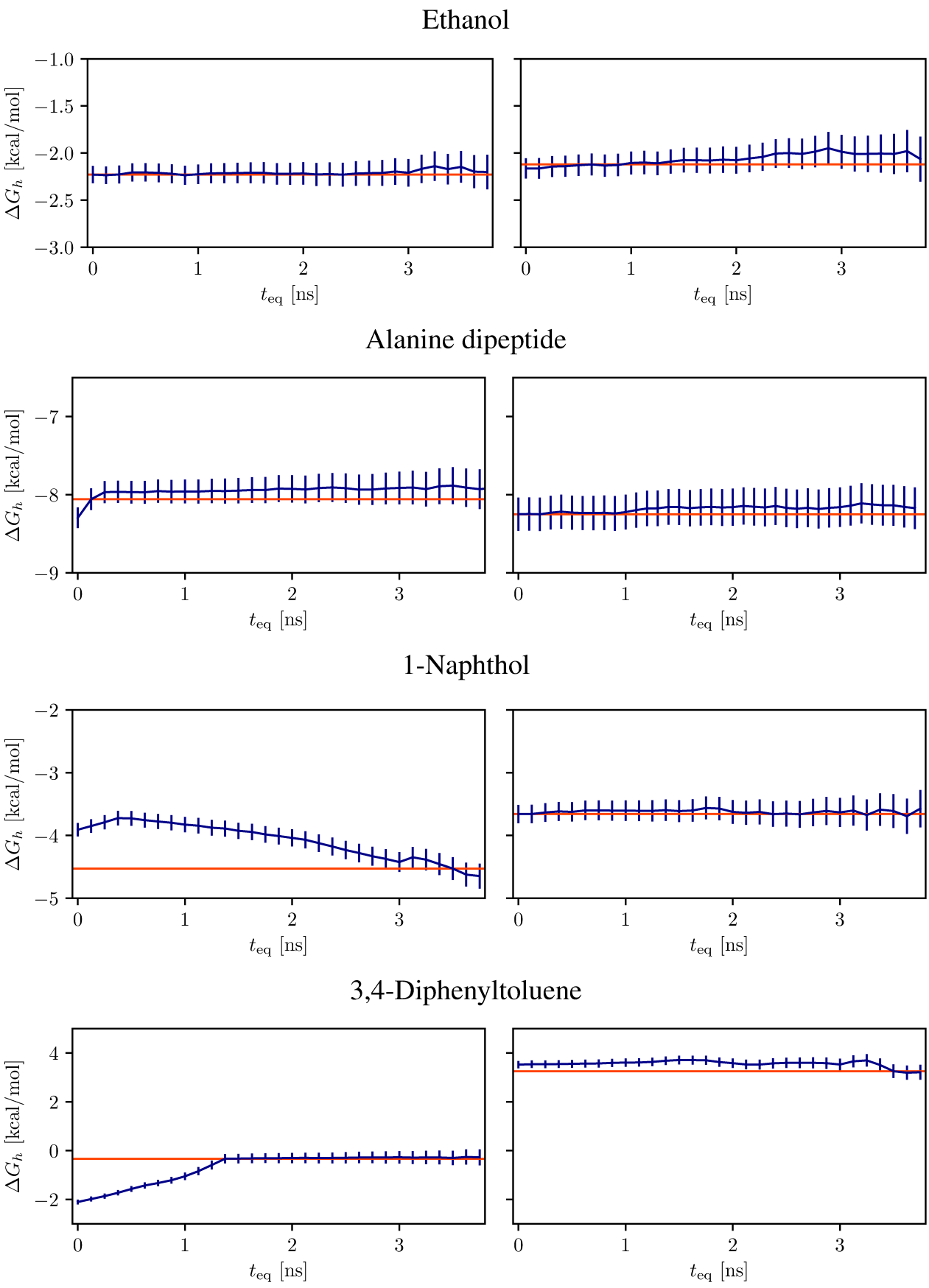

FIG. 7. Reverse cumulative equilibration profiles of the hydration free energies of the solutes indicated with the linear (left) and softplus (right) alchemical perturbation potentials. The horizontal line corresponds to the hydration free energy estimate for each case, defined as the hydration free energy at the earliest equilibration time that corresponds to an estimate statistically consistent with those of all of the following equilibration times.

optimized softplus potentials can enable rapid equilibration and convergence by promoting hydration/dehydration transitions.

IV. DISCUSSION

Alchemical hydration free energy calculations are widely employed to predict the water solubilities of substances,[4,46] to test force field and free energy protocols,[47–49] and to estimate absolute and relative binding free energies of molecular complexes.[3,50] While alternatives have been proposed,[13,51] the guidelines that are generally accepted to avoid slow convergence and numerical instabilities involve splitting the alchemical hydration process into two steps. In the first step, volume-exclusion and dispersion solute-solvent interactions are turned on, followed by a second step in which electrostatic interactions are established.[9] In addition, especially during the volume-exclusion phase, it is customary to employ soft-core interatomic pair potentials.[12,24,52]

These free energy practices, which are widely implemented in popular molecular simulation packages,[32,53–56] are generally successful and robust.[57] Nevertheless, it would be beneficial to explore concerted alternatives with improved equilibration and convergence characteristics. The software implementations of free energy methods can also be cumbersome and challenging to integrate and maintain alongside other molecular simulation algorithms. König et al.[52] have recently discussed the limitations of conventional non-linear and non-algebraic forms of soft-core pair potentials and have proposed a family of alchemical hybrid perturbation potentials that addresses some of the challenges. A recent study by Lee et al.[13] discussed the downsides of multi-step free energy methods and called for a streamlined concerted approach that would be more easily integrated with extended ensemble and self-adjusting conformational sampling algorithms, as well as non-equilibrium approaches.[58–60] Lee et al.[13] proposed a family of soft-core pair potentials and non-linear alchemical hybrid perturbations that facilitate the calculation of hydration free energies in a concerted fashion.

As shown in this work, concerted alchemical protocols present unique computational challenges that have likely discouraged progress in earlier attempts. We have shown, for example, that conventional linear alchemical interpolating potentials can introduce severe entropic bottlenecks that preclude the rapid equilibration between the hydrated and dehydrated states of the solute, which result in strongly biased free energy estimates.

In a recent study,[20] we identified pseudo first-order phase transitions along the alchemical path as the critical and fundamental obstacles to the application of concerted alchemical processes for molecular binding. In the same study, we addressed the issue by applying Hamiltonian-shaping techniques inspired by non-Boltzmann sampling methodologies developed by Straub and collaborators[23,61] for the study of temperature-driven first-order phase transitions. Using this approach, we were able to identify alchemical paths that avoid or soften biphasic behavior in the context of binding with an implicit description of the solvent.[20] In the present study, we demonstrate that these similar techniques can also be applied to the estimation of hydration free energies with explicit solvation. Critical to this approach is a new family of alchemical perturbation functions presented here that is potentially of broader applicability. We also illustrate a theoretical framework and a graphical procedure to optimize them in order to systematically improve conformational sampling and the rate of equilibration and convergence of free energy estimates.

In this work, we considered the hydration of solutes into a water droplet, a simplified system that is a first, yet representative validation of the method. The droplet model was selected mainly for expediency in order to validate the theory without extensive modifications of our Single Decoupling Method (SDM) software that was originally developed alongside an implicit description of the solvent.[20] In SDM, the perturbation from the coupled state to the uncoupled state of the molecular complex is achieved by displacing the ligand out of the receptor binding pocket and into the surrounding implicit solvent bulk. The same approach would not work for decoupling the solute molecule from an explicit solvent system with periodic boundary conditions because a periodic boundary system does not have an outer region where the solute can be placed in. A software implementation of our method with periodic boundary conditions would require the implementation of solute-solvent decoupling through the tuning of force field parameters. While OpenMM supports this procedure,[62] our immediate focus is to build upon the results of this work in order to develop concerted binding free energy estimation protocols with explicit solvation and periodic boundary conditions. This work is ongoing and will be reported in upcoming publications.

The theoretical, algorithmic, and numerical simulation strategies presented here and the promising results illustrated here indicate a possible roadmap to streamline the implementation of alchemical protocols in molecular simulation packages. The first suggestion is to replace soft-core pair potentials with, as done here, a soft-core function applied to the solute-solvent interaction energy [e.g. using Eqs. (7) and (8)]. This approach would greatly simplify the implementation of alchemical protocols in molecular dynamics packages by making it unnecessary to modify the core energy and force subroutines. While the method used here requires the reprocessing of the forces, this is cleanly and efficiently accomplished in our implementation by a $O(N)$ loop coded in a separate small subroutine of the molecular dynamics integrator. The family of alchemical potentials introduced here also simplifies the thermodynamic reweighting procedure with MBAR or UWHAM[15,25] for free energy estimation. Alchemical implementations using the molecular dynamics engine with soft-core pair potentials require the re-evaluation of each sample's perturbation energy for each λ state on the alchemical path.[14] This cumbersome process is completely bypassed in our implementation. The alchemical potential introduced here is a one-to-one function of the solute-solvent interaction energy, which can be saved together with each trajectory frame and manipulated algebraically with minimal effort, using Eq. (3), to yield the perturbation energies at any of the alchemical

states.

The second suggestion emerging from this study is the use of a non-linear alchemical perturbation that, like the softplus potential we propose [Eq. 5], judiciously warps the alchemical path in critical regions to avoid phase transitions while simultaneously retaining a linear character away from the problematic regions. Furthermore, we suggest the use of the λ-function formalism [Eqs. (15) and (16)] and adaptive maximum likelihood-based approaches to systematically optimize not only the alchemical schedule, but also the next generation of alchemical algebraic perturbation functions[13,52] in order to enhance the rate of equilibration and convergence.

V. CONCLUSIONS

We employ a statistical analytical theory of molecular association[22] and a single-decoupling alchemical method for binding free energy estimation with implicit solvation[20,23] to develop a concerted alchemical protocol for the calculation of hydration free energies of small to medium-sized molecules with explicit solvation. The concerted alchemical hydration protocol introduced here involves a single alchemical transformation rather than the more commonly employed pair of distinct decoupling transformations for electrostatic and steric/dispersion interactions. The approach and its benefits are illustrated for model systems involving the hydration of four diverse solutes in a water droplet. This proof of principle study paves the way for a new generation of streamlined concerted free energy estimation algorithms in condensed phases.

VI. DATA AVAILABILITY STATEMENT

The software used to generate the data is available on public Github repositories as indicated in Computational Details. The input files and the molecular dynamics trajectories that support the findings of this study are available from the corresponding author upon request.

ACKNOWLEDGMENTS

We acknowledge support from the National Science Foundation (NSF CAREER 1750511 to E.G.). Molecular simulations were conducted on the Comet GPU supercomputer cluster at the San Diego Supercomputing Center supported by NSF XSEDE award TG-MCB150001.

Appendix A: Appendix

1. Parameters of the Analytical Model for $p_0(u_{sc})$

See Table II.

2. Alchemical Schedule and Parameters of the Softplus Perturbation Function

See Tables III and V.

[1] A. Ben Naim. *Water and Aqueous Solutions*. Plenum, New York, 1974.
[2] Christel AS Bergström and Per Larsson. Computational prediction of drug solubility in water-based systems: qualitative and quantitative approaches used in the current drug discovery and development setting. *Inter. J. Pharm.*, 540(1-2):185–193, 2018.
[3] M. K. Gilson, J. A. Given, B. L. Bush, and J. A. McCammon. The statistical-thermodynamic basis for computation of binding affinities: A critical review. *Biophys. J.*, 72:1047–1069, 1997.
[4] David L Mobley and J Peter Guthrie. Freesolv: a database of experimental and calculated hydration free energies, with input files. *J. Comp. Aid. Mol. Des.*, 28(7):711–720, 2014.
[5] Chipot and Pohorille (Eds.). *Free Energy Calculations. Theory and Applications in Chemistry and Biology*. Springer Series in Chemical Physics. Springer, Berlin Heidelberg, Berlin Heidelberg, 2007.
[6] Anita de Ruiter and Chris Oostenbrink. Free energy calculations of protein–ligand interactions. *Current opinion in chemical biology*, 15(4):547–552, 2011.
[7] Thomas Simonson. Free energy of particle insertion: an exact analysis of the origin singularity for simple liquids. *Molecular Physics*, 80(2):441–447, 1993.
[8] David L Mobley. Lets get honest about sampling. *J. Comp. Aided Mol. Des.*, 26:93–95, 2012.
[9] Pavel V Klimovich, Michael R Shirts, and David L Mobley. Guidelines for the analysis of free energy calculations. *J. Comp. Aid. Mol. Des.*, 29(5):397–411, 2015.
[10] Tai-Sung Lee, Bryce K Allen, Timothy J Giese, Zhenyu Guo, Pengfei Li, Charles Lin, T Dwight McGee Jr, David A Pearlman, Brian K Radak, Yujun Tao, Hsu-Chun Tsai, Huafeng Xu, Woody Sherman, and Darrin M York. Alchemical binding free energy calculations in AMBER20: Advances and best practices for drug discovery. *J. Chem. Inf. Model.*, 2020.
[11] Thomas Steinbrecher, David L Mobley, and David A Case. Nonlinear scaling schemes for lennard-jones interactions in free energy calculations. *J Chem Phys*, 127:214108, 2007.
[12] Jed W Pitera and Wilfred F van Gunsteren. A comparison of non-bonded scaling approaches for free energy calculations. *Molecular Simulation*, 28(1-2):45–65, 2002.
[13] Tai-Sung Lee, Zhixiong Lin, Bryce K Allen, Charles Lin, Brian K Radak, Yujun Tao, Hsu-Chun Tsai, Woody Sherman, and Darrin M York. Improved alchemical free energy calculations with optimized smoothstep softcore potentials. *J. Chem. Theory Comput.*, 2020.
[14] Yaozong Li and Kwangho Nam. Repulsive soft-core potentials for efficient alchemical free energy calculations. *J. Chem. Theory Comput.*, 16(8):4776–4789, 2020.
[15] Michael R Shirts and John D Chodera. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.*, 129:124105, 2008.
[16] Zoe Cournia, Bryce K Allen, Thijs Beuming, David A Pearlman, Brian K Radak, and Woody Sherman. Rigorous free energy simulations in virtual screening. *Journal of Chemical Information and Modeling*, 2020.
[17] Baofeng Zhang, Michael P. DErasmo, Ryan P. Murelli, and Emilio Gallicchio. Free energy-based virtual screening and optimization of RNase H inhibitors of HIV-1 reverse transcriptase. *ACS Omega.*, 1:435–447, 2016.
[18] Rajat Kumar Pal, Steve Ramsey, Satishkumar Gadhiya, Pierpaolo Cordone, Lauren Wickstrom, Wayne W. Harding, Tom Kurtzman, and Emilio Gallicchio. Inclusion of enclosed hydration effects in the binding free energy estimation of dopamine D3 receptor complexes. *PLoS One*, 14(9):e0222092, 2019.
[19] Denise Kilburg and Emilio Gallicchio. Assessment of a single decoupling alchemical approach for the calculation of the absolute binding free energies of protein-peptide complexes. *Frontiers in Molecular Biosciences*, 5:22, 2018.
[20] Rajat K Pal and Emilio Gallicchio. Perturbation potentials to overcome order/disorder transitions in alchemical binding free energy calculations. *J. Chem. Phys.*, 151(12):124116, 2019.

[21] Qing Lu, Jaegil Kim, James D Farrell, David J Wales, and John E Straub. Investigating the solid-liquid phase transition of water nanofilms using the generalized replica exchange method. *J. Chem. Phys.*, 141(18):18C525, 2014.

[22] Denise Kilburg and Emilio Gallicchio. Analytical model of the free energy of alchemical molecular binding. *J. Chem. Theory Comput.*, 14(12):6183–6196, 2018.

[23] Jaegil Kim and John E Straub. Generalized simulated tempering for exploring strong phase transitions. *J. Chem. Phys.*, 133(15):154101, 2010.

[24] Thomas Steinbrecher, InSuk Joung, and David A. Case. Soft-core potentials in thermodynamic integration: Comparing one- and two-step transformations. *J. Comput. Chem.*, 32:3253–3263, 2011.

[25] Zhiqiang Tan, Emilio Gallicchio, Mauro Lapelosa, and Ronald M. Levy. Theory of binless multi-state free energy estimation with applications to protein-ligand binding. *J. Chem. Phys.*, 136:144102, 2012.

[26] E. Gallicchio, M. M. Kubo, and R. M. Levy. Entropy-enthalpy compensation in solvation and ligand binding revisited. *J. Am. Chem. Soc.*, 120:4526–27, 1998.

[27] Tom L. Beck, Michael E. Paulaitis, and Lawrence R. Pratt. *The Potential Distribution Theorem and Models of Molecular Solutions*. Cambridge University Press, New York, 2006.

[28] Emilio Gallicchio, Mauro Lapelosa, and Ronald M. Levy. Binding energy distribution analysis method (BEDAM) for estimation of protein-ligand binding affinities. *J. Chem. Theory Comput.*, 6:2961–2977, 2010.

[29] Emilio Gallicchio and Ronald M Levy. Recent theoretical and computational advances for modeling protein-ligand binding affinities. *Adv. Prot. Chem. Struct. Biol.*, 85:27–80, 2011.

[30] Tai-Sung Lee, Brian K Radak, Anna Pabis, and Darrin M York. A new maximum likelihood approach for free energy profile construction from molecular simulations. *J. Chem. Theory Comput.*, 9(1):153–164, 2012.

[31] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[32] Peter Eastman, Jason Swails, John D Chodera, Robert T McGibbon, Yutong Zhao, Kyle A Beauchamp, Lee-Ping Wang, Andrew C Simmonett, Matthew P Harrigan, Chaya D Stern, et al. Openmm 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comp. Bio.*, 13(7):e1005659, 2017.

[33] Albert C Pan, Huafeng Xu, Timothy Palpant, and David E Shaw. Quantitative characterization of the binding and unbinding of millimolar drug fragments with molecular dynamics simulations. *J. Chem. Theory Comput.*, 13(7):3372–3377, 2017.

[34] Christoph Ohlknecht, Jan Walther Perthold, Bettina Lier, and Chris Oostenbrink. Charge-changing perturbations and path sampling via classical molecular dynamic simulations of simple guest–host systems. *J. Chem. Theory Comput.*, Article ASAP, 2020.

[35] Junmei Wang, Romain M Wolf, James W Caldwell, Peter A Kollman, and David A Case. Development and testing of a general amber force field. *J. Comp. Chem.*, 25(9):1157–1174, 2004.

[36] Emilio Gallicchio, Junchao Xia, William F Flynn, Baofeng Zhang, Sade Samlalsingh, Ahmet Mentes, and Ronald M Levy. Asynchronous replica exchange software for grid and heterogeneous computing. *Computer Physics Communications*, 196:236–246, 2015.

[37] Yuji Sugita, Akio Kitao, and Yuko Okamoto. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.*, 113:6042–6051, 2000.

[38] A. K. Felts, Y. Harano, E. Gallicchio, and R. M. Levy. Free energy surfaces of beta-hairpin and alpha-helical peptides generated by replica exchange molecular dynamics with the AGBNP implicit solvent model. *Proteins: Struct. Funct. Bioinf.*, 56:310–321, 2004.

[39] K.P. Ravindranathan, E. Gallicchio, R. A. Friesner, A. E. McDermott, and R. M. Levy. Conformational equilibrium of cytochrome P450 BM-3 complexed with N-palmitoylglycine: A replica exchange molecular dynamics study. *J. Am. Chem. Soc.*, 128:5786–5791, 2006.

[40] Hisashi Okumura, Emilio Gallicchio, and Ronald M Levy. Conformational populations of ligand-sized molecules by replica exchange molecular dynamics and temperature reweighting. *J. Comput. Chem.*, 31:1357–1367, 2010.

[41] Christopher J. Woods, Jonathan W. Essex, and Michael A. King. The development of replica-exchange-based free-energy methods. *J. Phys. Chem. B*, 107:13703–13710, 2003.

[42] Steven W. Rick. Increasing the efficiency of free energy calculations using parallel tempering and histogram reweighting. *J. Chem. Theory Comput.*, 2:939–946, 2006.

[43] Thomas Simonson. Gaussian fluctuations and linear response in an electron transfer protein. *Proc. Natl. Acad. Sci.*, 99(10):6544–6549, 2002.

[44] Wei Yang, Ryan Bitetti-Putzer, and Martin Karplus. Free energy simulations: use of the reverse cumulative averaging to determine the equilibrated region and the time required for convergence. *J. Chem. Phys.*, 120(6):2618–2628, 2003.

[45] John D Chodera. A simple method for automated equilibration detection in molecular simulations. *J. Chem. Theory Comput.*, 12(4):1799–1805, 2016.

[46] Tang-Qing Yu, Pei-Yang Chen, Ming Chen, Amit Samanta, Eric Vanden-Eijnden, and Mark Tuckerman. Order-parameter-aided temperature-accelerated sampling for the exploration of crystal polymorphism and solid-liquid phase transitions. *J. Chem. Phys.*, 140(21):06B603_1, 2014.

[47] David L Mobley, Shaui Liu, David S Cerutti, William C Swope, and Julia E Rice. Alchemical prediction of hydration free energies for sampl. *J. Comp. Aid. Mol. Des.*, 26(5):551–562, 2012.

[48] David L Mobley, Elise Dumont, John D Chodera, and Ken A Dill. Comparison of charge models for fixed-charge force fields: small-molecule hydration free energies in explicit solvent. *J Phys Chem B*, 111:2242–2254, 2007.

[49] Devleena Shivakumar, Joshua Williams, Yujie Wu, Wolfgang Damm, John Shelley, and Woody Sherman. Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the opls force field. *J. Chem. Theory Comput.*, 6(5):1509–1519, 2010.

[50] John D. Chodera, David L. Mobley, Michael R. Shirts, Richard W. Dixon, Kim Branson, and Vijay S. Pande. Alchemical free energy methods for drug discovery: Progress and challenges. *Curr. Opin. Struct. Biol.*, 21:150–160, 2011.

[51] Clara D Christ and Wilfred F van Gunsteren. Multiple free energies from a single simulation: Extending enveloping distribution sampling to nonoverlapping phase-space distributions. *The Journal of chemical physics*, 128(17):174112, 2008.

[52] Gerhard König, Nina Glaser, Benjamin Schroeder, Alžbeta Kubincová, Philippe H Hünenberger, and Sereina Riniker. An alternative to conventional λ-intermediate states in alchemical free energy calculations: λ-enveloping distribution sampling. *J. Chem. Inf. Model.*, 2020.

[53] Bernard R Brooks, Charles L Brooks III, Alexander D Mackerell Jr, Lennart Nilsson, Robert J Petrella, Benoît Roux, Youngdo Won, Georgios Archontis, Christian Bartels, Stefan Boresch, et al. Charmm: the biomolecular simulation program. *Journal of computational chemistry*, 30(10):1545–1614, 2009.

[54] James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant Kale, and Klaus Schulten. Scalable molecular dynamics with namd. *J. Comp. Chem.*, 26(16):1781–1802, 2005.

[55] Sander Pronk, Szilárd Páll, Roland Schulz, Per Larsson, Pär Bjelkmar, Rossen Apostolov, Michael R Shirts, Jeremy C Smith, Peter M Kasson, David van der Spoel, Berk Hess, and Erik Lindahl. Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29:845–854, 2013.

[56] Xibing He, Shuhan Liu, Tai-Sung Lee, Beihong Ji, Viet H Man, Darrin M York, and Junmei Wang. Fast, accurate, and reliable protocols for routine calculations of protein–ligand binding affinities in drug design projects using AMBER GPU-TI with ff14SB/GAFF. *ACS Omega*, 5(9):4611–4619, 2020.

[57] Andrea Rizzi, Travis Jensen, David R Slochower, Matteo Aldeghi, Vytautas Gapsys, Dimitris Ntekoumes, Stefano Bosisio, Michail Papadourakis, Niel M Henriksen, Bert L De Groot, et al. The SAMPL6 SAMPLing challenge: Assessing the reliability and efficiency of binding free energy calcu-

lations. *J. Comp. Aid. Mol. Des.*, pages 1–33, 2020.

[58] Eric Darve, David Rodríguez-Gómez, and Andrew Pohorille. Adaptive biasing force method for scalar and vector free energy calculations. *The Journal of chemical physics*, 128(14):144120, 2008.

[59] Piero Procacci. Solvation free energies via alchemical simulations: let's get honest about sampling, once more. *Phys. Chem. Chem. Phys.*, 2019.

[60] Marina Macchiagodena, Marco Pagliai, Maurice Karrenbrock, Guido Guarnieri, Francesco Iannone, and Piero Procacci. Virtual double-system single-box: A nonequilibrium alchemical technique for absolute binding free energy calculations: Application to ligands of the SARS-CoV-2 Main Protease. *Journal of Chemical Theory and Computation*, 2020.

[61] Qing Lu, Jaegil Kim, and John E Straub. Order parameter free enhanced sampling of the vapor-liquid transition using the generalized replica exchange method. *J. Chem. Phys.*, 138(10):104119, 2013.

[62] Kai Wang, John D Chodera, Yanzhi Yang, and Michael R Shirts. Identifying ligand binding sites and poses using GPU-accelerated Hamiltonian replica exchange molecular dynamics. *J. Comp. Aided Mol. Des.*, 27(12):989–1007, 2013.

TABLE II. Optimized parameters for the analytical model of molecular association for the two systems studied in this work. Uncertainties are implied by the number of reported significant figures.

	weight	p_b	$\bar{u}_b$[a]	σ_b^a	ε^a	$\tilde{u}^a$	n_l
			Ethanol				
mode 1	3.41×10^{-3}	1.13×10^{-6}	−7.30	2.68	5.83	−0.07	4.08
mode 2	2.09×10^{-1}	9.40×10^{-7}	−3.70	2.70	28.6	1.90	4.41
mode 3	6.11×10^{-1}	2.27×10^{-9}	5.83	1.02	22.7	22.7	9.84
			Alanine dipeptide				
mode 1	1.16×10^{-10}	5.33×10^{-8}	−16.32	4.32	4.38	1.28	5.20
mode 2	2.03×10^{-7}	1.62×10^{-8}	−14.31	3.88	3.5	10.8	5.93
mode 3	1.00	0	103.89	4.72	33.6	83.0	19.07
			1-Naphthol				
mode 1	2.92×10^{-8}	7.90×10^{-8}	−14.86	3.36	4.24	1.85	4.85
mode 2	9.92×10^{-7}	7.68×10^{-8}	−12.13	3.32	6.84	11.8	5.06
mode 3	1.00	0	198	3	9.50	250	19.1
			3,4-Diphenyltoluene				
mode 1	1.06×10^{-15}	4.54×10^{-5}	−14.7	3.80	3.66	13.8	3.14
mode 2	1.00	0	172	4.00	5.82	367	49.2

[a] kcal/mol

TABLE III. Alchemical schedule of the softplus perturbation function for the hydration of ethanol.

λ	λ_1	λ_2	α[a]	u_0[b]	w_0[c]
0.000	0.000	0.000	0.400	10.000	0.000
0.067	0.000	0.044	0.400	10.000	-0.521
0.133	0.000	0.089	0.400	10.000	-1.043
0.200	0.000	0.133	0.400	10.000	-1.564
0.267	0.000	0.178	0.400	10.000	-2.086
0.333	0.000	0.400	0.400	8.889	-4.249
0.400	0.000	0.400	0.400	6.667	-3.360
0.467	0.000	0.400	0.400	4.444	-2.471
0.533	0.000	0.400	0.400	2.222	-1.582
0.600	0.000	0.400	0.400	0.000	-0.693
0.667	0.167	0.400	0.400	0.000	-0.404
0.733	0.333	0.400	0.400	0.000	-0.116
0.800	0.500	0.500	0.400	0.000	0.000
0.867	0.667	0.667	0.400	0.000	0.000
0.933	0.833	0.833	0.400	0.000	0.000
1.000	1.000	1.000	0.400	0.000	0.000

[a] $(\text{kcal/mol})^{-1}$
[b] kcal/mol
[c] kcal/mol

TABLE IV. Alchemical schedule of the softplus perturbation function for the hydration of alanine dipeptide.

λ	λ_1	λ_2	α[a]	u_0[b]	w_0[c]
0.000	0.000	0.000	0.400	3.000	0.000
0.067	0.000	0.138	0.400	3.000	-0.778
0.133	0.000	0.258	0.400	3.000	-1.557
0.200	0.000	0.395	0.400	3.000	-2.335
0.267	0.000	0.544	0.400	2.200	-3.113
0.333	0.000	0.651	0.400	2.200	-3.892
0.400	0.000	0.759	0.400	2.200	-5.837
0.467	0.000	0.867	0.400	1.100	-2.947
0.533	0.067	0.867	0.400	0.000	-1.387
0.600	0.200	0.867	0.400	0.000	-1.156
0.667	0.333	0.867	0.400	0.000	-0.925
0.733	0.467	0.867	0.400	0.000	-0.694
0.800	0.600	0.867	0.400	0.000	-0.463
0.867	0.723	0.867	0.400	0.000	-0.232
0.933	0.867	0.867	0.400	0.000	0.000
1.000	1.000	1.000	0.400	0.000	0.000

[a] $(\text{kcal/mol})^{-1}$
[b] kcal/mol
[c] kcal/mol

TABLE V. Alchemical schedule of the softplus perturbation function for the hydration of 1-naphthol.

λ	λ_1	λ_2	α[a]	u_0[b]	w_0[c]
0.000	0.000	0.000	0.400	5.000	0.000
0.067	0.000	0.116	0.400	5.000	-0.778
0.133	0.000	0.231	0.400	5.000	-1.557
0.200	0.000	0.347	0.400	5.000	-2.335
0.267	0.000	0.462	0.400	5.000	-3.113
0.333	0.000	0.578	0.400	5.000	-3.892
0.400	0.000	0.867	0.400	5.000	-5.837
0.467	0.000	0.867	0.400	1.667	-2.947
0.533	0.067	0.867	0.400	0.000	-1.387
0.600	0.200	0.867	0.400	0.000	-1.156
0.667	0.333	0.867	0.400	0.000	-0.925
0.733	0.467	0.867	0.400	0.000	-0.694
0.800	0.600	0.867	0.400	0.000	-0.463
0.867	0.733	0.867	0.400	0.000	-0.232
0.933	0.867	0.867	0.400	0.000	0.000
1.000	1.000	1.000	0.400	0.000	0.000

[a] $(\text{kcal/mol})^{-1}$
[b] kcal/mol
[c] kcal/mol

TABLE VI. Alchemical schedule of the softplus perturbation function for the hydration of 3,4-diphenyltoluene.

λ	λ_1	λ_2	α[a]	u_0[b]	w_0[c]
0.000	0.000	0.000	0.400	4.800	0.000
0.067	0.000	0.131	0.400	4.800	-0.212
0.133	0.000	0.262	0.400	4.800	-1.272
0.200	0.000	0.392	0.400	4.800	-1.908
0.267	0.000	0.524	0.400	4.800	-2.120
0.333	0.000	0.655	0.400	4.800	-2.544
0.400	0.000	0.786	0.400	4.400	-3.392
0.467	0.000	0.917	0.400	4.400	-3.180
0.533	0.050	0.917	0.400	4.400	-2.968
0.600	0.156	0.917	0.400	4.400	-2.756
0.667	0.308	0.917	0.400	4.000	-2.332
0.733	0.460	0.917	0.400	4.000	-1.908
0.800	0.612	0.917	0.400	0.000	-1.484
0.867	0.764	0.917	0.400	0.000	-1.060
0.933	0.917	0.917	0.400	0.000	0.000
1.000	1.000	1.000	0.400	0.000	0.000

[a] $(\text{kcal/mol})^{-1}$
[b] kcal/mol
[c] kcal/mol